Stochastic Approximation Affine Dynamics and Group Consensus over Random Signed Networks

Ge Chen, *Member, IEEE*, Xiaoming Duan, Wenjun Mei, Francesco Bullo, *Fellow, IEEE*

arXiv:1708.06851v1 [math.OC] 22 Aug 2017

***Abstract*—This paper studies stochastic approximation models of affine dynamical systems and their application to multi-agent systems in engineering and sociology. As main contribution, we consider stochastic approximation affine dynamics and provide necessary and sufficient conditions for convergence under non-negative gains, non-positive gains, and arbitrary gains, respectively. We also characterize the system convergence rate, when the system is convergent. Using our general treatment, we provide necessary and sufficient conditions to reach consensus and group consensus for first-order discrete-time multi-agent system over random signed networks and with state-dependent noise. Finally, we extend our results to the setting of multi-dimensional stochastic approximation affine dynamics and characterize the behavior of the multi-dimensional Friedkin-Johnsen model over random interaction networks.**

***Index Terms*—stochastic approximation, affine systems, multi-agent systems, consensus, signed network**

I. Introduction

Distributed coordination of multi-agent systems has drawn much attention from various fields over the past decades. For example, engineers control the formations of mobile robots, satellites, unmanned aircraft, and automated highway systems [9], [26]; physicists and computer scientists model the collective behavior of animals [27], [32]; sociologists investigate the evolution of opinion, belief and social power over social networks [7], [12], [17]. Many models for distributed coordination have been proposed and analyzed; a common thread in all these works is the study of a group of interacting agents trying to achieve a collective behavior by using neighborhood information allowed by the network topology.

Affine dynamical systems are basic first-order dynamical models with application to many practical problems in multi-agent systems, including distributed consensus of multi-agent systems, computation of PageRank, sensor localization of wireless networks, and belief evolution on social networks [12], [25]. If the linear operator in the affine system is time-invariant, then the study of these systems is straightforward. However, practical systems are very often subject to random fluctuations, so that the linear operator in an affine system is time-variant and the affine system may not convergence. To overcome this deficiency and eliminate the effects of uncertainty, a feasible approach is to adopt models based on the stochastic approximation (SA) algorithm.

The main idea of the SA algorithm is as follows: each agent has a memory of its current state. At each time step, each agent updates its state according to a convex combination of its current state and the information received from its neighbors. Critically, the weight accorded to its own state tends to 1 as time grows (as a way to model the accumulation of experience). The earliest SA algorithms were proposed by Robbins and Monro [28] who aimed to solve root finding problems. SA algorithms have then attracted much interest due to many applications such as the study of reinforcement learning [30], consensus protocols in multi-agent systems [5], and fictitious play in game theory [14]. A main tool in the study of SA algorithms (see [19, Chapter 5]) is the ordinary differential equations (ODE) method, which transforms the analysis of asymptotic properties of a discrete-time stochastic process into the analysis of a continuous-time deterministic process.

In this paper, we consider the SA algorithm for a discrete-time affine dynamics with random linear operator; these models are basic first-order protocols with numerous applications in engineering and sociology. A key analysis method for SA algorithms is the so-called ODE method; this method however requires (i) the state of each agent to be uniformly bounded almost surely and (ii) the limit points to be deterministic; e.g., see [19, Chapter 5]. These requirements lead to sufficient conditions for the convergence of SA affine dynamics, but not to necessary and sufficient conditions. This paper develops appropriate discrete-time analysis methods and provides necessary and sufficient conditions for the convergence of SA affine dynamics. Under our necessary and sufficient conditions, the state of the system will converge to a random vector. An additional restriction of traditional SA algorithms is that only non-negative gain schedules are considered. This paper relaxes also this requirement and provides necessary and sufficient conditions for convergence of SA affine dynamics under non-positive gains and arbitrary gains. In addition, we analyze the convergence rate of the system when it is convergent.

Our general theoretical results are directly applicable to certain multi-agent systems. The first application is to the study of consensus problems in multi-agent systems. As it is well known, numerous works provide sufficient conditions for consensus in time-varying multi-agent systems with row-

This material is based upon work supported by, or in part by, the U.S. Army Research Laboratory and the U.S. Army Research Office under grant number W911NF-15-1-0577. The research of G. Chen was supported in part by the National Natural Science Foundation of China under grants 91427304 and 61673373.

Ge Chen is with the National Center for Mathematics and Interdisciplinary Sciences & Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China, chenge@amss.ac.cn

Xiaoming Duan, Wenjun Mei, and Francesco Bullo are with the Department of Mechanical Engineering and the Center of Control, Dynamical-Systems and Computation, University of California at Santa Barbara, CA 93106-5070, USA. xiaomingduan.zju@gmail.com; meiwenjunbd@gmail.com; bullo@engineering.ucsb.edu

stochastic interaction matrices; an incomplete list of references is [4], [5], [8], [20], [22], [29]; see also the classic works [3], [6], [31]. Recently, motivated by the study of antagonistic interactions in social networks, novel concepts of bipartite, group, and cluster consensus have been studied over signed networks (mainly focusing on continuous-time dynamical models); see [1], [21], [24], [33]. In this paper, we apply and extend our results on SA affine dynamics to the setting of first-order discrete-time multi-agent system over random signed networks and with state-dependent noise; for such models, we provide novel necessary and sufficient conditions to reach consensus and group consensus.

As the second application of our results, we study the Friedkin-Johnsen (FJ) model of opinion dynamics in social networks. The FJ model was first proposed in [11], where each agent is assumed to be susceptible to other agents' opinions but also to be anchored to his own initial opinion with a certain level of stubbornness. Ravazzi *et al.* proposed a gossip version of the FJ model in [25], whereby each link in the network is sampled uniformly and the agents associated with the link meet and update their opinions. The agents' opinions were proven to converge in mean square. Frasca *et al.* considered a symmetric pairwise randomization of FJ in [10], whereby a pair of agents are chosen to update their opinions. Our work, by exploiting stochastic approximation, largely relaxes the conditions for convergence when applied to FJ model over random interaction networks. The sociological meaning of stochastic approximated FJ model is that agents have cumulative memory about their previous opinions. The adoption of SA models in the study of human behavior is widely adopted in game theory and economics; e.g., see [14].

The main contributions of this paper are summarized as follows.

1) We propose a stochastic approximation version of a time-varying affine system and provide necessary and sufficient conditions to guarantee convergence by developing appropriate discrete-time methods. The convergence rate is also obtained when the system is convergent.
2) Using our results, we get the necessary and sufficient conditions to reach consensus and group consensus of the first-order discrete-time multi-agent system over random signed networks and with state-dependent noise for the first time.
3) We extend our results to the multi-dimensional SA affine dynamics and provide applications to the multi-dimensional FJ model over random interaction networks.

Organization: The remainder of this paper is organized as follows. We briefly review the time-varying affine dynamics and propose a stochastic approximation version of it in Section II. The main results are presented in Section III. In particular, we introduce some preliminaries and assumptions in Subsection III-A. Sufficient conditions that guarantee the convergence of the affine dynamical system are obtained in Subsection III-B. We provide the results on convergence rate in the same subsection. In Subsection III-C, we prove that the sufficient condition is also necessary. The necessary and sufficient conditions for convergence are then summarized in Subsection III-D. We generalize the results to multi-dimensional models and discuss their application to group consensus and the FJ model in Section IV. Section V concludes the paper.

II. Affine Systems

A. Review of time-varying affine dynamics

In [13], [25] a time-varying affine system was considered as follows:

$$x(s+1) = P(s)x(s) + u(s), \quad s = 0, 1, \dots, \tag{1}$$

where $P(s) \in \mathbb{R}^{n\times n}$ is a matrix associated to the communication network between agents, and $u(s) \in \mathbb{R}^n$ is an input vector. Given a matrix $A \in \mathbb{R}^{n\times n}$, let $\rho(A)$ denote its spectral radius, i.e., $\rho(A) = \max_i |\lambda_i(A)|$, where $\lambda_i(A)$ is an eigenvalue of A. For system (1), if $P(s) \equiv P$, $u(s) \equiv u$, and $\rho(A) < 1$, then it is immediate to see that $x(s)$ converges to $(I_n - P)^{-1}u$. However, when $P(s)$ or $u(s)$ is time-varying the system (1) does not necessarily converge. As an alternative, Ravazzi *et al.* [25] investigate the ergodicity of system (1) as follows.

Proposition II.1 (Theorem 1 in [25]): Consider system (1) and assume $\{P(s)\}$ and $\{u(s)\}$ are i.i.d. and have finite first moments. Suppose there exists a constant $\alpha \in (0, 1]$ such that

$$\mathbb{E}[P(s)] = (1-\alpha)I_n + \alpha P, \quad \mathbb{E}[u(s)] = \alpha u, \quad \forall s \geq 0.$$

If $\rho(P) < 1$, then $x(s)$ converges to a random variable in distribution, and $\frac{1}{s}\sum_{k=0}^{s-1} x(k)$ converges to $(I_n - P)^{-1}u$ almost surely.

In a separate line of inquiry, Han *et al.* [13] provide some sufficient conditions for cluster consensus with some special settings on $P(s)$ and $u(s)$.

In this paper we adopt the stochastic approximation method to average the effect of the time-varying $P(s)$ and $u(s)$ to the state $x(s)$. In this case we study the sufficient and necessary conditions for convergence of $x(s)$, and also obtain a convergence rate.

B. SA affine dynamics over time-varying networks

In this subsection we consider the stochastic-approximation version of system (1), formulated as:

$$\begin{aligned} x(s+1) &= (1-a(s))x(s) \\ &\quad + a(s)[P(s)x(s) + u(s)], \quad s = 0, 1, \dots, \end{aligned} \tag{2}$$

where $a(s) \in \mathbb{R}$ is the gain function. Compared to system (1), each agent updates its state depending not only on the affine map $P(s)x(s) + u(s)$ but also on its own current state. If $a(s) = \frac{1}{s+1}$, then $x(s+1)$ equals the approximate average value of the previous s affine maps because $x(s)$ carries the information of the previous $s-1$ affine maps. Such a system is a basic first-order discrete-time multi-agent system with many applications like computation of PageRank [34], sensor localization of wireless networks [18], distributed consensus of multi-agent systems, and belief evolution on social networks.

We will study the mean-square convergence of system (2) whose definition is given as follows:

Definition II.1: For an n-dimensional random vector x, we say $x(s)$ converges to x in mean square if

$$\mathbb{E}\|x\|_2^2 < \infty \quad \text{and} \quad \lim_{s\to\infty} \mathbb{E}\|x(s) - x\|_2^2 = 0. \tag{3}$$

Also, we say $\{x(s)\}$ is mean-square convergent if there exists an n-dimensional random vector x such that (3) holds.

III. Main results

A. Informal statement of main results

We start with some notation. Given a matrix $A \in \mathbb{R}^{n\times n}$, define $\widetilde{\rho}_{\max}(A) := \max_i \mathrm{Re}(\lambda_i(A))$ and $\widetilde{\rho}_{\min}(A) := \min_i \mathrm{Re}(\lambda_i(A))$ to be the maximum and minimum values of the real parts of the eigenvalues of A respectively. It is easy to show that $|\widetilde{\rho}_{\max}(A)| \le \rho(A)$.

For $\{P(s)\}$ and $\{u(s)\}$, we relax the i.i.d. condition in [25] to the following assumption:

(A1) Suppose there exist a matrix $P \in \mathbb{R}^{n\times n}$ and a vector $u \in \mathbb{R}^n$ such that $\mathbb{E}[P(s)\,|\,x(s)] = P$ and $\mathbb{E}[u(s)\,|\,x(s)] = u$ for any $s \ge 0$ and $x(s) \in \mathbb{R}^n$. Also, assume $\mathbb{E}[\|P(s)\|_2^2\,|\,x(s)]$ and $\mathbb{E}[\|u(s)\|_2^2\,|\,x(s)]$ are uniformly bounded.

For $\{a(s)\}$, generally SA algorithms use the following assumption:

(A2) Assume $\{a(s)\}$ are non-negative real numbers independent with $\{x(s)\}$, and satisfying $\sum_{s=0}^{\infty} a(s) = \infty$ and $\sum_{s=0}^{\infty} a^2(s) < \infty$.

We will consider the following alternative assumption.

(A2') Assume $\{a(s)\}$ are non-positive real numbers independent with $\{x(s)\}$, and satisfying $\sum_{s=0}^{\infty} a(s) = -\infty$ and $\sum_{s=0}^{\infty} a^2(s) < \infty$.

Under the assumptions (A1) and (A2), we later show the necessary and sufficient condition for the convergence of $x(s)$ in system (2) is $\widetilde{\rho}_{\max}(P) < 1$, or $\widetilde{\rho}_{\max}(P) = 1$ together with the following condition for P and u:

(A3) Assume any eigenvalue of P whose real part is 1 equals 1, and the eigenvalue 1 has the same algebraic and geometric multiplicities, and $\xi u = 0$ for any left eigenvector ξ (row vector) of P corresponding to the eigenvalue 1.

Similarly, under (A1) and (A2') the necessary and sufficient condition for the convergence of $x(s)$ is $\widetilde{\rho}_{\min}(P) > 1$, or $\widetilde{\rho}_{\min}(P) = 1$ with (A3).

Also, we will study the convergence rates when $x(s)$ is convergent, and the convergence conditions when $\{a(s)\}$ are arbitrary real numbers.

B. Sufficient convergence conditions and convergence rates

Recall that P and u are the expectations of $P(s)$ and $u(s)$ respectively. Let

$$P = H^{-1}\mathrm{diag}(J_1, \ldots, J_K)H := H^{-1}DH, \tag{4}$$

where $H \in \mathbb{C}^{n\times n}$ is an invertible matrix, and D is the Jordan normal form of P with

$$J_i = \begin{bmatrix} \lambda_{i'}(P) & 1 & & \\ & \lambda_{i'}(P) & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_{i'}(P) \end{bmatrix}_{m_i\times m_i}$$

for $1 \le i \le K$.

Let r be the algebraic multiplicity of the eigenvalue 1 of P. We first consider the case $\widetilde{\rho}_{\max}(P) = 1$ (or $\widetilde{\rho}_{\min}(P) = 1$) with (A3), which implies that $r \ge 1$ and that the geometric multiplicity of the eigenvalue 1 is equal to r. We choose a suitable H such that $\lambda_1(P) = \cdots = \lambda_r(P) = 1$. Then the Jordan normal form D can be written as

$$D = \begin{bmatrix} I_r & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(n-r)\times r} & \underline{D}_{(n-r)\times(n-r)} \end{bmatrix} \in \mathbb{C}^{n\times n}, \tag{5}$$

where $\underline{D} := \mathrm{diag}(J_{r+1}, \ldots, J_K) \in \mathbb{C}^{(n-r)\times(n-r)}$. For any vector $y \in \mathbb{C}^n$, throughout this subsection we set $\bar{y} := (y_1, \ldots, y_r)^\top$ and $\underline{y} := (y_{r+1}, \ldots, y_n)^\top$.

Theorem III.1: (Convergence of SA affine system at critical point) Consider the system (2) satisfying (A1), (A2), and (A3) with $\widetilde{\rho}_{\max}(P) = 1$, or satisfying (A1), (A2'), and (A3) with $\widetilde{\rho}_{\min}(P) = 1$. Let H be the matrix defined by (4) such that the Jordan normal form D has the form of (5). Then, for any initial state, $x(s)$ converges to $H^{-1}y$ in mean square, where $\bar{y}$ is a random vector satisfying $\mathbb{E}\bar{y} = \overline{Hx(0)}$ and $\mathbb{E}\|\bar{y}\|_2^2 < \infty$, and $\underline{y} = (I_{n-r} - \underline{D})^{-1}\underline{Hu}$.

Proof. Let $y(s) := Hx(s)$, $v(s) := Hu(s)$ and $D(s) := HP(s)H^{-1}$, then by (2) we have

$$\begin{aligned} &H^{-1}y(s+1) \\ &= (1 - a(s))H^{-1}y(s) + a(s)[P(s)H^{-1}y(s) + u(s)], \end{aligned} \tag{6}$$

which implies

$$y(s+1) = y(s) + a(s)[(D(s) - I_n)y(s) + v(s)]. \tag{7}$$

Let $v := Ev(s) = Hu$. From (4) we have $HP = DH$, which implies $H_iP = H_i$ for $1 \le i \le r$, where H_i is the i-th row of the matrix H. Thus, H_i, $1 \le i \le r$, is a left eigenvector corresponding to the eigenvalue 1. By (A3) we have

$$v_i = H_i u = 0, \quad \forall 1 \le i \le r. \tag{8}$$

Recall that $\underline{v} = (v_{r+1}, \ldots, v_n)^\top$. Also, $I_{n-r} - \underline{D}$ is an invertible matrix, so we can set

$$z := \begin{bmatrix} \mathbf{0}_{r\times 1} \\ (I_{n-r} - \underline{D})^{-1}\underline{v} \end{bmatrix} \in \mathbb{C}^n.$$

From (5) and (8) we have

$$(D - I_n)z + v = \mathbf{0}_{n\times 1}. \tag{9}$$

Set $\theta(s) := y(s) - z$. From (7) we obtain

$$\theta(s+1) = \theta(s) + a(s)[(D(s) - I_n)(\theta(s) + z) + v(s)]. \tag{10}$$

We first consider the case when $\widetilde{\rho}_{\max}(P) = 1$, which implies that $\underline{D} - I_{n-r}$ is a Hurwitz matrix. Thus, by the stability theory of continuous Lyapunov equation (see [15, Corollary 2.2.4]), there exists a Hermitian positive definite matrix $A \in \mathbb{C}^{(n-r)\times(n-r)}$ such that

$$(\underline{D} - I_{n-r})^* A + A(\underline{D} - I_{n-r}) = -I_{n-r}, \tag{11}$$

where $(\cdot)^*$ denotes the conjugate transpose of the matrix or vector. Set

$$A_1 := \begin{bmatrix} I_r & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(n-r)\times r} & A_{(n-r)\times(n-r)} \end{bmatrix} \in \mathbb{C}^{n\times n},$$

then A_1 is still a Hermitian positive definite matrix. Define the Lyapunov function $V_1(\theta) := \theta^* A_1 \theta$. By (10), (A1) and (11), for any $\theta(s)$ we have

$$\begin{aligned}&\mathbb{E}[V_1(\theta(s+1))|\theta(s)] \\ &\quad\le V_1(\theta(s)) + a(s)\theta^*(s)\big[(D-I_n)^* A_1 + A_1(D-I_n)\big]\theta(s) \\ &\qquad + O\big(a^2(s)(\|\theta(s)\|_2^2+1)\big)^1. \end{aligned} \tag{12}$$

From (5) and (11), we obtain

$$\begin{aligned}&(D-I_n)^* A_1 + A_1(D-I_n) \\ &= \begin{bmatrix} \mathbf{0}_{r\times r} & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(n-r)\times r} & (\underline{D}-I_{n-r})^* A + A(\underline{D}-I_{n-r}) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{0}_{r\times r} & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(n-r)\times r} & -I_{n-r} \end{bmatrix}, \end{aligned} \tag{13}$$

so (12) implies

$$\mathbb{E}[V_1(\theta(s+1))] \le [1 + c_1 a^2(s)]\mathbb{E}[V_1(\theta(s))] + c_2 a^2(s), \tag{14}$$

where c_1 and c_2 are two positive constants. Using (14) repeatedly we get

$$\begin{aligned}&\mathbb{E}[V_1(\theta(s+1))] \\ &\le \prod_{i=0}^{s}[1+c_1a^2(i)] + \sum_{i=0}^{s} c_2 a^2(i) \prod_{j=i+1}^{s}[1+c_1a^2(j)] \\ &< \infty \quad \text{as} \quad s\to\infty, \end{aligned} \tag{15}$$

where the last inequality uses the condition that $\sum_{s=0}^{\infty} a^2(s) < \infty$. Also, because A_1 is a Hermitian positive definite matrix,

$$\frac{1}{\rho(A_1)} V_1(\theta(s)) \le \|\theta(s)\|_2^2 \le \frac{1}{\lambda_{\min}(A_1)} V_1(\theta(s)). \tag{16}$$

Combining (15) and (16) yields

$$\sup_s \mathbb{E}\|\theta(s)\|_2^2 \le \sup_s \frac{\mathbb{E}[V_1(\theta(s))]}{\lambda_{\min}(A_1)} < \infty. \tag{17}$$

Inequality (17) shows that $\theta(s)$ will not diverge, however we need to prove its convergence. We first consider the convergence of $\underline{\theta}(s)$. Set

$$A_2 := \begin{bmatrix} \mathbf{0}_{r\times r} & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(n-r)\times r} & A_{(n-r)\times(n-r)} \end{bmatrix} \in \mathbb{C}^{n\times n}$$

and define $V_2(\theta) := \theta^* A_2 \theta = \underline{\theta}^* A \underline{\theta}$. Similar to (12), we have

$$\begin{aligned}&\mathbb{E}[V_2(\theta(s+1))] \\ &\le \mathbb{E}\Big[V_2(\theta(s)) + a(s)\theta^*(s)(D-I_n)^* A_2 \theta(s) \\ &\quad + a(s)\theta^*(s) A_2 (D-I_n)\theta(s) + O\big(a^2(s)(\|\theta(s)\|_2^2+1)\big)\Big] \\ &= \mathbb{E}\Big[V_2(\theta(s)) + a(s)\theta^*(s)\begin{bmatrix} \mathbf{0}_{r\times r} & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(n-r)\times r} & -I_{n-r} \end{bmatrix}\theta(s)\Big] \\ &\quad + O(a^2(s)) \\ &\le \Big(1 - \frac{a(s)}{\rho(A)}\Big)\mathbb{E}[V_2(\theta(s))] + O\big(a^2(s)\big), \end{aligned} \tag{18}$$

where the forth line uses (13) and (17), and the last inequality does a similar computation as (16). By (18) and Lemma A.1 in Appendix A, we obtain $\lim_{s\to\infty}\mathbb{E}[V_2(\theta(s))] = 0$, which implies

$$\lim_{s\to\infty} \mathbb{E}\|\underline{\theta}(s)\|_2^2 = 0. \tag{19}$$

It remains to consider the convergence of $\bar{\theta}(s)$. Set

$$A_3 := \begin{bmatrix} I_r & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(n-r)\times r} & \mathbf{0}_{(n-r)\times(n-r)} \end{bmatrix} \in \mathbb{R}^{n\times n},$$

and define $V_3(\theta) = \theta^* A_3 \theta = \bar{\theta}^* \bar{\theta}$. By (5) and (8) we get $A_3(D-I_n) = \mathbf{0}_{n\times n}$ and $A_3 v = \mathbf{0}_{n\times 1}$, thus by (A1) for any $i<j$ we have

$$\begin{aligned}&E\Big[[(D(i)-I_n)(\theta(i)+z)+v(i)]^* A_3 \\ &\qquad \times [(D(j)-I_n)(\theta(j)+z)+v(j)]\Big] \\ &= E\Big[[(D(i)-I_n)(\theta(i)+z)+v(i)]^* A_3 \\ &\qquad \times [(D-I_n)(\theta(j)+z)+v]\Big] = 0. \end{aligned} \tag{20}$$

Similarly, the equation (20) still holds for $i>j$. From these and (10) we get for any $s_2 > s \ge 0$,

$$\begin{aligned}&\mathbb{E}[V_3(\theta(s_2)-\theta(s))] \\ &= \mathbb{E}\Big[V_3\Big(\sum_{i=s}^{s_2-1}[\theta(i+1)-\theta(i)]\Big)\Big] \\ &= \mathbb{E}\Big[\Big(\sum_{i=s}^{s_2-1} a(i)[(D(i)-I_n)(\theta(i)+z)+v(i)]\Big)^* A_3 \\ &\quad \times \Big(\sum_{i=s}^{s_2-1} a(i)[(D(i)-I_n)(\theta(i)+z)+v(i)]\Big)\Big] \\ &= \sum_{i=s}^{s_2-1} a^2(i)\mathbb{E}\big[[(D(i)-I_n)(\theta(i)+z)+v(i)]^* \\ &\quad \times A_3[(D(i)-I_n)(\theta(i)+z)+v(i)]\big] \\ &= O\Big(\sum_{i=s}^{s_2-1} a^2(i)\Big), \end{aligned} \tag{21}$$

where the last line uses (A1) and (17). Since $\sum_{i=0}^{\infty} a^2(i) < \infty$, from (21) we have

$$\begin{aligned}&\lim_{s\to\infty}\lim_{s_2\to\infty} \mathbb{E}\|\bar{\theta}(s_2)-\bar{\theta}(s)\|_2^2 \\ &\qquad = \lim_{s\to\infty}\lim_{s_2\to\infty} \mathbb{E}[V_3(\theta(s_2)-\theta(s))] = 0. \end{aligned} \tag{22}$$

By the Cauchy criterion (see [16, page 58]), $\bar{\theta}(s)$ has a mean square limit $\bar{\theta}(\infty)$. Also, from (10), (A1) and (8) we have

$$\begin{aligned}&\mathbb{E}[A_3\theta(s+1)] \\ &\quad = \mathbb{E}\big[\mathbb{E}[A_3\theta(s+1)\,|\,\theta(s)]\big] \\ &\quad = \mathbb{E}\big[A_3\theta(s) + a(s)A_3[(D-I_n)(\theta(s)+z)+v]\big] \\ &\quad = \mathbb{E}[A_3\theta(s)] = \cdots = A_3\theta(0), \end{aligned} \tag{23}$$

which is followed by

$$\mathbb{E}\bar{\theta}(\infty) = \bar{\theta}(0) = \bar{y}(0) = \overline{Hx}(0). \tag{24}$$

We remark that $x(s) = H^{-1}[\theta(s)+z]$. Let y be a vector satisfying $\underline{y} = \underline{z} = (I_{n-r} - \underline{D})^{-1}\underline{v}$ and $\bar{y} = \bar{\theta}(\infty) + \bar{z} =$

[1] Given two sequences of positive numbers $\{g_1(s)\}_{s=0}^{\infty}$ and $\{g_2(s)\}_{s=0}^{\infty}$, we say $g_1(s) = O(g_2(s))$ if there exist a constants $c>0$ such that $g_1(s) \le c g_2(s)$ for all $s \ge 0$.

$\bar{\theta}(\infty)$. By (19) and (22) we have that $x(s)$ converges to $H^{-1}y$ in mean square. By (24) and (17) we get $\mathbb{E}\bar{y} = \overline{H}x(0)$ and $\mathbb{E}\|\bar{y}\|_2^2 < \infty$.

For the case that $\widetilde{\rho}_{\min}(P) = 1$, which implies $I_{n-r} - \underline{D}$ is a Hurwitz matrix. Set $b(s) = -a(s) \geq 0$ and substitute it to (10) we obtain

$$\theta(s+1) = \theta(s) + b(s)[(I_n - D(s))(\theta(s) + z) - v(s)].$$

Finally, a process similar to that from (11) to (24) yields our result. □

For the case when $\widetilde{\rho}_{\max}(P) < 1$ or $\widetilde{\rho}_{\min}(P) > 1$, from the proof of Theorem III.1 we have the following proposition:

Proposition III.1: Consider the system (2) satisfying (A1), (A2) and $\widetilde{\rho}_{\max}(P) < 1$, or satisfying (A1), (A2') and $\widetilde{\rho}_{\min}(P) > 1$. Then, for any initial state, $x(s)$ converges to $(I_n - P)^{-1}u$ in mean square.

Proof. We can set $r = 0$ in the proof of Theorem III.1, then we obtain that $x(s)$ converges to $H^{-1}(I_n - D)^{-1}Hu = (I_n - P)^{-1}u$ in mean square. □

Remark 1: If $\|x(s)\|$ is uniformly bounded a.s. in addition to the conditions of Proposition III.1, then we can obtain that $x(s)$ converges to $(I_n - P)^{-1}u$ almost surely by the ODE method in stochastic approximation (Theorem 5.2.1 in [19] or Theorem 2.2 in [2]).

Next, we give the convergence rate when $x(s)$ is mean-square convergent.

Theorem III.2: (Convergence rate of SA affine system) Consider the system (2) satisfying (A1) and one of the following four cases: i) $\widetilde{\rho}_{\max}(P) < 1$; ii) $\widetilde{\rho}_{\min}(P) > 1$; iii) $\widetilde{\rho}_{\max}(P) = 1$ with (A3); and iv) $\widetilde{\rho}_{\min}(P) = 1$ with (A3). Let $\beta > 0, \gamma \in (\frac{1}{2}, 1]$, and α be a large positive number. Choose $a(s) = \frac{\alpha}{(s+\beta)^\gamma}$ if $\widetilde{\rho}_{\max}(P) \leq 1$, and $a(s) = \frac{-\alpha}{(s+\beta)^\gamma}$ if $\widetilde{\rho}_{\min}(P) \geq 1$. Then for any initial state,

$$\mathbb{E}\big\|x(s) - x\big\|_2^2 = \begin{cases} O(s^{-\gamma}), & \text{if } \widetilde{\rho}_{\max}(P) < 1 \text{ or } \widetilde{\rho}_{\min}(P) > 1 \\ O(s^{1-2\gamma}), & \text{if } \widetilde{\rho}_{\max}(P) = 1 \text{ or } \widetilde{\rho}_{\min}(P) = 1 \end{cases}$$

where x is a mean square limit of $x(s)$ whose expression is provided by Theorem III.1 and Proposition III.1.

The proof of this theorem is postponed to Appendix B.

C. Sufficient conditions for non-convergence

We first consider the conditions of non-convergence under the assumptions (A1) and (A2) or (A2'):

Theorem III.3: Consider the system (2) satisfying (A1). Then:
i) If $\widetilde{\rho}_{\max}(P) > 1$, or $\widetilde{\rho}_{\max}(P) = 1$ but (A3) does not hold, there exist some initial states such that $x(s)$ is not mean-square convergent for any $\{a(s)\}$ satisfying (A2).
ii) If $\widetilde{\rho}_{\min}(P) < 1$, or $\widetilde{\rho}_{\min}(P) = 1$ but (A3) does not hold, there exist some initial states such that $x(s)$ is not mean-square convergent for any $\{a(s)\}$ satisfying (A2').

The proof of this theorem is postponed to Appendix C.

The condition of non-convergence in Theorem III.3 has a constraint that the gain function $\{a(s)\}$ must satisfy the assumption (A2) or (A2'). An interesting problem is to understand what happens if $\{a(s)\}$ are chosen as arbitrary real numbers. Obviously, from protocol (2) if $\{a(s)\}$ has only finite non-zero elements, then $x(s)$ will converge to a random variable. Thus, we only consider the setting whereby $x(s)$ does not converge to a deterministic vector for arbitrary gains.

Recall that

$$P = H^{-1}\text{diag}(J_1, \ldots, J_K)H = H^{-1}DH,$$

where $H \in \mathbb{C}^{n\times n}$ is an invertible matrix, and D is the Jordan normal form of P. For $1 \leq i \leq K$, define

$$\widetilde{I}_i = \text{diag}(0, \ldots, I_{m_i}, \ldots, 0) \in \mathbb{R}^{n\times n}, \tag{25}$$

which corresponds to the Jordan block J_i and then $D\widetilde{I}_i = \text{diag}(0, \ldots, J_i, \ldots, 0)$. To study the condition for non-convergence of system (2), we need the following two assumptions:

(A4) Assume there is a Jordan block J_j in D associated with the eigenvalue $\lambda_{j'}(P)$ such that $\text{Re}(\lambda_{j'}(P)) = 1$ and

$$\mathbb{E}\big[\|\widetilde{I}_j H[(P(s) - P)x(s) + u(s) - u]\|_2^2 \,|\, x(s)\big] \geq c_1\|x(s)\|_2^2 + c_2 \tag{26}$$

for any $s \geq 0$ and $x(s) \in \mathbb{R}^n$, where P, u, H, D and $\widetilde{I}_j$ are defined by (A1), (4), and (25), and c_1 and c_2 are constants satisfying $c_1 \geq 0$, $c_2 \geq 0$, and $c_1 + c_2 > 0$.

(A4') Assume there are two Jordan blocks J_{j_1} and J_{j_2} associated with the eigenvalues $\lambda_{j_1'}(P)$ and $\lambda_{j_2'}(P)$ respectively such that $\text{Re}(\lambda_{j_1'}(P)) < 1 < \text{Re}(\lambda_{j_2'}(P))$ and (26) holds for $j = j_1, j_2$.

Theorem III.4: Consider the system (2) satisfying (A1) and (A4) or (A4'). In addition, assume there exists a constant $c_3 > 0$ such that for any $s \geq 0$ and $x(s) \in \mathbb{R}^n$,

$$\mathbb{E}\big[\|(P(s) - P)x(s) + u(s) - u\|_2^2 | x(s)\big] \geq c_3. \tag{27}$$

Then for any deterministic vector $b \in \mathbb{R}^n$, any initial state $x(0) \neq b$, and any real number sequence $\{a(s)\}_{s\geq 0}$ independent with $\{x(s)\}_{s\geq 0}$, $x(s)$ cannot converge to b in mean square.

The proof of this theorem is postponed to Appendix D.

If $u(s)$ is a degenerate random vector which means that $\mathbb{E}\|u(s) - u\|_2^2 = 0$, then the condition (27) may not be satisfied.

Theorem III.5: Consider the system (2) satisfying (A1), and $\mathbb{E}[\|u(s) - u\|_2^2 \,|\, x(s)] = 0$ for any $s \geq 0$ and $x(s) \in \mathbb{R}^n$. Assume (A4) or (A4') holds but using

$$\mathbb{E}\big[\|\widetilde{I}_j H(P(s) - P)x(s)\|_2^2 \,|\, x(s)\big] \geq c_1\|x(s)\|_2^2 \tag{28}$$

instead of (26). For any deterministic vector $b \in \mathbb{R}^n$ and any initial state $x(0) \neq b$, if one of the following three conditions holds:
i) $u \neq \mathbf{0}_{n\times 1}$ and $x(0) \neq \mathbf{0}_{n\times 1}$;
ii) $u \neq \mathbf{0}_{n\times 1}$, $x(0) = \mathbf{0}_{n\times 1}$, and $b \neq \alpha u$ for any $\alpha \in \mathbb{R}$; or
iii) $u = \mathbf{0}_{n\times 1}$, and the eigenvalues $\lambda_{j'}(P)$ in (A4), or $\lambda_{j_1'}(P)$

and $\lambda_{j_2'}(P)$ in (A4') are not real numbers,
then $x(s)$ cannot converge to b in mean square for any real number sequence $\{a(s)\}_{s\geq 0}$ independent with $\{x(s)\}_{s\geq 0}$.

The proof of this theorem is postponed to Appendix E.

D. Necessary and sufficient conditions for convergence

From Theorems III.1 and III.3 and Proposition III.1, the following necessary and sufficient conditions for convergence with non-negative and non-positive gains are obtained immediately.

Theorem III.6: (Necessary and sufficient condition for convergence of SA affine system with non-negative gains) Consider the system (2) satisfying (A1) and (A2). Then $x(s)$ is mean-square convergent for any initial state if and only if $\widetilde{\rho}_{\max}(P)<1$, or $\widetilde{\rho}_{\max}(P)=1$ with (A3).

Theorem III.7: (Necessary and sufficient condition for convergence of SA affine system with non-positive gains) Consider the system (2) satisfying (A1) and (A2'). Then $x(s)$ is mean-square convergent for any initial state if and only if $\widetilde{\rho}_{\min}(P)>1$, or $\widetilde{\rho}_{\min}(P)=1$ with (A3).

Remark 2: Theorem III.6 cannot be obtained by the general tool of the ODE method in SA theory due to two reasons. First, the ODE method requires the limit points to be deterministic, but, by Theorem III.1, our sequence $x(s)$ may converge to a random vector when $\widetilde{\rho}_{\max}(P)=1$. Second, the ODE method requires the state of each agent to be uniformly bounded almost surely. This assumption makes it impossible to establish the necessary condition for convergence

Remark 3: Compared to Theorem 1 in [25], Theorem III.6 extends the convergence condition from $\rho(P)<1$ to the sufficient and necessary condition. In fact, for the basic affine protocol $x(s+1)=Px(s)+u$, $x(s)$ converges if and only if $\rho(P)<1$. However, if we consider the time-varying affine maps and adopt the SA method to eliminate the effect of fluctuation, then the convergence condition can be substantially weakened.

Theorems III.6 and III.7 have a constraint that the gain function $\{a(s)\}$ must satisfy the assumption (A2) or (A2'). Without this constraint we can get the following necessary and sufficient condition for convergence to a deterministic vector, but with some additional conditions on $\{u(s)\}$ or $\{P(s)\}$.

Theorem III.8 (Necessary and sufficient condition for convergence of SA affine dynamics with arbitrary gains): Consider the SA affine system (2) which satisfies (A1). Suppose there exists a constant $c\in(0,1)$ such that for any $s\geq 0$, $x(s)\in\mathbb{R}^n$, $\xi_1,\ldots,\xi_m\in\{P_{ij}(s),1\leq i,j\leq n;u_i(s),1\leq i\leq n\}$ and $c_1,\ldots,c_m\in\mathbb{C}$,

$$\mathbb{E}\Big[\Big|\sum_{i=1}^m c_i(\xi_i-\mathbb{E}\xi_i)\Big|^2\,|\,x(s)\Big] \geq c\sum_{i=1}^m |c_i|^2\mathbb{E}\big[(\xi_i-\mathbb{E}\xi_i)^2\,|\,x(s)\big]. \quad (29)$$

In addition, assume one of the following two conditions holds:
i) $\inf_{k,s}\mathbb{E}[(u_k(s)-u_k)^2\,|\,x(s)]>0$.
ii) $\mathbb{E}[\|u(s)-u\|^2\,|\,x(s)]=0$, $u\neq\mathbf{0}_{n\times 1}$, $x(0)\neq\mathbf{0}_{n\times 1}$, and $\inf_{i,j,s}\mathbb{E}[(P_{ij}(s)-P_{ij})^2\,|\,x(s)]>0$.
Then we can choose real number sequence $\{a(s)\}_{s\geq 0}$ independent with $\{x(s)\}_{s\geq 0}$ such that $x(s)$ converges to a deterministic vector different from $x(0)$ in mean square if and only if $\widetilde{\rho}_{\max}(P)<1$ or $\widetilde{\rho}_{\min}(P)>1$.

Proof. If $\widetilde{\rho}_{\max}(P)<1$ or $\widetilde{\rho}_{\min}(P)>1$, by Proposition III.1 we obtain that $x(s)$ converges to $(I_n-P)^{-1}u$ in mean square.

For $\widetilde{\rho}_{\min}(P)\leq 1\leq\widetilde{\rho}_{\max}(P)$, we set $\widetilde{P}(s):=P(s)-P$ and $\widetilde{u}(s):=u(s)-u$. Define H and K by (4), and define $\widetilde{I}_i$ by (25). For any $j\in\{1,\ldots,K\}$, since H is an invertible matrix, $\widetilde{I}_jH$ contains at least one non-zero row $H_{j'}$. Thus, for any $x(s)\in\mathbb{R}^n$ we have

$$\begin{aligned}&\mathbb{E}\big[\|\widetilde{I}_jH[\widetilde{P}(s)x(s)+\widetilde{u}(s)]\|_2^2\,|\,x(s)\big]\\ &\quad\geq\mathbb{E}\big[|H_{j'}[\widetilde{P}(s)x(s)+\widetilde{u}(s)]|^2\,|\,x(s)\big]\\ &\quad=\mathbb{E}\Big[\Big|\sum_{i,k}H_{j'i}\widetilde{P}_{ik}(s)x_k(s)+\sum_i H_{j'i}\widetilde{u}_i(s)\Big|^2\,|\,x(s)\Big]\\ &\quad\geq c\sum_{i,k}|H_{j'i}|^2\mathbb{E}\big[\widetilde{P}_{ik}^2(s)\,|\,x(s)\big]x_k^2(s)\\ &\qquad+c\sum_i|H_{j'i}|^2\mathbb{E}\big[\widetilde{u}_i^2(s)\,|\,x(s)\big],\end{aligned} \quad (30)$$

where the last inequality uses (29).

If Condition i) holds, we have there exists a constant $d_1>0$ such that $\mathbb{E}\big[\widetilde{u}_i^2(s)\,|\,x(s)\big]\geq d_1$ for $s\geq 0$ and $1\leq i\leq n$. Combing this with (30) and the assumption $\widetilde{\rho}_{\min}(P)\leq 1\leq\widetilde{\rho}_{\max}(P)$, we obtain that (27) and (A4) or (A4') hold. By Theorem III.4, $x(s)$ cannot converge to a deterministic vector different from $x(0)$ in mean square.

If Condition ii) holds, we have $\mathbb{E}[\|\widetilde{u}(s)\|_2^2\,|\,x(s)]=0$ and there exists a constant $d_2>0$ such that $\mathbb{E}\big[\widetilde{P}_{ik}^2(s)\,|\,x(s)\big]\geq d_2$ for $s\geq 0$ and $1\leq i,k\leq n$. By (30) we obtain

$$\begin{aligned}&\mathbb{E}\big[\|\widetilde{I}_jH[\widetilde{P}(s)x(s)]\|_2^2\,|\,x(s)\big]\\ &\quad\geq cd_2\sum_{i,k}|H_{j'i}|^2x_k^2(s)=cd_2\|x(s)\|_2^2\sum_i|H_{j'i}|^2,\end{aligned}$$

which is followed by (28). By Theorem III.5 i) $x(s)$ cannot converge to a deterministic vector different from $x(0)$ in mean square. □

IV. Some Applications and Extension

A. Necessary and sufficient conditions for group consensus over random signed networks and with state-dependent noise

As we discuss in the Introduction, consensus problems in multi-agent systems have drawn a lot of attention from various fields including physics, biology, engineering and mathematics in the past two decades. Typically, a general assumption is adopted that the interaction matrix associated with the network is row-stochastic at every time. Recently, motivated by the possible antagonistic interaction in social networks, bipartite/group/cluster consensus problems have been studied over signed networks (focusing on continuous-time dynamic models), e.g., see [1], [21], [24], [33]. Interestingly, from

Theorems III.1, III.6 and III.7 we can obtain some results for the group consensus of the discrete-time system over random signed networks and with state-dependent noise.

Consider a discrete-time first-order system containing n agents. Each agent i has a state $x_i(s) \in \mathbb{R}$ at time s which can represent the opinion, social power or others, and is updated according to the current state and the interaction from the others. In detail, for $1 \leq i \leq n$ and $s \geq 0$, the state of agent i is updated by

$$\begin{aligned} x_i(s+1) &= (1-a(s))x_i(s) \\ &+ a(s) \sum_{j \in \mathcal{N}_i(s)} P_{ij}(s) \left[x_j(s) + f_{ji}(x(s)) w_{ji}(s) \right], \quad (31) \end{aligned}$$

where $a(s) \geq 0$ is the gain at time s, $\mathcal{N}_i(s)$ is the neighbors of node i at time s, $P_{ij}(s)$ is the weight of the edge (j,i) at time s, and $f_{ji}(x(s))w_{ji}(s)$ is the noise of agent i receiving information from agent j at time s. Here we consider the noise may be state-dependent which means that $f_{ji}(x(s))$ is a function of the state vector $x(s)$. Let $P_{ij}(s) = 0$ if $j \notin \mathcal{N}_i(s)$, and set

$$u_i(s) := \sum_{j \in \mathcal{N}_i(s)} P_{ij}(s) f_{ji}(x(s)) w_{ji}(s),$$

then system (31) can be rewritten as

$$x(s+1) = (1-a(s))x(s) + a(s) \left[P(s)x(s) + u(s) \right].$$

If $P_{ij}(s)$ is a stationary stochastic process with uniformly bounded variance, and $w_{ji}(s)$ is a zero-mean noise with uniformly bounded variance for any $x(s)$, $P_{ji}(s)$, and $j \in \mathcal{N}_i(s)$, then (A1) is satisfied with $u = \mathbf{0}_{n\times 1}$.

We say the subsets $S_1, \ldots, S_{r'} (r' \geq 1)$ is *a partition of* $\{1, \ldots, n\}$ if $\emptyset \subset S_i \subseteq \{1, \ldots, n\}$ for $1 \leq i \leq r'$, $S_i \cap S_j = \emptyset$ for $i \neq j$, and $\cup_{i=1}^{r'} S_i = \{1, \ldots, n\}$. Following [33] with some modifications we introduce the definition for group consensus:

Definition IV.1: Let the subsets $S_1, \ldots, S_{r'}$ be a partition of $\{1, \ldots, n\}$. If $x(s)$ is mean-square convergent, and $\lim_{s\to\infty} \mathbb{E}|x_i(s) - x_j(s)| = 0$ when i and j belong to a same subset, then we say $x(s)$ asymptotically reaches $\{S_i\}_{i=1}^{r'}$-group consensus in mean square.

From Definition IV.1 we can know that consensus is a special case of the $\{S_i\}_{i=1}^{r'}$-group consensus with $r' = 1$. Before the statement of our results, we need to introduce some notations and an assumption:

For a partition $S_1, \ldots, S_{r'}$ of $\{1, \ldots, n\}$, let $\mathbb{1}^i \in \mathbb{R}^n (1 \leq i \leq r')$ denote the column vector satisfying $\mathbb{1}^i_k = 1$ if $k \in S_i$ and $\mathbb{1}^i_k = 0$ otherwise. A linear combination of $\{\mathbb{1}^i\}_{i=1}^{r'}$ is $c_1 \mathbb{1}^1 + \ldots + c_{r'} \mathbb{1}^{r'}$ with $c_1, \ldots, c_{r'} \in \mathbb{C}$ being constants.

(A5) Assume any eigenvalue of P whose real part is 1 equals 1, and the algebraic and geometric multiplicities of the eigenvalue 1 equal $r \in [1, r']$, and any right eigenvector of P corresponding to the eigenvalue 1 can be written as a linear combination of $\{\mathbb{1}^i\}_{i=1}^{r'}$.

With Theorems III.1, III.6 and Proposition III.1 we obtain the following result:

Theorem IV.1: (Necessary and sufficient condition for group consensus with non-negative gains) Consider the system (2) or (31) satisfying (A1) with $u = \mathbf{0}_{n\times 1}$ and (A2). Let $S_1, \ldots, S_{r'}$ be a partition of $\{1, \ldots, n\}$. Then $x(s)$ asymptotically reaches $\{S_i\}_{i=1}^{r'}$-group consensus in mean square for any initial state if and only if $\widetilde{\rho}_{\max}(P) < 1$, or (A5) holds with $\widetilde{\rho}_{\max}(P) = 1$.

Proof. Before proving our result, we introduce some notes first. For any matrix $A \in \mathbb{C}^{n\times n}$, let A_i and A^i denote the i-th row and i-th column of A respectively. Set $A^{[i,j]} = (A^i, A^{i+1}, \ldots, A^j) \in \mathbb{C}^{n\times(j-i+1)}$.

We first consider the sufficient part. If $\widetilde{\rho}_{\max}(P) < 1$, by Proposition III.1 and the fact $u = \mathbf{0}_{n\times 1}$ we obtain that $x(s)$ converges to $\mathbf{0}_{n\times 1}$ in mean square for all initial states. Hence, the $\{S_i\}$-group consensus can be reached.

If (A5) holds with $\widetilde{\rho}_{\max}(P) = 1$, which implies that (A3) holds together with the fact $u = \mathbf{0}_{n\times 1}$. Let $P = H^{-1}DH$, where H is an invertible matrix, and D is the Jordan normal form of P with the same expression as (5). Then, by Theorem III.1, for any initial state there exist random variables $y_1, \ldots, y_r$ such that in mean square

$$x(s) \to y_1[H^{-1}]^1 + \cdots + y_r[H^{-1}]^r \quad \text{as} \quad s \to \infty. \qquad (32)$$

Also, from $PH^{-1} = DH^{-1}$ and (5) we have

$$P[H^{-1}]^i = [H^{-1}]^i, \quad 1 \leq i \leq r. \qquad (33)$$

Hence, by (32) and (A5), there exist random variables $z_1, \ldots, z_{r'}$ such that in mean square

$$x(s) \to z_1 \mathbb{1}^1 + \cdots + z_{r'} \mathbb{1}^{r'} \quad \text{as} \quad s \to \infty,$$

which implies that $x(s)$ asymptotically reaches $\{S_i\}_{i=1}^{r'}$-group consensus in mean square for any initial state.

Next we prove the necessary part. Since $x(s)$ asymptotically reaches $\{S_i\}_{i=1}^{r'}$-group consensus in mean square for any initial state, then, by Definition IV.1, $x(s)$ is mean-square convergent for any initial state. Hence, by Theorem III.6, we obtain that $\widetilde{\rho}_{\max}(P) < 1$, or (A3) holds with $\widetilde{\rho}_{\max}(P) = 1$.

It remains to show (A5) holds for the case when (A3) holds. For any complex eigenvector $\mathbf{a} + \mathbf{b}i \in \mathbb{C}^n$ of P corresponding to eigenvalue 1, we have $P\mathbf{a} = \mathbf{a}$ and $P\mathbf{b} = \mathbf{b}$, which implies that $\mathbf{a}$ and $\mathbf{b}$ are real eigenvectors of P corresponding to eigenvalue 1. Thus, any complex eigenvector of P corresponding to the eigenvalue 1 can be written as a linear combination of real eigenvectors. Also, from (5) we have $PH^{-1} = H^{-1}D$ if and only if (33) and $P[H^{-1}]^{[r+1,n]} = [H^{-1}]^{[r+1,n]}\underline{D}$ hold. Thus, we can choose suitable H such that $P = H^{-1}DH$ and $[H^{-1}]^1, \ldots, [H^{-1}]^r$ are real vectors. By Theorem III.1, we have

$$\lim_{s\to\infty} \mathbb{E}x(s) = \sum_{i=1}^{r} H_i x(0) \cdot [H^{-1}]^i \qquad (34)$$

Also, from $HH^{-1} = I_n$ we have $H_i[H^{-1}]^j$ equals 1 if $i = j$ and 0 otherwise. If we choose $x(0) = [H^{-1}]^i$ $(1 \leq i \leq r)$, by (34) we have $\lim_{s\to\infty} \mathbb{E}x(s) = [H^{-1}]^i$. Because for any initial state, $x(s)$ asymptotically reaches $\{S_i\}_{i=1}^{r'}$-group consensus in mean square, which implies $Ex(s)$ also asymptotically reaches $\{S_i\}_{i=1}^{r'}$-group consensus, $[H^{-1}]^i (1 \leq i \leq r)$ can be written as a linear combination of $\{\mathbb{1}^j\}_{j=1}^{r'}$. From the linear independence of $[H^{-1}]^1, \ldots, [H^{-1}]^r$ we have $r \le r'$, and $[H^{-1}]^1, \ldots, [H^{-1}]^r$ is a basis of the eigenspace $\mathcal{R}^1$ which consists of all the right eigenvectors of P corresponding to the eigenvalue 1 and together with the zero vector. Hence, any vector in $\mathcal{R}^1$ can be written as a linear combination of $[H^{-1}]^1, \ldots, [H^{-1}]^r$, and thus a linear combination of $\{\mathbb{1}^j\}_{j=1}^{r'}$. □

Similar to Theorem IV.1 we have the following theorem:

Theorem IV.2: (Necessary and sufficient condition for group consensus with non-positive gains) Consider the system (2) or (31) satisfying (A1) with $u = \mathbf{0}_{n\times 1}$ and (A2'). Let $S_1, \ldots, S_{r'}$ be a partition of $\{1, \ldots, n\}$. Then $x(s)$ asymptotically reaches $\{S_i\}_{i=1}^{r'}$-group consensus in mean square for any initial state if and only if $\widetilde{\rho}_{\min}(P) > 1$, or (A5) holds with $\widetilde{\rho}_{\min}(P) = 1$.

By Theorems IV.1 and IV.2 with $r' = 1$, we immediately obtain the following two corollaries for consensus:

Corollary IV.1: Consider the system (2) or (31) satisfying (A1) with $u = \mathbf{0}_{n\times 1}$ and (A2). Then $x(s)$ asymptotically reaches consensus in mean square for any initial state if and only if one of the following condition holds:
i) $\widetilde{\rho}_{\max}(P) < 1$;
ii) The sum of of each row of P equals 1, and P has $n-1$ eigenvalues whose real parts are all less than 1.

Corollary IV.2: Consider the system (2) or (31) satisfying (A1) with $u = \mathbf{0}_{n\times 1}$ and (A2'). Then $x(s)$ asymptotically reaches consensus in mean square for any initial state if and only if one of the following condition holds:
i) $\widetilde{\rho}_{\min}(P) > 1$;
ii) The sum of of each row of P equals 1, and P has $n-1$ eigenvalues whose real parts are all bigger than 1.

We give the following two examples for consensus and group consensus respectively over signed networks.

Example 1: By the Perron-Frobenius Theorem, if P is an irreducible row-stochastic matrix then P satisfies the condition ii) in Corollary IV.1, which implies that $x(s)$ asymptotically reaches mean-square consensus. However, the reverse is not always true. For example, let

$$P := \begin{bmatrix} 1.1 & -0.1 & 0 \\ 0.2 & 0.6 & 0.2 \\ -0.1 & 0.7 & 0.4 \end{bmatrix}.$$

The eigenvalues of P are $1, 0.9772, 0.1228$, which can guarantee consensus of $x(s)$ but P is not a row-stochastic matrix.

Example 2: Different from consensus, the group consensus does not require that the sum of each row of P equals 1. For example, if

$$P = \begin{bmatrix} 0.3 & 0.5 & 0.5 & -0.4 \\ 0.5 & 0.3 & -0.4 & 0.5 \\ -0.1 & 0.5 & 0.4 & 0.4 \\ 0.5 & -0.1 & 0.4 & 0.4 \end{bmatrix},$$

then $P[1,1,2,2]^\top = [1,1,2,2]^\top$, and the eigenvalues of P are $1, 0.6, -0.1+0.728i, -0.1-0.728i$. Let $S_1 = \{1,2\}$ and $S_2 = \{3,4\}$, by Theorem IV.1 $x(s)$ can asymptotically reach $\{S_1, S_2\}$-group consensus in mean square for any initial state.

B. An extension to multidimensional affine dynamics

Our results in Section III can be extended to multidimensional affine dynamics in which the state of each agent is a m-dimensional vector. The dynamics is

$$\begin{aligned} X(s+1) &= (1-a(s))X(s) \\ &\quad + a(s)\big[P(s)X(s)C^\top(s) + U(s)\big],\ s \ge 0, \end{aligned} \tag{35}$$

where $X(s) \in \mathbb{R}^{n\times m}$ is the state matrix, $P(s) \in \mathbb{R}^{n\times n}$ is still an interaction matrix, $C \in \mathbb{R}^{m\times m}$ is an interdependency matrix, and $U(s) \in \mathbb{R}^{n\times m}$ is an input vector.

The system (35) seems complex, however it can be transformed to one dimensional system (2) by the following way:

Given a pair of matrices $A \in \mathbb{R}^{n\times m}$, $B \in \mathbb{R}^{p\times q}$, their Kronecker product is defined by

$$A \otimes B = \begin{bmatrix} A_{11}B & \cdots & A_{1m}B \\ \cdots & \cdots & \cdots \\ A_{n1}B & \cdots & A_{nm}B \end{bmatrix} \in \mathbb{R}^{np\times mq}.$$

Let $Q(s) := P(s) \otimes C(s)$. From (35) we have

$$\begin{aligned} X_{ij}(s+1) &= (1-a(s))X_{ij}(s) \\ &\quad + a(s)\Big[\sum_{k_1,k_2} P_{ik_1}(s)X_{k_1,k_2}(s)C_{jk_2}(s) + U_{ij}(s)\Big]. \end{aligned} \tag{36}$$

for any $s \ge 0$, $1 \le i \le n$, and $1 \le j \le m$. Let

$$y(s) := (X_{11}(s), \ldots, X_{1m}(s), \ldots, X_{n1}(s), \ldots, X_{nm}(s))^\top$$

and

$$v(s) := (U_{11}(s), \ldots, U_{1m}(s), \ldots, U_{n1}(s), \ldots, U_{nm}(s))^\top$$

be the vector in $\mathbb{R}^{nm}$ transformed from the matrices $X(s)$ and $U(s)$ respectively. By (36) we have

$$\begin{aligned} y_{(i-1)m+j}(s+1) &= X_{ij}(s+1) \\ &= (1-a(s))X_{ij}(s) \\ &\quad + a(s)\Big[\sum_{k_1,k_2} P_{ik_1}(s)X_{k_1,k_2}(s)C_{jk_2}(s) + U_{ij}(s)\Big] \\ &= (1-a(s))y_{(i-1)m+j}(s) + \\ &\quad a(s)\Big[\sum_{k_1,k_2} Q_{(i-1)m+j,(k_1-1)m+k_2}(s)y_{(k_1-1)m+k_2}(s) \\ &\qquad + v_{(i-1)m+j}(s)\Big], \end{aligned}$$

which implies

$$y(s+1) = (1-a(s))y(s) + a(s)[Q(s)y(s) + v(s)].$$

The system (37) has the same form as the system (2), so the results in Section III can be applied to the multidimensional affine dynamics.

C. SA Friedkin-Johnsen model over time-varying interaction network

The Friedkin-Johnsen (FJ) model proposed by [11] considers a community of n social actors (or agents) whose opinion column vector is $x(s) = (x_1(s), \ldots, x_n(s))^\top \in \mathbb{R}^n$ at time s. The FJ model also contains a row-stochastic matrix of interpersonal influences $P \in \mathbb{R}^{n\times n}$ and a diagonal matrix of actors' susceptibilities to the social influence $\Lambda \in \mathbb{R}^{n\times n}$ with $\mathbf{0}_{n\times n} \le \Lambda \le I_n$. The state of the FJ model is updated by

$$x(s+1) = \Lambda P x(s) + (I_n - \Lambda)x(0), \quad s = 0, 1, \ldots. \tag{37}$$

By [23], if $\mathbf{0}_{n\times n} \le \Lambda < I_n$, then

$$\lim_{s\to\infty} x(s) = (I_n - \Lambda P)^{-1}(I_n - \Lambda)x(0). \tag{38}$$

However, if the interpersonal influences are affected by noise, then the system (37) may not converge.

The FJ model (37) was extended to the multidimensional case in [12], [23]. The multidimensional FJ model still contains n individuals, but each individuals has beliefs on m truth statements. Let $X(s) \in \mathbb{R}^{n\times m}$ be the matrix of n individuals' beliefs on m truth statements at time s. Following [12], it is updated by

$$X(s+1) = \Lambda P X(s) C^\top + (I_n - \Lambda)X(0) \tag{39}$$

for $s = 0, 1, \ldots,$ where $\Lambda, P \in \mathbb{R}^{n\times n}$ are the same matrices in (37), and $C \in \mathbb{R}^{m\times m}$ is a row-stochastic matrix of interdependencies among the m truth statements. The convergence of system (39) has been analyzed in [23]. Similar to (37) it is easy to see that if system (39) is affected by noise, then it will not converge. We will adopt the stochastic-approximation method to smooth the effects of the noise.

Proposition IV.1: Consider the system

$$X(s+1) = (1 - a(s))X(s) + a(s)[\Lambda(s)P(s)X(s)C(s)^\top + (I_n - \Lambda(s))X(0)], \tag{40}$$

for $s = 0, 1, \ldots$, where $\Lambda(s) \in \mathbb{R}^{n\times n}$, $P(s) \in \mathbb{R}^{n\times n}$ and $C(s) \in \mathbb{R}^{m\times m}$ are independent matrix sequence with invariant expectation Λ, P, and C respectively. Suppose P and C are row-stochastic matrix, and $\mathbf{0}_{n\times n} \le \Lambda < I_n$, and the gain function $a(s)$ satisfies (A2). Then for any initial state, $X(s)$ converges to X^* in mean square, where X^* is the unique solution of the equation

$$X = \Lambda P X C^\top + (I_n - \Lambda)X(0). \tag{41}$$

Proof. Since P and C are row-stochastic matrices, $P \otimes C$ is still a row-stochastic matrix. Together with the condition that $\mathbf{0}_{n\times n} \le \Lambda < I_n$, we have that the sum of each row of $(\Lambda P) \otimes C$ is less than 1. Thus, using the Gershgorin disk theorem we obtain $\widetilde{\rho}_{\max}((\Lambda P)\otimes C) < 1$. Let $Q := (\Lambda P)\otimes C$, $U(s) := (I_n - \Lambda(s))X(0)$,

$$y(s) := (X_{11}(s), \ldots, X_{1m}(s), \ldots, X_{n1}(s), \ldots, X_{nm}(s))^\top,$$

$$v(s) := (U_{11}(s), \ldots, U_{1m}(s), \ldots, U_{n1}(s), \ldots, U_{nm}(s))^\top,$$

and $v := Ev(s)$. By Proposition III.1 and the transformation from (35) to (37), we obtain that $y(s)$ converges to $(I_{mn} - Q)^{-1}v$ in mean square.

It remains to discuss the relation between $(I_{mn} - Q)^{-1}v$ and X^*. Let

$$y^* := (X^*_{11}, \ldots, X^*_{1m}, \ldots, X^*_{n1}, \ldots, X^*_{nm})^\top \in \mathbb{R}^{nm}.$$

By (41), similar to (37) we have $y^* = Qy^* + v$, which has a unique solution $y^* = (I_{mn} - Q)^{-1}v$ since $I_{mn} - Q$ is an invertible matrix by $\widetilde{\rho}_{\max}(Q) < 1$. Thus, with the fact that $y(s)$ converges to $(I_{mn} - Q)^{-1}v$ in mean square we obtain that $X(s)$ converges to X^* in mean square. □

Remark 4: According to Theorem III.1 and Proposition III.1, the conditions of $\Lambda(s), P(s)$ and $C(s)$ in Proposition IV.1 can be further relaxed for convergence, such as P and C are not row-stochastic matrices, and $\mathbf{0}_{n\times n} \le \Lambda < I_n$ may be extended to $\Lambda < \mathbf{0}_{n\times n}$ or $\Lambda \ge I_n$.

V. Conclusion

In this paper, we consider a time-varying affine dynamical system, where the state of the system features persistent oscillation and does not converge. We propose a stochastic approximation-based approach and obtain necessary and sufficient conditions to guarantee mean-square convergence. Our theoretical results largely extend the conditions on the spectrum of the expectation of the system matrix and thus can be applied in a much broader range of applications. We also derived the convergence rate of the system. To illustrate the theoretical results, we applied them in two different applications: group consensus in multi-agent systems and FJ model with time-varying interactions in social networks.

This work leaves various problems for future research. First, the system matrix and input are assumed to have constant expectations in this paper. However, it would be more interesting, yet challenging, to study systems with time-varying expectation of the system matrix and input. Second, we only considered basic affine systems in this paper. How and whether the proposed framework can be extended to nonlinear system are important and intriguing questions. Finally, we have illustrated our results in two different application scenarios; there are other possible applications such as gossip algorithms for consensus.

Appendix A

Lemma A.1: Suppose the non-negative real number sequence $\{y_s\}_{s\ge1}$ satisfies

$$y_{s+1} \le (1 - a_s)y_s + b_s, \tag{42}$$

where $b_s \ge 0$ and $a_s \in [0, 1)$ are real numbers. If $\sum_{s=1}^\infty a_s = \infty$ and $\lim_{s\to\infty} b_s/a_s = 0$, then $\lim_{s\to\infty} y_s = 0$ for any $y_1 \ge 0$.

Proof: Repeating (42) we obtain

$$y_{s+1} \le y(1)\prod_{t=1}^{s}(1 - a_t) + \sum_{i=1}^{s} b_i \prod_{t=i+1}^{s}(1 - a_t).$$

Here we define $\prod_{t=i}^{s}(\cdot) := 1$ when $i > s$. Since $\sum_{t=1}^{\infty} a_t = \infty$, we have $\prod_{t=1}^{\infty}(1-a_t)$. To obtain our result, we just need to prove that

$$\lim_{s\to\infty}\sum_{i=1}^{s} b_i \prod_{t=i+1}^{s}(1-a_t) = 0. \tag{43}$$

Since $\lim_{s\to\infty} b_s/a_s = 0$, for any real number $\varepsilon > 0$, there exists an integer $s^* > 0$ such that $b_s \le \varepsilon a_s$ when $s \ge s^*$. Thus,

$$\begin{aligned}&\sum_{i=1}^{s} b_i \prod_{t=i+1}^{s}(1-a_t) \qquad (44)\\ &\le \sum_{i=1}^{s^*-1} b_i \prod_{t=i+1}^{s}(1-a_t) + \sum_{i=s^*}^{s}\varepsilon a_i \prod_{t=i+1}^{s}(1-a_t)\\ &= \sum_{i=1}^{s^*-1} b_i \prod_{t=i+1}^{s}(1-a_t) + \varepsilon\Big(1-\prod_{t=s^*}^{s}(1-a_t)\Big)\\ &\to \varepsilon \;\text{ as }\; s\to\infty,\end{aligned}$$

where the first equality uses the equality (54). Let ε decrease to 0, then (44) is followed by (43). □

Appendix B
Proof of Theorem III.2

We prove this theorem under the following three cases:

Case I: $\widetilde{\rho}_{\max}(P) < 1$. Define $\theta(s)$, A and A_2 as in the proof of Theorem III.1 but with $r = 0$. Set $V(\theta) := \theta^* A\theta$ for any $\theta \in \mathbb{C}^n$, where θ^* denotes the conjugate transpose of θ. We remark that $A_2 = A \in \mathbb{C}^{n\times n}$ under the case $r = 0$, so that, by (18), we have

$$\begin{aligned}&\mathbb{E}[V(\theta(s+1))] \qquad (45)\\ &\le \Big(1-\frac{\alpha}{\rho(A)(s+\beta)^\gamma}\Big)\mathbb{E}[V(\theta(s))] + O\Big(\frac{1}{(s+\beta)^{2\gamma}}\Big).\end{aligned}$$

Set

$$\Phi(s,i) := \prod_{k=i}^{s}\Big(1-\frac{\alpha}{\rho(A)(k+\beta)^\gamma}\Big)$$

and define $\prod_{k=i}^{s}(\cdot) := 1$ if $s < i$. We compute

$$\begin{aligned}\Phi(s,i) &= O\Big(\exp\Big[\sum_{k=i}^{s} -\frac{\alpha}{\rho(A)(k+\beta)^\gamma}\Big]\Big) \qquad (46)\\ &= O\Big(\exp\Big(\int_i^s -\frac{\alpha}{\rho(A)(k+\beta)^\gamma}dk\Big)\Big)\\ &= \begin{cases} O\Big(\big(\frac{s+\beta}{i+\beta}\big)^{-\alpha/\rho(A)}\Big), \text{ if } \gamma = 1,\\ O\Big(\exp\big(\frac{-\alpha}{(1-\gamma)\rho(A)}[(s+\beta)^{1-\gamma}-(i+\beta)^{1-\gamma}]\big)\Big),\\ \qquad \text{if } \frac{1}{2}<\gamma<1.\end{cases}\end{aligned}$$

Also, using (45) repeatedly we obtain

$$\begin{aligned}&\mathbb{E}[V(\theta(s+1))] \qquad (47)\\ &\le \Phi(s,0)\mathbb{E}[V(\theta(0))] + \sum_{i=0}^{s}\Phi(s,i+1)O\Big(\frac{1}{(i+\beta)^{2\gamma}}\Big).\end{aligned}$$

Assume $\alpha \ge \rho(A)$. We first consider the case that $\gamma = 1$. From (46) and (47) we have

$$\begin{aligned}&\mathbb{E}[V(\theta(s+1))] \qquad (48)\\ &= o\Big(\frac{1}{s}\Big) + O\Big(\sum_{i=0}^{s}\frac{(s+\beta)^{-\alpha/\rho(A)}}{(i+\beta)^{2-\frac{\alpha}{\rho(A)}}}\Big) = O\Big(\frac{1}{s}\Big).\end{aligned}$$

For the case when $\gamma \in (\frac{1}{2}, 1)$, we take $b = \frac{\alpha}{(1-\gamma)\rho(A)}$, and from (46) and (47) we can obtain

$$\begin{aligned}&\mathbb{E}[V(\theta(s+1))]\\ &= e^{-b(s+\beta)^{1-\gamma}}\cdot O\Big(1+\sum_{i=0}^{s}\frac{e^{b(i+\beta)^{1-\gamma}}}{(i+\beta)^{2\gamma}}\Big)\\ &= e^{-b(s+\beta)^{1-\gamma}}\cdot O\Big(\sum_{i=0}^{s}\sum_{k=0}^{\infty}\frac{b^k(i+\beta)^{(1-\gamma)k-2\gamma}}{k!}\Big)\\ &= e^{-b(s+\beta)^{1-\gamma}}\cdot O\Big(\sum_{k=0}^{\infty}\frac{b^k}{k!}\sum_{i=0}^{s}(i+\beta)^{(1-\gamma)k-2\gamma}\Big)\\ &= e^{-b(s+\beta)^{1-\gamma}}\cdot O\Big(\sum_{k=0}^{\infty}\frac{b^k(s+\beta)^{(1-\gamma)k-2\gamma+1}}{k![(1-\gamma)k-2\gamma+1]}\Big)\\ &= \frac{e^{-b(s+\beta)^{1-\gamma}}}{(s+\beta)^\gamma}\cdot O\Big(\sum_{k=0}^{\infty}\frac{b^{k+1}(s+\beta)^{(1-\gamma)(k+1)}}{(k+1)!}\Big)\\ &= O(s^{-\gamma}). \qquad (49)\end{aligned}$$

By (48) and (49), we have $\mathbb{E}[V(\theta(s))] = O(s^{-\gamma})$ for $\frac{1}{2} < \gamma \le 1$. Combing this with the definition of $\theta(s)$ yields our result.

Case II: $\widetilde{\rho}_{\max}(P) = 1$. Let $\theta(s)$, $\underline{\theta}(s)$, $\bar{\theta}(s)$, $\bar{\theta}(\infty)$, H, y and z be the same variables as in the proof of Theorem III.1. With (18) and following the similar process from (45) to (49), we have $\mathbb{E}\|\underline{\theta}(s)\|_2^2 = O(s^{-\gamma})$. Also, from (21) we have

$$\begin{aligned}\mathbb{E}\|\bar{\theta}(\infty)-\bar{\theta}(s)\|_2^2 &= O\Big(\sum_{k=s}^{\infty}a^2(k)\Big) = O\Big(\sum_{k=s}^{\infty}a^2(k)\Big)\\ &= O\Big(\sum_{k=s}^{\infty}\frac{1}{(s+\beta)^{2\gamma}}\Big) = O\Big(\frac{1}{s^{2\gamma-1}}\Big).\end{aligned}$$

Since $x(s) = H^{-1}[\theta(s)+z]$ and $H^{-1}y$ is a mean square limit of $x(s)$, the arguments above imply

$$\begin{aligned}\mathbb{E}\|x(s)-H^{-1}y\|_2^2 &= \max\big\{O(s^{-\gamma}), O(s^{1-2\gamma})\big\}\\ &= O(s^{1-2\gamma}).\end{aligned}$$

Case III: $\widetilde{\rho}_{\min}(P) \ge 1$. The protocol (2) is written as

$$x(s+1) = x(s) + \frac{\alpha}{(s+\beta)^\gamma}[(I_n - P(s))x(s) - u(s)].$$

Because $\widetilde{\rho}_{\max}(I_n - P) \le 0$, arguments similar to that for Cases I) and II) yield our result.

Appendix C
Proof of Theorem III.3

i) As same as Subsection III-B, the Jordan normal form of H is

$$D = \mathrm{diag}(J_1,\dots,J_k) = HPH^{-1}.$$

We also set $y(s) := Hx(s)$, $v(s) := Hu(s)$, $D(s) := HP(s)H^{-1}$, $D = \mathbb{E}D(s) = HPH^{-1}$, and $v = Ev(s) = Hu$. By (7) and (A1) we have

$$\begin{aligned}\mathbb{E}y(s+1) &= \mathbb{E}\big[\mathbb{E}[y(s+1)\,|\,y(s)]\big]\\ &= \mathbb{E}y(s) + a(s)[(D-I_n)Ey(s)+v]. \end{aligned} \tag{50}$$

Let $B(s) := I_n + a(s)(D - I_n)$. Using (50) repeatedly we obtain

$$\begin{aligned}&\mathbb{E}[y(s+1)] \\ &\quad = B(s)\cdots B(0)y(0) + \sum_{t=0}^{s} a(t)B(s)\cdots B(t+1)v.\end{aligned} \tag{51}$$

We will continue the proof under the following two cases:
Case I: $\widetilde{\rho}_{\max}(P) > 1$. Without loss of generality we assume $\mathrm{Re}(\lambda_1(P)) > 1$. Let J_1 be a Jordan block in D corresponding to $\lambda_1(P)$. Let m_1 be the row index of D corresponding to the last line of J_1, i.e.,

$$D_{m_1} = (0, \dots, 0, \lambda_1(P), 0, \dots, 0). \tag{52}$$

Then by (51)

$$\begin{aligned}&\mathbb{E}[y_{m_1}(s+1)]\\ &\quad = y_{m_1}(0)\prod_{t=0}^{s}[1-a(i)[1-\lambda_1(P)]] + \frac{v_{m_1}}{1-\lambda_1(P)}\\ &\qquad \times \sum_{t=0}^{s} a(t)[1-\lambda_1(P)] \prod_{k=t+1}^{s}(1-a(k)[1-\lambda_1(P)])\\ &\quad = y_{m_1}(0)\prod_{t=0}^{s}(1-a(t)[1-\lambda_1(P)]) + \frac{v_{m_1}}{1-\lambda_1(P)}\\ &\qquad \times \Big(1 - \prod_{t=0}^{s}(1-a(t)[1-\lambda_1(P)])\Big),\end{aligned} \tag{53}$$

where the last equality uses the classical equality

$$\sum_{t=0}^{s} c_t \prod_{k=t+1}^{s}(1-c_k) = 1 - \prod_{t=0}^{s}(1-c_t) \tag{54}$$

with $\{c_t\}$ being any complex numbers, which can be obtained by induction. Here we define $\prod_{k=s_1}^{s_2}(\cdot) = 1$ if $s_2 < s_1$. Since $\sum_s a(s) = \infty$,

$$\begin{aligned}&\prod_{t=0}^{\infty}|1-a(t)[1-\lambda_1(P)]|^2\\ &\qquad\qquad \geq \prod_{t=0}^{\infty}\{1 + 2a(t)[\mathrm{Re}(\lambda_1(P)) - 1]\} = \infty.\end{aligned}$$

Hence, from (53), if $y_{m_1}(0) \neq \frac{v_{m_1}}{1-\lambda_1(P)}$, then

$$\lim_{s\to\infty}|\mathbb{E}[y_{m_1}(s)]| = \infty, \tag{55}$$

which implies $\lim_{s\to\infty}\mathbb{E}\|x(s)\|_2^2 = \infty$.
Case II: $\widetilde{\rho}_{\max}(P) = 1$. Under this case we consider the following three situations:
(a) There is an eigenvalue $\lambda_j(P) = 1 + \mathrm{Im}(\lambda_j(P))i$ with $\mathrm{Im}(\lambda_j(P)) \neq 0$, where $\mathrm{Im}(\lambda_j(P))$ denotes the imaginary part of $\lambda_j(P)$. Similar to (52), we can choose a row $D_{j'}$ of D which is equal to $(0, \dots, 0, \lambda_j(P), 0, \dots, 0)$. Similar to (53), we have

$$\begin{aligned}\mathbb{E}[y_{j'}(s+1)] &= y_{j'}(0)\prod_{t=0}^{s}[1-a(t)[1-\lambda_j(P)]]\\ &\quad + \frac{v_{j'}}{1-\lambda_j(P)}\cdot\Big(1-\prod_{t=0}^{s}[1-a(t)[1-\lambda_j(P)]]\Big).\end{aligned} \tag{56}$$

We write

$$\begin{aligned}1-a(t)[1-\lambda_j(P)] &= 1 + a(t)\mathrm{Im}(\lambda_j(P))i\\ &= r_t e^{i\varphi_t} = r_t(\cos\varphi_t + i\sin\varphi_t),\end{aligned}$$

where $r_t = \sqrt{1 + a^2(t)\mathrm{Im}^2(\lambda_j(P))}$ and

$$\begin{aligned}\varphi_t &= \arctan[a(t)\mathrm{Im}(\lambda_j(P))]\\ &= a(t)\mathrm{Im}(\lambda_j(P)) + \sum_{k=1}^{\infty}\frac{(-1)^k}{2k+1}[a(t)\mathrm{Im}(\lambda_j(P))]^{2k+1},\end{aligned} \tag{57}$$

so

$$\prod_{t=0}^{s}(1-a(t)[1-\lambda_j(P)]) = \exp\Big(i\sum_{t=0}^{s}\varphi_t\Big)\prod_{t=0}^{s}r_t. \tag{58}$$

Assume $y_{j'}(0) \neq \frac{v_{j'}}{1-\lambda_j(P)}$. Since $\sum_{t=0}^{\infty}a(t) = \infty$, equations (56), (58), and (57) imply

$$\overline{\lim}_{s\to\infty}\overline{\lim}_{s_2\to\infty}|\mathbb{E}[y_{j'}(s_2) - y_{j'}(s)]| > 0. \tag{59}$$

Next we consider the convergence of $x(s)$. Because $x(s) = H^{-1}y(s)$, using Jensen's inequality we have

$$\begin{aligned}\mathbb{E}\|x(s_2)-x(s)\|_2^2 &= \mathbb{E}\|H^{-1}[y(s_2)-y(s)]\|_2^2\\ &\geq \sigma_n^2(H^{-1})\mathbb{E}\|y(s_2)-y(s)\|_2^2\\ &\geq \sigma_n^2(H^{-1})\mathbb{E}|y_{j'}(s_2)-y_{j'}(s)|^2\\ &\geq \sigma_n^2(H^{-1})|\mathbb{E}[y_{j'}(s_2)-y_{j'}(s)]|^2,\end{aligned} \tag{60}$$

where $\sigma_n(H^{-1}) = \inf_{\|x\|_2=1}\|H^{-1}x\|_2$ denotes the least singular value of H^{-1}. Because H^{-1} is invertible, we have $\sigma_n(H^{-1}) > 0$. Hence, by (59) and (60), we obtain

$$\overline{\lim}_{s\to\infty}\overline{\lim}_{s_2\to\infty}\mathbb{E}\|x(s_2)-x(s)\|_2^2 > 0.$$

By the Cauchy criterion (see [16, page 58]), $x(s)$ is not mean square convergent.
(b) The geometric multiplicity of the eigenvalue 1 is less than its algebraic multiplicity. By (a), we only need to consider the case when any eigenvalue of P with 1 as real part has zero imaginary part. Thus, the Jordan normal form D contains a Jordan block

$$J_j = \begin{bmatrix}1 & 1 & & \\ & 1 & \ddots & \\ & & \ddots & 1\\ & & & 1\end{bmatrix}_{m_j\times m_j}$$

with $m_j \geq 2$. Let j' be the row index of D corresponding to the second line from the bottom of J_j. It can be computed that

$$[B(s)\cdots B(t)]_{j',j'+1} = \sum_{k=t}^{s}a(k).$$

Since $\sum_{k=0}^{\infty} a(k) = \infty$, from (51), there are some initial states such that $\lim_{s\to\infty} |\mathbb{E}[y_{j'}(s)]| = \infty$, which is followed by $\lim_{s\to\infty} \mathbb{E}\|x(s)\|_2^2 = \infty$.
(c) There is a left eigenvector ξ (row vector) of P corresponding to the eigenvalue 1 such that $\xi u \neq 0$. By (2) and (A1) we have

$$\begin{aligned}\xi\mathbb{E}x(s+1) &= (1-a(s))\xi\mathbb{E}x(s) + a(s)[\xi P\mathbb{E}x(s) + \xi u]\\ &= \xi\mathbb{E}x(s) + a(s)\xi u\\ &= \cdots = \xi x(0) + \sum_{k=0}^{s} a(k)\xi u,\end{aligned}$$

which implies $\lim_{s\to\infty} \mathbb{E}\|x(s)\|_2^2 = \infty$ by $\sum_{k=0}^{\infty} a(k) = \infty$.
ii) It can be obtained by the similar method as i).

Appendix D
Proof of Theorem III.4

We prove our result by contradiction: Suppose that there exists a real number sequence $\{a(s)\}_{s\geq 0}$ independent with $\{x(s)\}$ such that

$$\lim_{s\to\infty} \mathbb{E}\big\|x(s)-b\big\|_2^2 = 0. \tag{61}$$

We assert that $\lim_{s\to\infty} a(s) = 0$. This assertion will be proved still by contradiction: Assume that there exists a subsequence $\{a(s_k)\}_{k\geq 0}$ which does not converge to zero. Let $\widetilde{P}(s) := P(s) - P$ and $\widetilde{u}(s) = u(s) - u$ for any $s \geq 0$, then by (2), (A1) and (27) we have

$$\begin{aligned}&\mathbb{E}\Big[\big\|x(s_k+1)-b\big\|_2^2 \,|\, x(s_k)\Big]\\ &\quad= \mathbb{E}\Big[\big\|\xi + a(s_k)\big(\widetilde{P}(s_k)x(s_k) + \widetilde{u}(s_k)\big)\big\|_2^2 \,|\, x(s_k)\Big]\\ &\quad= \|\xi\|_2^2 + a^2(s_k)\mathbb{E}\Big[\big\|\widetilde{P}(s_k)x(s_k) + \widetilde{u}(s_k)\big\|_2^2 \,|\, x(s_k)\Big]\\ &\quad\geq a^2(s_k)c_3,\end{aligned} \tag{62}$$

where

$$\xi := (1-a(s_k))x(s_k) + a(s_k)(Px(s_k)+u) - b.$$

From (62) we know that $\mathbb{E}\|x(s_k+1)-b\|_2^2$ will not converge to 0 as k grows to infinity, which is in contradiction with (61).

Since $x(0)\neq b$, to guarantee the convergence of $x(s)$, the gain function $\{a(s)\}_{s\geq 0}$ must at least contain one non-zero element. Also, from (62), we can obtain that the number of the non-zero elements in the sequence $\{a(s)\}_{s\geq 0}$ must be infinite. Thus, together with the assertion of $\lim_{s\to\infty} a(s) = 0$, there exists an integer $s^* > 0$ such that $a(s^*-1)\neq 0$, $\{a(i)\}_{i=0}^{s^*-2}$ contains non-zero element, and

$$2|a(s)(1-\mathrm{Re}(\lambda_j(P))| < 1,\ \forall s\geq s^*, 1\leq j\leq n. \tag{63}$$

Let $A(s) := (1-a(s))I_n + a(s)P(s)$. By (2) we have

$$x(s+1) = A(s)x(s) + a(s)u(s), \quad s\geq s^*.$$

By (A1), we obtain

$$\begin{aligned}&\mathbb{E}[x(s+1)\,|\,x(s^*)] - (I_n-P)^{-1}u\\ &\quad= \big[I_n - a(s)(I_n-P)\big]\mathbb{E}[x(s)\,|\,x(s^*)]\\ &\qquad+ a(s)u - (I_n-P)^{-1}u\\ &\quad= \big[I_n - a(s)(I_n-P)\big]\big(\mathbb{E}[x(s)\,|\,x(s^*)] - (I_n-P)^{-1}u\big)\\ &\quad= \cdots = \Big(\prod_{k=s^*}^{s}\mathbb{E}[A(k)]\Big)\big(x(s^*) - (I_n-P)^{-1}u\big),\end{aligned}$$

which implies

$$\begin{aligned}&\mathbb{E}[x(s+1)|x(s^*)]\\ &\quad= H^{-1}\Big(\prod_{i=s^*}^{s}[I_n - a(i)(I_n-D)]\Big)\\ &\qquad\times H\big(x(s^*) - (I_n-P)^{-1}u\big) + (I_n-P)^{-1}u\end{aligned} \tag{64}$$

from (4). Set

$$\begin{aligned}z(s) :=& \Big(\prod_{i=s^*}^{s}[I_n - a(i)(I_n-D)]\Big)H\\ &\cdot\big(x(s^*) - (I_n-P)^{-1}u\big) + H(I_n-P)^{-1}u - Hb.\end{aligned} \tag{65}$$

Using Jensen's inequality and (64) we have

$$\begin{aligned}&\mathbb{E}\big[\|x(s+1)-b\|_2^2\,|\,x(s^*)\big]\\ &\quad\geq \big\|\mathbb{E}\big[(x(s+1)-b)\,|\,x(s^*)\big]\big\|_2^2\\ &\quad= \big\|\mathbb{E}[(x(s+1)\,|\,x(s^*)] - b\big\|_2^2\\ &\quad= \|H^{-1}z(s)\|_2^2 \geq \sigma_n^2(H^{-1})\|z(s)\|_2^2,\end{aligned} \tag{66}$$

where $\sigma_n(H^{-1}) = \inf_{\|x\|_2=1}\|H^{-1}x\|_2$ denotes the least singular value of H^{-1}. Because H^{-1} is invertible, $\sigma_n(H^{-1}) > 0$. Define

$$w_j(s) := \prod_{i=s^*}^{s}(1 - a(i)[1-\lambda_j(P)]) \tag{67}$$

and

$$M_j := \prod_{i=s^*}^{\infty}[I_{m_j} - a(i)(I_{m_j} - J_j)]. \tag{68}$$

We can compute that

$$\begin{aligned}|w_j(s)|^2 &= \prod_{i=s^*}^{s}|1 - a(i)[1-\lambda_j(P)]|^2\\ &= \prod_{i=s^*}^{s}\big\{1 - 2a(i)[1-\mathrm{Re}(\lambda_j(P))]\\ &\qquad + a^2(i)[1 - 2\mathrm{Re}(\lambda_j(P)) + |\lambda_j(P)|^2\big\}.\end{aligned}$$

From this and (63) we have $w_j(s)\neq 0$ for any finite s. Also, if $w_j(\infty) = 0$, then $[1-\mathrm{Re}(\lambda_j(P))]\sum_{i=s^*}^{\infty} a(i) = \infty$. Hence, by (A4) or (A4'), there exists a Jordan block J_{j_1} associated with the eigenvalue $\lambda_{j_1'}(P)$ such that $w_{j_1'}(\infty)\neq 0$ and (26) holds. Because M_{j_1} is an upper triangular matrix whose diagonal elements are all $w_{j_1'}(\infty)\neq 0$, we can obtain the least singular value

$$\sigma_{m_{j_1}}(M_{j_1}) > 0.$$

Also, by (66) and (25), we obtain

$$\begin{aligned}\mathbb{E}\|x(\infty)-b\|_2^2 &= \mathbb{E}\big[\mathbb{E}[\|x(\infty)-b\|_2^2\,|\,x(s^*)]\big]\\ &\geq \sigma_n^2(H^{-1})\mathbb{E}\|z(\infty)\|_2^2\\ &\geq \sigma_n^2(H^{-1})\mathbb{E}\|z(\infty)-Ez(\infty)\|_2^2\\ &\geq \sigma_n^2(H^{-1})\mathbb{E}\big\|\widetilde{I}_{j_1}[z(\infty)-Ez(\infty)]\big\|_2^2\\ &= \sigma_n^2(H^{-1})\mathbb{E}\Big\|\widetilde{I}_{j_1}\Big(\prod_{i=s^*}^{\infty}[I_n-a(i)(I_n-D)]\Big)\\ &\quad\times H\big(x(s^*)-Ex(s^*)\big)\Big\|_2^2\\ &= \sigma_n^2(H^{-1})\mathbb{E}\big\|M_{j_1}\widetilde{I}_{j_1}H\big(x(s^*)-Ex(s^*)\big)\big\|_2^2\\ &\geq \sigma_n^2(H^{-1})\sigma_{m_{j_1}}^2(M_{j_1})\\ &\quad\times \mathbb{E}\big\|\widetilde{I}_{j_1}H\big(x(s^*)-Ex(s^*)\big)\big\|_2^2. \end{aligned}\tag{69}$$

Using (2) and (26) we have

$$\begin{aligned}&\mathbb{E}\left\{\big\|\widetilde{I}_{j_1}H(x(s^*)-\mathbb{E}x(s^*))\big\|_2^2\,|\,x(s^*-1)\right\}\\ &= a^2(s^*-1)\mathbb{E}\Big\{\big\|\widetilde{I}_{j_1}H\widetilde{P}(s^*-1)x(s^*-1)\\ &\quad+\widetilde{I}_{j_1}H\widetilde{u}(s^*-1)\big\|_2^2\,|\,x(s^*-1)\Big\}\\ &\geq a^2(s^*-1)\left(c_1\|x(s^*-1)\|_2^2+c_2\right). \end{aligned}\tag{70}$$

Because c_1 and c_2 cannot be zero at the same time, we consider the case when $c_2>0$ first. With the fact that $a(s^*-1)\neq 0$ and (70) we obtain

$$\begin{aligned}&\mathbb{E}\big\|\widetilde{I}_{j_1}H(x(s^*)-\mathbb{E}x(s^*))\big\|_2^2\\ &= \mathbb{E}\left\{\mathbb{E}\big\|\widetilde{I}_{j_1}H(x(s^*)-\mathbb{E}x(s^*))\big\|_2^2\,|\,x(s^*-1)\right\}>0.\end{aligned}$$

Substituting this into (69) yields $\mathbb{E}\|x(\infty)-b\|_2^2>0$, which is contradictory with (61).

For the case when $c_1>0$, by (69) and (70), we have

$$\begin{aligned}&\mathbb{E}\|x(\infty)-b\|_2^2\\ &\geq \sigma_n^2(H^{-1})\sigma_{m_{j_1}}^2(M_{j_1})a^2(s^*-1)\\ &\quad\cdot\mathbb{E}\left(\big\|\widetilde{I}_{j_1}H\widetilde{P}(0)x(s^*-1)\big\|_2^2\,|\,x(s^*-1)\right)\\ &\geq \sigma_n^2(H^{-1})\sigma_{m_{j_1}}^2(M_{j_1})a^2(s^*-1)c_1\mathbb{E}\|x(s^*-1)\|_2^2. \end{aligned}\tag{71}$$

Because $\{a(i)\}_{i=0}^{s^*-2}$ contains non-zero elements, we set s' to be the biggest number such that $s'\leq s^*-2$ and $a(s')\neq 0$. By (62) we have

$$\begin{aligned}\mathbb{E}\big\|x(s^*-1)\big\|_2^2 &= \mathbb{E}\big\|x(s'+1)\big\|_2^2\\ &\geq a^2(s')EE\big[\big\|\widetilde{P}(s')x(s')+\widetilde{u}(s')\big\|_2^2\,|\,x(s')\big]\\ &\geq a^2(s')c_3>0.\end{aligned}$$

Substituting this into (71) we get $\mathbb{E}\|x(\infty)-b\|_2^2>0$, which is contradictory with (61).

Appendix E
Proof of Theorem III.5

Similar to the proof of Theorem III.4 we prove our result by contradiction: Suppose that there exists a real number sequence $\{a(s)\}_{s\geq 0}$ independent with $\{x(s)\}$ such that (61) holds. Since $x(0)\neq b$, by (61) $\{a(s)\}_{s\geq 0}$ must contain non-zero elements. We consider the following three cases respectively to deduce the contradiction:

Case I: The condition i) is satisfied. Similar to the proof of Theorem III.4, we first prove $\lim_{s\to\infty}a(s)=0$ by contradiction: Suppose there exists a subsequence $\{a(s_k)\}$ that does not converge to zero. For the case when $b\neq \mathbf{0}_{n\times 1}$, by (61), there exists a time $s_1\geq 0$ such that

$$\mathbb{E}\big\|x(s)-b\big\|_2^2\leq \frac{1}{4}\|b\|_2,\quad \forall s>s_1. \tag{72}$$

Because for any $x(s_k)$,

$$\begin{aligned}\|b\|_2^2 &\leq \big(\|x(s_k)\|_2+\big\|b-x(s_k)\big\|_2\big)^2\\ &\leq 2\big(\|x(s_k)\|_2^2+\big\|b-x(s_k)\big\|_2^2\big),\end{aligned}$$

(72) is followed by

$$\mathbb{E}\|x(s_k)\|_2^2\geq \frac{1}{2}\|b\|_2-\mathbb{E}\big\|x(s_k)-b\big\|_2^2\geq \frac{1}{4}\|b\|_2 \tag{73}$$

for large k. By (62), (28) and (73) we obtain

$$\begin{aligned}\mathbb{E}\big\|x(s_k+1)-b\big\|_2^2 &\geq a^2(s_k)\mathbb{E}\|\widetilde{P}(s_k)x(s_k)\|_2^2\\ &= a^2(s_k)\mathbb{E}\|H^{-1}H\widetilde{P}(s_k)x(s_k)\|_2^2\\ &\geq a^2(s_k)\sigma_n^2(H^{-1})\mathbb{E}\|H\widetilde{P}(s_k)x(s_k)\|_2^2\\ &\geq a^2(s_k)\sigma_n^2(H^{-1})c_1\mathbb{E}\|x(s_k)\|_2^2\\ &\geq \frac{1}{4}a^2(s_k)\sigma_n^2(H)c_1\|b\|_2, \end{aligned}\tag{74}$$

which is contradictory with (61).

For the case when $b=\mathbf{0}_{n\times 1}$, by (62) and (74), we have

$$\begin{aligned}&\mathbb{E}\big\|x(s_k+1)\big\|_2^2\\ &\geq \mathbb{E}\|(1-a(s_k))x(s_k)+a(s_k)(Px(s_k)+u)\|_2^2\\ &\quad+a^2(s_k)\sigma_n^2(H^{-1})c_1\mathbb{E}\|x(s_k)\|_2^2. \end{aligned}\tag{75}$$

If $\|(1-a(s_k))I_n+a(s_k)P\|_2\mathbb{E}\|x(s_k)\|_2>\frac{1}{2}\|a(s_k)u\|_2$, by (75) and Jensen's inequality we have

$$\begin{aligned}&\mathbb{E}\big\|x(s_k+1)\big\|_2^2\geq a^2(s_k)\sigma_n^2(H^{-1})c_1(\mathbb{E}\|x(s_k)\|_2)^2\\ &\geq \frac{a^4(s_k)\sigma_n^2(H^{-1})c_1\|u\|^2}{4\|(1-a(s_k))I_n+a(s_k)P\|_2^2}\\ &\nrightarrow 0 \text{ if } a(s_k)\nrightarrow 0. \end{aligned}\tag{76}$$

Otherwise,

$$\begin{aligned}&\mathbb{E}\|(1-a(s_k))x(s_k)+a(s_k)(Px(s_k)+u)\|_2\\ &\geq \|a(s_k)u\|_2-\mathbb{E}\|(1-a(s_k))x(s_k)+a(s_k)Px(s_k)\|_2\\ &\geq \|a(s_k)u\|_2-\mathbb{E}\|(1-a(s_k))I_n+a(s_k)P\|_2\|x(s_k)\|_2\\ &\geq \frac{1}{2}\|a(s_k)u\|_2,\end{aligned}$$

Hence, using (75) and Jensen's inequality again, we obtain

$$\begin{aligned}&\mathbb{E}\big\|x(s_k+1)\big\|_2^2\\ &\geq \big(\mathbb{E}\|(1-a(s_k))x(s_k)+a(s_k)(Px(s_k)+u)\|_2\big)^2\\ &\geq \|a(s_k)u\|_2^2/4. \end{aligned}\tag{77}$$

Combining (76) and (77) yields $\mathbb{E}\big\|x(s_k+1)\big\|_2^2$. This quantity does not converge to zero, which is in contradiction with (61). By summarizing the arguments above we prove the assertion of $\lim_{s\to\infty}a(s)=0$.

Because $\lim_{s\to\infty} a(s) = 0$ and because $\{a(s)\}_{s\geq 0}$ contains non-zero elements, there exists an integer $s^* > 0$ such that $a(s^*-1) \neq 0$ and (63) holds. Define $w_j(s)$ and M_j by (67) and (68) respectively. With the arguments similar to the proof of Theorem III.4, we can find a Jordan block J_{j_1} associated with the eigenvalue $\lambda_{j_1'}(P)$ such that $w_{j_1'}(\infty) \neq 0$ and (28) holds. Similar to (71) we obtain

$$\begin{aligned}&\mathbb{E}\|x(\infty)-b\|_2^2\\ &\quad \geq \sigma_n^2(H)\sigma_{m_{j_1}}^2(M_{j_1})a^2(s^*-1)c_1\mathbb{E}\|x(s^*-1)\|_2^2. \quad (78)\end{aligned}$$

By (75) we have that if $\mathbb{E}\|x(s)\|_2^2 > 0$, then $\mathbb{E}\|x(s+1)\|_2^2 > 0$ for any $a(s) \in \mathbb{R}$. Then with the condition $x(0) \neq \mathbf{0}_{n\times 1}$, we have $\mathbb{E}\|x(s^*-1)\|_2^2 > 0$. Using this and (78) we get $\mathbb{E}\|x(\infty)-b\|_2^2 > 0$, which is contradictory with (61).

Case II: The condition ii) is satisfied. Since $\{a(s)\}_{s\geq 0}$ contains non-zero elements, we define s_1 to be the first s such that $a(s) \neq 0$. Then $x(s_1+1) = a(s_1)u \neq b$ almost surely. Let s_1+1 be the initial time and by the same arguments as in Case I we obtain $\mathbb{E}\|x(\infty)-b\|_2^2 > 0$.

Case III: The condition iii) is satisfied. If $x(0) = \mathbf{0}_{n\times 1}$, we obtain $\mathbb{E}\|x(s)\|_2^2 = 0$ for any $s \geq 0$, which is contradictory with (61). Thus, we just need to consider the case when $x(0) \neq \mathbf{0}_{n\times 1}$. Since $\{a(s)\}_{s\geq 0}$ contains non-zero elements, we define s_1 to be the first s such that $a(s) \neq 0$.

Set $x^* := x_1 + 1$. Define $w_j(s)$ and M_j by (67) and (68) respectively. If $\lambda_j(P)$ is not a real number, then $w_j(s)$ cannot be equal to 0 for any finite s. By the similar arguments as in the proof of Theorem III.4, there exists a Jordan block J_{j_1} associated with the eigenvalue $\lambda_{j_1'}(P)$ such that $w_{j_1'}(\infty) \neq 0$ and (28) holds. By (78) we have

$$\begin{aligned}&\mathbb{E}\|x(\infty)-b\|_2^2\\ &\quad \geq \sigma_n^2(H^{-1})\sigma_{m_{j_1}}^2(M_{j_1})a^2(s^*-1)c\mathbb{E}\|x(s^*-1)\|_2^2\\ &\quad = \sigma_n^2(H^{-1})\sigma_{m_{j_1}}^2(M_{j_1})a^2(s_1)c\|x(0)\|_2^2 > 0,\end{aligned}$$

which is contradictory with (61).

References

[1] C. Altafini. Consensus problems on networks with antagonistic interactions. *IEEE Transactions on Automatic Control*, 58(4):935–946, 2013. doi:10.1109/TAC.2012.2224251.

[2] V. S. Borkar and S. P. Meyn. The O.D.E. method for convergence of stochastic approximation and reinforcement learning. *SIAM Journal on Control and Optimization*, 38(2):447–469, 2000. doi:10.1137/S0363012997331639.

[3] S. Chatterjee and E. Seneta. Towards consensus: Some convergence theorems on repeated averaging. *Journal of Applied Probability*, 14(1):89–97, 1977. doi:10.2307/3213262.

[4] G. Chen, Z. Liu, and L. Guo. The smallest possible interaction radius for synchronization of self-propelled particles. *SIAM Review*, 56(3):499–521, 2014. doi:10.1137/140961249.

[5] G. Chen, L. Y. Wang, C. Chen, and G. Yin. Critical connectivity and fastest convergence rates of distributed consensus with switching topologies and additive noises. *IEEE Transactions on Automatic Control*, 2017. to appear. doi:10.1109/TAC.2017.2696824.

[6] R. Cogburn. The ergodic theory of Markov chains in random environments. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 66(1):109–128, 1984. doi:10.1007/BF00532799.

[7] M. H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974. doi:10.1080/01621459.1974.10480137.

[8] F. Fagnani and S. Zampieri. Randomized consensus algorithms over large scale networks. *IEEE Journal on Selected Areas in Communications*, 26(4):634–649, 2008. doi:10.1109/JSAC.2008.080506.

[9] J. A. Fax and R. M. Murray. Information flow and cooperative control of vehicle formations. *IEEE Transactions on Automatic Control*, 49(9):1465–1476, 2004. doi:10.1109/TAC.2004.834433.

[10] P. Frasca, H. Ishii, C. Ravazzi, and R. Tempo. Distributed randomized algorithms for opinion formation, centrality computation and power systems estimation: A tutorial overview. *European Journal of Control*, 24:2–13, 2015. doi:10.1016/j.ejcon.2015.04.002.

[11] N. E. Friedkin and E. C. Johnsen. Social influence networks and opinion change. In S. R. Thye, E. J. Lawler, M. W. Macy, and H. A. Walker, editors, *Advances in Group Processes*, volume 16, pages 1–29. Emerald Group Publishing Limited, 1999.

[12] N. E. Friedkin, A. V. Proskurnikov, R. Tempo, and S. E. Parsegov. Network science on belief system dynamics under logic constraints. *Science*, 354(6310):321–326, 2016. doi:10.1126/science.aag2624.

[13] Y. Han, W. Lu, and T. Chen. Cluster consensus in discrete-time networks of multi-agents with inter-cluster nonidentical inputs. *IEEE Transactions on Neural Networks and Learning Systems*, 24(4):566–578, 2013. doi:10.1109/TNNLS.2013.2237786.

[14] J. Hofbauer and W. H. Sandholm. On the global convergence of stochastic fictitious play. *Econometrica*, 70(6):2265–2294, 2002. doi:10.1111/j.1468-0262.2002.00440.x.

[15] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1994.

[16] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Dover Publications, 2007.

[17] P. Jia, A. MirTabatabaei, N. E. Friedkin, and F. Bullo. Opinion dynamics and the evolution of social power in influence networks. *SIAM Review*, 57(3):367–397, 2015. doi:10.1137/130913250.

[18] U. A. Khan, S. Kar, and J. M. F. Moura. Distributed sensor localization in random environments using minimal number of anchor nodes. *IEEE Transactions on Signal Processing*, 57(5):2000–2016, 2009. doi:10.1109/TSP.2009.2014812.

[19] H. J. Kushner and G. G. Yin. *Stochastic Approximation and Recursive Algorithms and Applications*. Springer, 1997. doi:10.1007/b97441.

[20] T. Li and J. F. Zhang. Consensus conditions of multi-agent systems with time-varying topologies and stochastic communication noises. *IEEE Transactions on Automatic Control*, 55(9):2043–2057, 2010. doi:10.1109/TAC.2010.2042982.

[21] Z. Meng, G. Shi, K. H. Johansson, M. Cao, and Y. Hong. Behaviors of networks with antagonistic interactions and switching topologies. *Automatica*, 73:110–116, 2016. doi:10.1016/j.automatica.2016.06.022.

[22] L. Moreau. Stability of multiagent systems with time-dependent communication links. *IEEE Transactions on Automatic Control*, 50(2):169–182, 2005. doi:10.1109/TAC.2004.841888.

[23] S. E. Parsegov, A. V. Proskurnikov, R. Tempo, and N. E. Friedkin. Novel multidimensional models of opinion dynamics in social networks. *IEEE Transactions on Automatic Control*, 62(5):2270–2285, 2017. doi:10.1109/TAC.2016.2613905.

[24] J. Qin and C. Yu. Cluster consensus control of generic linear multi-agent systems under directed topology with acyclic partition. *Automatica*, 49(9):2898–2905, 2013. doi:10.1016/j.automatica.2013.06.017.

[25] C. Ravazzi, P. Frasca, R. Tempo, and H. Ishii. Ergodic randomized algorithms and dynamics over networks. *IEEE Transactions on Control of Network Systems*, 2(1):78–87, 2015. doi:10.1109/TCNS.2014.2367571.

[26] W. Ren and R. W. Beard. *Distributed Consensus in Multi-vehicle Cooperative Control*. Communications and Control Engineering. Springer, 2008.

[27] C. W. Reynolds. Flocks, herds, and schools: A distributed behavioral model. *Computer Graphics*, 21(4):25–34, 1987. doi:10.1145/37402.37406.

[28] H. Robbins and S. Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3):400–407, 1951. URL: http://www.jstor.org/stable/2236626.

[29] A. Tahbaz-Salehi and A. Jadbabaie. A necessary and sufficient condition for consensus over random networks. *IEEE Transactions on Automatic Control*, 53(3):791–795, 2008. doi:10.1109/TAC.2008.917743.

[30] J. N. Tsitsiklis. Asynchronous stochastic approximation and Q-learning. *Machine Learning*, 16(3):185–202, 1994. doi:10.1023/A:1022689125041.

[31] J. N. Tsitsiklis, D. P. Bertsekas, and M. Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Transactions on Automatic Control*, 31(9):803–812, 1986. doi:10.1109/TAC.1986.1104412.

[32] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet. Novel type of phase transition in a system of self-driven particles. *Physical Review Letters*, 75(6-7):1226–1229, 1995. doi:10.1103/PhysRevLett.75.1226.

[33] J. Yu and L. Wang. Group consensus in multi-agent systems with switching topologies and communication delays. *Systems & Control Letters*, 59(6):340–348, 2010. doi:10.1016/j.sysconle.2010.03.009.

[34] W. X. Zhao, H. F. Chen, and H. T. Fang. Convergence of distributed randomized PageRank algorithms. *IEEE Transactions on Automatic Control*, 58(12):3255–3259, 2013. doi:10.1109/TAC.2013.2264553.