



HAL
open science

Diffusion-Based Distributed Parameter Estimation Through Directed Graphs With Switching Topology: Application of Dynamic Regressor Extension and Mixing

Alexey S Matveev, Mostafa Almodarresi, Romeo Ortega, Anton Pyrkin, Siyu
Xie

► **To cite this version:**

Alexey S Matveev, Mostafa Almodarresi, Romeo Ortega, Anton Pyrkin, Siyu Xie. Diffusion-Based Distributed Parameter Estimation Through Directed Graphs With Switching Topology: Application of Dynamic Regressor Extension and Mixing. IEEE Transactions on Automatic Control, 2022, 67 (8), pp.4256-4263. 10.1109/TAC.2021.3115075 . hal-04177823

HAL Id: hal-04177823

<https://hal.science/hal-04177823v1>

Submitted on 6 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Diffusion-based Distributed Parameter Estimation Through Directed Graphs with Switching Topology: Application of Dynamic Regressor Extension and Mixing

Alexey S. Matveev, Mostafa Almodarresi, Romeo Ortega, Life Fellow, IEEE, Anton Pyrkin, Senior Member, IEEE, and Siyu Xie

Abstract—In this paper, we consider the problem of discrete-time, diffusion-based distributed parameter estimation with the agents connected via *directed* graphs with *switching* topologies and a self loop at each node. We show that, by incorporating the recently introduced dynamic regressor extension and mixing procedure to a classical gradient-descent algorithm, improved convergence properties can be achieved. In particular, it is shown that with this modification sufficient conditions for global convergence of *all* the estimators is that *one of the sensors* receives enough information to generate a consistent estimate and that this sensor is “well-connected”. The main feature of this result is that the excitation condition imposed on this distinguished sensor is strictly weaker than the classical persistent excitation requirement. The connectivity assumption is also very mild, requiring only that the union of the edges of all connectivity graphs over any time interval with an arbitrary but fixed length contains a spanning tree rooted at the information-rich node. In the case of non-switching topologies, this assumption is satisfied by strongly connected graphs, and not only by them.

Index Terms—Diffusion strategies, sensor networks, distributed adaptive filters, dynamic regressor extension and mixing, consensus algorithms.

I. INTRODUCTION

THE problem of cooperative estimation of an unknown parameter vector in sensor networks has a wide area of applications including signal processing and distributed control, see *e.g.*, [4, 10, 11] for a recent survey of the field. It is usually assumed that each sensor possesses only partial information of the unknown parameter process, which is generated via N linear regression equations (LRE), where N is the number of sensors. If the partial information available in some sensors is not sufficient to consistently estimate the parameters, the idea is to share information among the sensors to achieve this objective. In most practical scenarios, it is not possible to have centralized information, therefore, we have to consider some kind of distributed exchange of information, which is represented by a weighted graph. Another aspect to be considered in the problem is the distributed processing of the information that, as discussed in [10, 11], are classified into incremental, consensus and diffusion strategies. Given its

Alexey S. Matveev is with the Department of Mathematics and Mechanics, Saint Petersburg State University, St. Petersburg 198504, Russia. Email: almat1712@yahoo.com.

Mostafa Almodarresi is with the Department of Electrical Engineering, Yazd University, Yazd 89195-741, Iran. Email: almodarresi8@gmail.com.

Romeo Ortega is with Departamento Académico de Sistemas Digitales, ITAM, C.P. 01080, Ciudad de México, México. Email: romeo.ortega@itam.mx.

Anton Pyrkin is with the Department of Control Systems and Informatics, ITMO University, Saint Petersburg 197101, Russia. Email: a.pyrkin@gmail.com.

Siyu Xie is with the Department of Electrical and Computer Engineering, Wayne State University, Detroit, Michigan 48202, USA. Email: syxie@wayne.edu

proven advantages—stemming from its more efficient propagation of information—we consider in this paper the latter strategy, also adopted in [3, 10, 11].

In diffusion parameter estimation strategies the sensors exchange estimates with their neighbors and fuse the collected estimates via linearly combining them. Depending on the order of adaptation and combination, there are two different types of diffusion strategies: combine-then-adapt (CTA) or adapt-then-combine (ATC). In this paper, we consider both the CTA and ATC strategies. The effect of the measurement noise on both these strategies, within a rigorous stochastic analysis framework, has been extensively studied in the literature and we refer the interested reader to [10] for a recent review of the literature. In this paper, similarly to [3], we adopt a *purely deterministic* framework that leads to the, somehow simpler, stability analysis of a linear time-varying (LTV) system.

The main modification we introduce to the classical gradient (also called least-mean squares) scheme is the incorporation of the dynamic regressor extension and mixing (DREM) estimation procedure, recently introduced in [1]. The successful application of DREM to various theoretical and practical engineering problems is reported in the tutorial paper [7]. The main feature of DREM that we exploit in this paper is that it transforms the q -dimensional LRE, associated with a q -dimensional unknown vector, into q *scalar* LRE, one for each one of the unknown parameters. This feature turns out to have a major impact on the stability properties of the LTV systems that describe the behavior of the parameter errors.

The main contributions of our paper are the following.

C1 For both CTA and ATC strategies global convergence of all the DREM-based estimators is ensured *if one of the sensors*—which should be “well connected”—receives enough information to generate a consistent estimate.

C2 The excitation condition imposed on this distinguished sensor is strictly weaker than the classical persistent excitation (PE) requirement [10] or the relaxed “cooperative PE” condition introduced in [3].

C3 In contrast with most existing results, we handle *directed* graphs with *switching topologies*, that are required to contain a jointly spanning tree rooted at the information-rich node with a self loop at each node. For *constant* graphs the latter connectivity requirement is satisfied by *strongly connected* graphs—a condition often assumed in the literature [3, 10].

Although we consider a noise-free scenario, it should be underscored that it has been shown in [8] that DREM is Input-to-State Stable (ISS), hence robust to additive input noise. The remainder of the paper is organized as follows.

The problem formulation and standing assumptions are given in Section II. Section III briefly reviews the DREM-based distributed estimator. The main result is presented in Section IV. Simulation results, which illustrate the performance of the proposed estimator, even in the presence of noise, are presented in Section V. To enhance readability, the proof of the main result is given in Appendix A.

II. PROBLEM FORMULATION AND DESCRIPTION OF THE NETWORK TOPOLOGY

In this section we present the discrete-time, distributed estimation problem that is studied in the paper, and describe the information exchange pattern between the agents.

A. Distributed cooperative estimation problem

We consider the problem of discrete-time parameter estimation using a network consisting of N sensors distributed over a geographic region, which communicate with one another to obtain a consistent estimate of a certain parameter. We assume that at time $t \in \mathbb{N}_+$ sensor¹ $i \in \bar{N} := \{1, \dots, N\}$ has access to the measurements $y^i(t) \in \mathbb{R}$ and $\phi^i(t) \in \mathbb{R}^q$, and that they are related via the following LRE

$$y^i(t) = \theta^\top \phi^i(t), \quad t \in \mathbb{N}_+. \quad (1)$$

Here $\theta \in \mathbb{R}^q$ is an *unknown* constant parameter vector that should be estimated at the site of every sensor. We are specially interested in the situation where the measurements of a particular sensor i are insufficient for reconstructing θ in full; an example arises where all components of the vector $\phi^i(t) = \{\phi_j^i(t)\}_{j=1}^q \in \mathbb{R}^q$ are constantly zero $\phi_j^i(t) = 0$ for $j = 1, \dots, q'$ with a certain $q' \geq 1$. Then, this sensor i gets a chance to build a proper estimate of the entire θ only thanks to information received from the peers. Success of this endeavor is not given for granted and depends on a number of factors. They are partly concerned with both information exchange in the network, *e.g.*, the communication topology, and informativeness of the measurements performed by various sensors. The focus of the paper is on exactly these issues, studied within the framework of distributed estimation. In the face of this primary goal and paper length limit, we omit a somewhat classical research issue of effects from sensor noises on the estimate, and consider a “noise-free” scenario.

Specifically, we assume that each sensor receives information from some other sensors. How the sensors should cooperate in individually generating a consistent estimate, and under which conditions on the measurements this may happen?

B. Description of the communication pattern

The interconnection structure of the network is conventionally encoded by a *time-varying, directed, and weighted graph* [2]. It is defined as a triple $\mathcal{G}(t) = [\mathcal{V}, \mathcal{E}(t), \mathcal{A}(t)]$, where $\mathcal{V} = \{1, 2, \dots, N\}$ is the set of nodes, each associated with a sensor, $\mathcal{E}(t)$ the set of edges, where $(i, j) \in \mathcal{E}(t)$ if and only if

node j affects node i at time t , and $\mathcal{A}(t) = \{a_{ij}(t)\} \in \mathbb{R}^{N \times N}$ is the weighted adjacency matrix. Its entries are such that

$$0 \leq a_{ij}(t) \leq 1; \quad \sum_{j=1}^N a_{ij}(t) = 1, \quad \forall i, j \in \bar{N}, \quad (2)$$

and $a_{ij}(t) > 0 \Leftrightarrow (i, j) \in \mathcal{E}(t)$. Along with thereby reflecting the topology of the network, the *weights* $a_{ij}(t)$ indicate the strengths of interaction among various nodes. The *set of neighbors* of sensor i is defined as

$$\mathcal{N}_i(t) = \{j \in \mathcal{V} \mid (i, j) \in \mathcal{E}(t)\}.$$

In contrast with most of works on distributed estimation, we consider time-varying graphs and, moreover, *directed* graphs. Hence the matrix $\mathcal{A}(t)$ is not necessarily symmetric. Our assumptions about the graph are fully exposed in Section IV.

III. PROPOSED DREM-BASED DISTRIBUTED ESTIMATOR

In this paper, we adopt the diffusion estimation strategy, where the sensors share estimates with their neighbors via communication and fuse the collected data via linear combination. The estimate of θ built by sensor i at time t is denoted by $\hat{\theta}^i(t)$. Such an estimate is typically generated with a gradient algorithm. The main novelty of the paper stems from using, instead, a *DREM-based distributed estimator* [1, 7].

A. The underlying DREM estimator

To make the paper self-contained, we now briefly recall the construction of the DREM estimator for the isolated i -th sensor and highlight the main convergence property of this estimator.

To streamline the presentation, we introduce the following notations:

$$\begin{aligned} \Phi^i(t) &:= \begin{bmatrix} (\phi^i(t))^\top \\ (\phi^i(t-1))^\top \\ \vdots \\ (\phi^i(t-q+1))^\top \end{bmatrix} \in \mathbb{R}^{q \times q}, \\ Y^i(t) &= \begin{bmatrix} Y_1^i(t) \\ Y_2^i(t) \\ \vdots \\ Y_q^i(t) \end{bmatrix} := \text{adj}\{\Phi^i(t)\} \begin{bmatrix} y^i(t) \\ y^i(t-1) \\ \vdots \\ y^i(t-q+1) \end{bmatrix}, \\ \delta^i(t) &:= \det\{\Phi^i(t)\}, \end{aligned} \quad (3)$$

where we denote by $\text{adj}\{\cdot\}$ the adjugate matrix. Let θ_ℓ stand for the ℓ -th entry of θ . (The index ℓ always takes values in the set $\bar{q} := \{1, \dots, q\}$. This clarification is basically omitted in the sequel, unless it is not clear from the context.) The gradient estimate of this entry is generated via

$$\hat{\theta}_\ell^i(t+1) = \hat{\theta}_\ell^i(t) + \frac{\delta^i(t)}{\mu_i + (\delta^i(t))^2} \left(Y_\ell^i(t) - \delta^i(t) \hat{\theta}_\ell^i(t) \right), \quad (4)$$

where $\mu_i > 0$ is a parameter of the estimator. Given an initial estimate $\hat{\theta}_\ell^i(0)$, the subsequent estimates $\hat{\theta}_\ell^i(t), t \geq 1$ are uniquely determined via (4).

Proposition III.1 Consider the LRE (1) and the signals (3).

¹The index i takes always values in the set \bar{N} . Therefore, whenever clear from the context, this clarification is omitted in the sequel.

(i) For each entry θ_ℓ of θ and at any time t , we have

$$Y_\ell^i(t) = \delta^i(t)\theta_\ell. \quad (5)$$

(ii) The gradient estimate of the ℓ -th entry verifies

$$\lim_{t \rightarrow \infty} \hat{\theta}_\ell^i(t) = \theta_\ell, \quad \forall \hat{\theta}_\ell^i(0) \Leftrightarrow \delta^i(t) \notin \ell_2, \quad (6)$$

where ℓ_2 is the space of square summable sequences.

Proof: The proof of (i) follows immediately by noting that

$$\begin{bmatrix} y^i(t) \\ y^i(t-1) \\ \vdots \\ y^i(t-q+1) \end{bmatrix} = \Phi^i(t)\theta,$$

multiplying this equation by $\text{adj}\{\Phi^i(t)\}$, and recalling that for any (possibly singular) $q \times q$ matrix M , we have [5]:

$$\text{adj}\{M\}M = \det\{M\}I_q.$$

Claim (ii) is established by noting that the estimation errors $\tilde{\theta}_\ell^i(t) := \hat{\theta}_\ell^i(t) - \theta_\ell$ satisfy the recursive equation

$$\tilde{\theta}_\ell^i(t+1) = \frac{\mu_i}{\mu_i + (\delta^i(t))^2} \tilde{\theta}_\ell^i(t),$$

and observing that

$$\prod_{j=0}^{\infty} \left[\frac{\mu_i}{\mu_i + (\delta^i(j))^2} \right] = 0 \Leftrightarrow \delta^i(t) \notin \ell_2. \quad \square\square\square$$

B. Diffusion strategies

As indicated in the introduction, see also [10], depending on the order of adaptation and combination, there are two different types of diffusion strategies, CTA or ATC. Now we describe these strategies for the proposed DREM estimator.

Algorithm 1 CTA diffusion DREM algorithm

For any sensor i and parameter element θ_ℓ , pick an initial estimate $\hat{\theta}_\ell^{i,CTA}(0)$ and then do the following for $t = 0, 1, \dots$

1: Combine local estimates:

$$\hat{\beta}_\ell^{i,CTA}(t) = \sum_{j \in \mathcal{N}_i(t)} a_{ij}(t) \hat{\theta}_\ell^{j,CTA}(t).$$

2: Adapt the local estimate:

$$\begin{aligned} \hat{\theta}_\ell^{i,CTA}(t+1) &= \hat{\beta}_\ell^{i,CTA}(t) \\ &+ \frac{\delta^i(t)}{\mu_i + (\delta^i(t))^2} \left(Y_\ell^i(t) - \delta^i(t) \hat{\theta}_\ell^{i,CTA}(t) \right). \end{aligned}$$

C. Derivation of the error equations

To simplify the error equations, we introduce the following notations:

Algorithm 2 ATC diffusion DREM algorithm

For any sensor i and parameter element θ_ℓ , pick an initial estimate $\hat{\theta}_\ell^{i,ATC}(0)$ and then do the following for $t = 0, 1, \dots$

1: Adapt the local estimate:

$$\begin{aligned} \hat{\beta}_\ell^{i,ATC}(t+1) &= \hat{\theta}_\ell^{i,ATC}(t) \\ &+ \frac{\delta^i(t)}{\mu_i + (\delta^i(t))^2} \left(Y_\ell^i(t) - \delta^i(t) \hat{\theta}_\ell^{i,ATC}(t) \right). \end{aligned}$$

2: Combine local estimates:

$$\hat{\theta}_\ell^{i,ATC}(t+1) = \sum_{j \in \mathcal{N}_i(t)} a_{ij}(t) \hat{\beta}_\ell^{j,ATC}(t+1).$$

$$\hat{\Theta}_\ell^{CTA}(t) := \text{col} \left(\hat{\theta}_\ell^{1,CTA}(t), \dots, \hat{\theta}_\ell^{N,CTA}(t) \right) \in \mathbb{R}^N,$$

$$\Theta_\ell := \text{col}(1, \dots, 1)\theta_\ell \in \mathbb{R}^N, \quad (7)$$

$$\mathbf{X}_\ell(t) := \hat{\Theta}_\ell^{CTA}(t) - \Theta_\ell, \quad (8)$$

$$\hat{\Theta}_\ell^{ATC}(t) := \text{col} \left(\hat{\theta}_\ell^{1,ATC}(t), \dots, \hat{\theta}_\ell^{N,ATC}(t) \right) \in \mathbb{R}^N,$$

$$\mathbf{Z}_\ell(t) := \hat{\Theta}_\ell^{ATC}(t) - \Theta_\ell, \quad (9)$$

$$B(t) := \text{diag} \left\{ \frac{1}{1 + \frac{(\delta^1(t))^2}{\mu_1}}, \dots, \frac{1}{1 + \frac{(\delta^N(t))^2}{\mu_N}} \right\} \in \mathbb{R}^{N \times N}. \quad (10)$$

We are in a position to present the error equations.

Lemma 1 For any entry θ_ℓ of the estimated parameter θ , the error vectors (8) and (9) evolve as follows

$$\mathbf{X}_\ell(t+1) = B(t)\mathcal{A}(t)\mathbf{X}_\ell(t), \quad (11)$$

$$\mathbf{Z}_\ell(t+1) = \mathcal{A}(t)B(t)\mathbf{Z}_\ell(t), \quad (12)$$

where $\mathcal{A}(t)$ is the weighted adjacency matrix of the graph.

Proof: First of all, we put

$$\mathbf{Y}_\ell(t) := \text{col}(Y_\ell^1(t), \dots, Y_\ell^N(t)),$$

$$\Delta(t) := \text{diag}\{\delta^1(t), \dots, \delta^N(t)\},$$

$$\hat{\mathbf{B}}_\ell^{CTA}(t) := \text{col}(\hat{\beta}_\ell^{1,CTA}(t), \dots, \hat{\beta}_\ell^{N,CTA}(t)),$$

$$\hat{\mathbf{B}}_\ell^{ATC}(t) := \text{col}(\hat{\beta}_\ell^{1,ATC}(t), \dots, \hat{\beta}_\ell^{N,ATC}(t)),$$

$$\mathbf{L}(t) := \text{diag} \left\{ \frac{\delta^1(t)}{\mu_1 + (\delta^1(t))^2}, \dots, \frac{\delta^N(t)}{\mu_N + (\delta^N(t))^2} \right\}.$$

Based on (5) and (7), we see that

$$\mathbf{Y}_\ell(t) = \Delta(t)\Theta_\ell. \quad (13)$$

For the CTA diffusion DREM algorithm, we have

$$\hat{\mathbf{B}}_\ell^{CTA}(t) = \mathcal{A}(t)\hat{\Theta}_\ell^{CTA}(t)$$

$$\hat{\Theta}_\ell^{CTA}(t+1) = \hat{\mathbf{B}}_\ell^{CTA}(t) + \mathbf{L}(t) \left(\mathbf{Y}_\ell(t) - \Delta(t)\hat{\mathbf{B}}_\ell^{CTA}(t) \right).$$

Using (8) and substituting (13) into the last equation, we get

$$\begin{aligned} \mathbf{X}_\ell(t+1) &= \mathcal{A}(t)\hat{\Theta}_\ell^{CTA}(t) - \Theta_\ell \\ &+ \mathbf{L}(t) \left(\Delta(t)\Theta_\ell - \Delta(t)\mathcal{A}(t)\hat{\Theta}_\ell^{CTA}(t) \right). \quad (14) \end{aligned}$$

Since $\mathcal{A}(t)\Theta_\ell = \Theta_\ell$ due to (2) and (7), we see that

$$\begin{aligned} \mathbf{X}_\ell(t+1) &= \left(I_N - \mathbf{L}(t)\Delta(t)\right)\mathcal{A}(t)\mathbf{X}_\ell(t) \\ &= B(t)\mathcal{A}(t)\mathbf{X}_\ell(t), \end{aligned}$$

where $B(t)$ is defined in (10).

For the ATC diffusion DREM algorithm, we have

$$\begin{aligned} \widehat{\mathbf{B}}_\ell^{ATC}(t+1) &= \widehat{\Theta}_\ell^{ATC}(t) + \mathbf{L}(t) \left(\mathbf{Y}_\ell(t) - \Delta(t)\widehat{\Theta}_\ell^{ATC}(t)\right) \\ \widehat{\Theta}_\ell^{ATC}(t+1) &= \mathcal{A}(t)\widehat{\mathbf{B}}_\ell^{ATC}(t+1). \end{aligned} \quad (15)$$

By invoking (9), substituting (13) into (15) and using $\mathcal{A}(t)\Theta_\ell = \Theta_\ell$ once more, we arrive at

$$\mathbf{Z}_\ell(t+1) = \mathcal{A}(t) \left(I_N - \mathbf{L}(t)\Delta(t)\right) \mathbf{Z}_\ell(t) = \mathcal{A}(t)B(t)\mathbf{Z}_\ell(t),$$

which completes the proof. $\square\square\square$

IV. MAIN RESULT

Our main result provides conditions for global asymptotic stability of the LTV systems (11) and (12). They describe the evolution of the estimation errors that arise when using the CTA and ATC DREM-based diffusion strategies, respectively. Interestingly, the conditions are the same for both strategies. Specifically, there are five conditions. The first three of them are very mild, even self-evident in the present context: namely, that the signals appearing in the LREs (1) are bounded, that the positive elements of the adjacency matrix remain bounded away from zero and that all agents use their own information to generate their estimate. These assumptions are grouped together into one and are formally articulated as follows.

- Assumption 1**
- i) In (3), $\delta^i(t)$ are bounded over t .
 - ii) There is $\underline{\alpha} \in (0, 1)$ such that $a_{ij}(t) > 0 \Rightarrow a_{ij}(t) \geq \underline{\alpha}$.
 - iii) Each node contains a *self loop* at any time.

Here i) holds true if all the regressors $\phi^i(t)$ are bounded. For time-invariant graphs, ii) is always true. In general, ii) means that the strength of any existing influence does not degenerate to zero as time progresses.

The remaining two conditions pertain, respectively, to the information content of the regressor vectors $\phi^i(t)$ and the scope for the dissemination of this information among agents via the network. The former condition is consonant to excitation assumptions that are typically imposed on the signals when treating parameter estimation and state observation tasks. With respect to DREM and according to Proposition III.1, now such an assumption is concerned with the non-square summability of the determinant of the extended regressor matrix $\Phi^i(t)$.² In the networked scenario examined in this paper, convergence of *all* estimators holds even if only *one* of the agents verifies this condition, provided that this agent is “suitably connected” to all the others. The last is fleshed out by our fifth condition. To state it, we need the following notation: For any discrete interval $\mathcal{I} = [t_0 : t_1] := \{t = t_0, \dots, t_1 - 1\}$, we denote by $\mathcal{G}(\mathcal{I})$ the graph with nodes

²In [7, Proposition 3] it is shown that this assumption on the lack of square summability is *strictly weaker* than the classical PE requirement of gradient estimators.

$i \in \bar{N}$ that results from the union of the edges of all graphs $\mathcal{G}(t_0), \mathcal{G}(t_0 + 1), \dots, \mathcal{G}(t_1 - 1)$.

Assumption 2

- i) There exists a node $r \in \bar{N}$ such that the sequence $\delta^r(t) \notin \ell_2$.

- ii) There is a natural number p such that the graph $\mathcal{G}[[tp, (t+1)p]]$ contains a spanning tree rooted at r for any $t \in \mathbb{N}_+$.

For example, ii) holds *with any* $r \in \bar{N}$ if there exists p such that the graph $\mathcal{G}[[tp, (t+1)p]]$ is strongly connected for any $t \in \mathbb{N}_+$.

We are in position to present the main result of the paper.

Theorem 1 Consider a network of N sensors, each verifying the LRE (1), whose information exchange is characterized by the graph $\mathcal{G}(t)$. Suppose that Assumptions 1 and 2 hold. Starting from any initial conditions, the parameter estimation errors of the CTA and the ATC DREM-based diffusion algorithms converge to zero in *all nodes*, that is,

$$\lim_{t \rightarrow \infty} \mathbf{X}_\ell(t) = 0 \text{ and } \lim_{t \rightarrow \infty} \mathbf{Z}_\ell(t) = 0, \forall \ell \in \bar{q}.$$

V. SIMULATION RESULTS

In this section, we illustrate the performance of the proposed distributed estimation scheme with simulation examples. We consider a network of four sensors with a time-varying interaction graph $\mathcal{G}(t)$ that switches among three topologies shown in Fig. 1. The adjacency matrices of the graphs are set as

$$\begin{aligned} \mathcal{A}_1 &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1/3 & 2/3 & 0 & 0 \\ 0 & 1/6 & 5/6 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \mathcal{A}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1/3 & 2/3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1/6 & 5/6 \end{bmatrix} \\ \mathcal{A}_3 &= \begin{bmatrix} 2/3 & 0 & 0 & 1/3 \\ 0 & 1 & 0 & 0 \\ 0 & 1/6 & 5/6 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

The pattern of switchings among the topologies is given by

$$\mathcal{A}(t) = \begin{cases} \mathcal{A}_1 & t = 0, 1, 2, 3 \\ \mathcal{A}_2 & t = 4, 5, 6, 7 \\ \mathcal{A}_3 & t = 8, 9, 10, 11 \end{cases}, \quad \mathcal{A}(t+12) = \mathcal{A}(t).$$

As can be easily seen, then the union of the graphs over every time interval with a length at least 12 is directed and strongly connected. Assume that a constant parameter vector $\theta \in \mathbb{R}^2$ should be cooperatively estimated. Consider the following regression vectors

$$\phi^1(t) = \begin{bmatrix} a(t) \\ 1 \end{bmatrix}, \phi^2(t) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \phi^3(t) = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \phi^4(t) = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

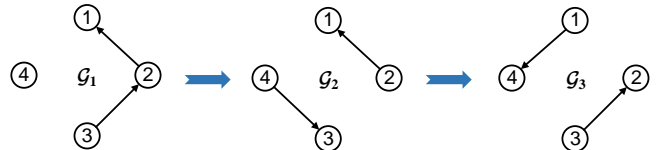


Fig. 1: The set of switched topologies. The graph retains each topology for 4 time steps and then moves to the next topology.

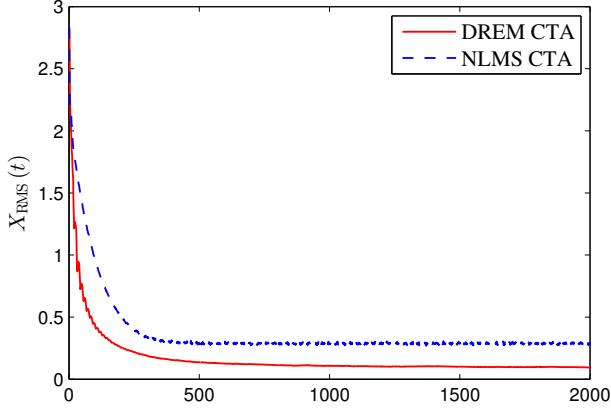


Fig. 2: Root mean square of the overall parameter estimation error.

where $a(t)$ is generated via the recursion $a(t) = a(t-1) + 2(-1)^t/\sqrt{t}$ and $a(0) = -1$. It follows that

$$\delta^1(t) = \frac{2(-1)^t}{\sqrt{t}} \notin \ell_2, \quad \delta^2(t) = \delta^3(t) = \delta^4(t) = 0.$$

Other simulation parameters are chosen as,

$$\theta = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \hat{\theta}^{i,CTA}(0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mu_i = 1.$$

The output was corrupted with an additive i.i.d. Gaussian noise with zero mean and variance 0.1.

To show the superiority of our method over the method recently proposed in [10], simulation tests were carried out for both. According to the Monte Carlo simulation method, we repeated the simulation $m = 500$ times with various noise realizations and then calculated the root mean square error $\mathbf{X}_{\text{RMS}}(t) = (\frac{1}{m} \sum_{j=1}^m \|\mathbf{X}_j(t)\|^2)^{1/2}$, where $\mathbf{X}_j = [\mathbf{X}_{j,1}^\top \ \mathbf{X}_{j,2}^\top]^\top$ and subscript j denotes the j th simulation result. Fig. 2 shows $\mathbf{X}_{\text{RMS}}(t)$ for the proposed DREM CTA algorithm and for the normalized least-mean-squares (NLMS) CTA algorithm in [10]. As can be seen from the figure, the former algorithm visibly outperforms the latter. The average error ≈ 0.1 of the DREM algorithm is fairly understandable and excusable in the face of the above sensor error variance 0.1, whereas the error of the method from [10] is approximately 2.5 times worse. Fig. 3 also displays the standard deviation of $\|\mathbf{X}_j(t)\|$. Along with DREM-CTA and NLMS-CTA, simulations were also performed for DREM-ATC and NLMS-ATC. We omit detailed presentation of the respective results and limit ourselves to the remark that the error behaviors were nearly the same as in the case of CTA.

The next simulation test aims at highlighting some theoretical findings of this paper. To this end, we consider two scenarios: where all $\delta^i(t)$'s are in ℓ_2 and, respectively, just one $\delta^j(t)$ is not in ℓ_2 . The graph, adjacency matrices, and initial conditions are taken from the previous simulation; the parameters $\mu_i = 0.1$. For the first scenario,

$$\delta^1(t) = \frac{2}{t}, \quad \delta^2(t) = \frac{1}{t^2}, \quad \delta^3(t) = \left(\frac{1}{3}\right)^t, \quad \delta^4(t) = \left(\frac{1}{4}\right)^t;$$

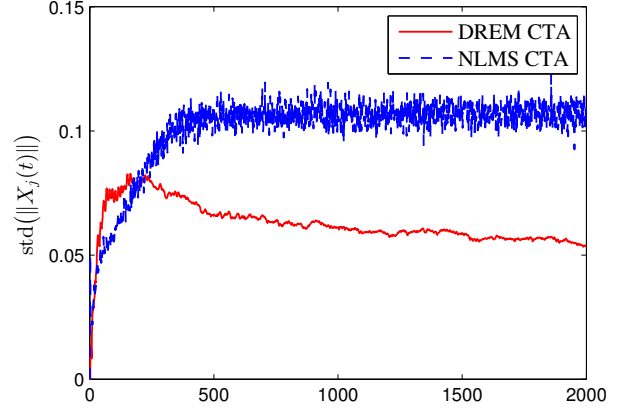


Fig. 3: Standard deviation of $\|\mathbf{X}_j(t)\|$.

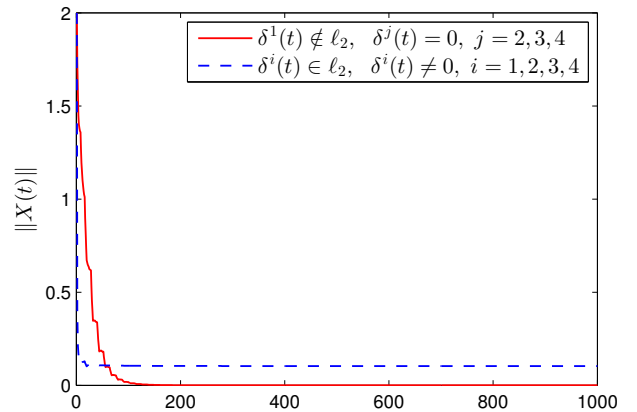


Fig. 4: Parameter estimation error.

all these sequences are in ℓ_2 . For the second scenario,

$$\delta^1(t) = \frac{2}{\sqrt{t}}, \quad \delta^2(t) = \delta^3(t) = \delta^4(t) = 0.$$

Parameters estimation errors for these cases are shown in Fig. 4. This example underscores the relevance of point i) from Assumption 2 as an “almost necessary” assumption to ensure the consistency of the parameter estimates.

VI. CONCLUSION

It has been shown in the paper that the use of the DREM method allows us to relax the assumptions and improve the convergence properties of diffusion-based distributed parameter estimation. The main contributions of the work are, on one hand, that the excitation conditions to ensure convergence are strictly weaker than the classical PE requirement. On the other hand, the connectivity assumptions of the switching topologies are very mild, while in the case of fixed topologies they are satisfied by strongly connected graphs—that is the standing assumption in the existing literature.

In contrast with other papers, e.g., [10, 11], and similarly to [3], we have adopted a *deterministic* setup of the problem. Given this situation, the analysis of the effect of the noise is limited to proving properties like ISS that, as shown in [8], holds true for DREM. As is well-known, the ISS property

ensures the robustness to additive noise of the key asymptotic stability property. In any case, we have provided a Monte Carlo simulation which includes several realizations of the noise, to show the performance improvement due to the use of DREM. Stochastic stability analysis of the proposed method in the presence of additive noise is a future line of our research.

Acknowledgments This work, except for Section IV, is supported by the Russian Science Foundation grant (project No 19-19-00403). Section IV is supported by the Ministry of Science and Higher Education of the Russian Federation (Project No. 075-15-2021-573).

REFERENCES

- [1] S. Aranovskiy, A. Bobtsov, R. Ortega and A. Pyrkin, "Performance enhancement of parameter estimators via dynamic regressor extension and mixing," *IEEE Trans. Automatic Control*, vol. 62, pp. 3546-3550, 2017.
- [2] B. Bollobas, *Modern Graph Theory*. Springer, NY, 1998.
- [3] W. Chen, C. Wen, S. Hua, and C. Sun, "Distributed cooperative adaptive identification and control for a group of continuous-time systems with a cooperative PE condition via consensus," *IEEE Trans. Automatic Control*, vol. 59, no. 1, pp. 91–106, 2014.
- [4] C. Li, H. Dong, J. Li, and F. Wang, "Distributed Kalman filtering for sensor network with balanced topology," *Systems & Control Letters*, vol. 131, 2019.
- [5] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 2012.
- [6] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Automatic Control*, vol. 48, no. 6, pp. 988–1001, 2003.
- [7] R. Ortega, S. Aranovskiy, A. Pyrkin, A. Astolfi and A. Bobtsov, "New results on parameter estimation via dynamic regressor extension and mixing: continuous and discrete-time cases," *IEEE Trans. Automatic Control*, vol. 66, no. 5, pp. 2265-2272, 2021.
- [8] J. Wang, D. Efimov, A. Bobtsov, "On robust parameter estimation in finite-time without persistence of excitation" *IEEE Trans. Automatic Control*, vol. 65, no. 4, pp. 1731-1738, 2019.
- [9] J. Wolfowitz, "Products of indecomposable, aperiodic, stochastic matrices," *Proceedings of the American Mathematical Society*, vol. 14, no. 5, pp. 733–737, 1963.
- [10] S. Xie and L. Guo, "Analysis of distributed adaptive filters based on diffusion strategies over sensor networks," *IEEE Trans. Automatic Control*, vol. 63, no. 11, pp. 3643–3658, 2018.
- [11] S. Xie and L. Guo, "Analysis of normalized least mean squares-based consensus adaptive filters under a general information condition," *SIAM Journal on Control and Optimization*, vol. 56, no. 5, pp. 3404–3431, 2018.

APPENDIX

We first make the observation that the vector sequences are related via the invertible transformation

$$\mathbf{X}_\ell(t+1) = B(t)\mathbf{Z}_\ell(t).$$

Consequently, in the sequel we will concentrate only on the stability of one LTV system, namely

$$\mathbf{X}_\ell(t+1) = B(t)\mathcal{A}(t)\mathbf{X}_\ell(t). \quad (16)$$

The inequalities between vectors from \mathbb{R}^N are meant componentwise. Integer intervals are denoted by $[k_1 : k_2] := \{t \text{ is an integer and } k_1 \leq t \leq k_2\}$; the symbols $[k_1 : k_2)$, $(k_1 : k_2]$, etc. are defined likewise. In this section, Assumptions 1–2 are supposed to be true.

A. General Preliminaries

Lemma 2 Suppose that an infinite sequence $\{\mu(t)\}_{t=0}^\infty$ with nonnegative entries $\mu(t) \geq 0$ obeys iterative inequalities of the following form

$$\mu(t+1) \leq \frac{1}{1+\delta(t)^2}\mu(t) + \eta(t), \quad (17)$$

where $\eta(t) \geq 0 \forall t$ and $\sup_{t \geq 0} \delta(t) < \infty$. If

$$\{\delta(t)\}_{t=0}^\infty \notin \ell_2 \quad \text{and} \quad \{\eta(t)\}_{t=0}^\infty \in \ell_1, \quad (18)$$

where ℓ_1 is the set of absolute summable sequences, then

$$\mu(t) \rightarrow 0 \quad \text{as} \quad t \rightarrow \infty. \quad (19)$$

Proof: Arguing via induction on t and based on (17), it is easy to see that for $t \geq 1$,

$$\mu(t) \leq \Xi(0, t)\mu(0) + \sum_{k=0}^{t-1} \Xi(k+1, t)\eta(k),$$

where

$$\Xi(p, q) := \prod_{k=p}^{q-1} \frac{1}{1+\delta(k)^2},$$

and any product $\prod_{k_1}^{k_2} \dots$ with $k_2 < k_1$ is defined to be 1. We put $B := 1 + \sup_{t \geq 0} \delta(t) < \infty$, $c := \frac{\ln(1+B^2)}{B^2}$ and note that $\ln(1+y) \geq cy$ whenever $0 \leq y \leq B^2$ since the logarithm is a concave function. Hence

$$-\ln \Xi(p, q) = \sum_{k=p}^{q-1} \ln(1+\delta(k)^2) \geq c \sum_{k=p}^{q-1} \delta(k)^2,$$

$$\Xi(p, q) \leq \exp \left[-c \sum_{k=p}^{q-1} \delta(k)^2 \right] \stackrel{(18)}{\Rightarrow}$$

$$\Xi(p, q) \rightarrow 0 \quad \text{as} \quad q \rightarrow \infty \quad \text{for any fixed } p.$$

We also note that $0 \leq \Xi(p, q) \leq 1$. By picking some integer $T > 0$, we have for all $t > T$

$$\begin{aligned} \mu(t) &\leq \Xi(0, t)\mu(0) + \sum_{k=0}^{T-1} \Xi(k+1, t)\eta(k) + \sum_{k=T}^{t-1} \Xi(k+1, t)\eta(k) \\ &\leq \Xi(0, t)\mu(0) + \sum_{k=0}^{T-1} \Xi(k+1, t)\eta(k) + \sum_{k=T}^{t-1} \eta(k) \\ &\Rightarrow \overline{\lim}_{t \rightarrow \infty} \mu(t) \leq \sum_{k=T}^{\infty} \eta(k). \end{aligned}$$

The proof is completed by letting $T \rightarrow \infty$. •

Lemma 3 In order that (19) holds for any sequence $\{\delta(t)\}_{t=0}^{\infty}$ satisfying the first relation from (18), it is not only sufficient but also necessary that the sequence $\{\eta(t)\}_{t=0}^{\infty}$ satisfies the second relation from (18).

Proof: Suppose to the contrary that $\sum_{t=0}^{\infty} \eta(t) = \infty$. Then there exists an infinite time sequence $0 < t_1 < t_2 < \dots$ such that $\varsigma_i \stackrel{\text{def}}{=} \sum_{k=t_{i-1}-1}^{t_i-2} \eta(k) \rightarrow \infty$ as $i \rightarrow \infty$. Now we consider the sequence

$$\delta(t) = \begin{cases} 1 & \text{if } t = t_i - 1 \text{ for some } i \\ 0 & \text{otherwise,} \end{cases}$$

which evidently satisfies the first relation from (18). Also, we consider the solution $\{\mu(t)\}$ of the iterative equations that result from putting $=$ in place of \leq in (17) and starts with $\mu(0) = 0$. It is easy to see that

$$\begin{aligned} \mu(t_1 - 1) &= \varsigma_1, & \mu(t_2 - 1) &= \frac{1}{2}\varsigma_1 + \varsigma_2, \\ \mu(t_3 - 1) &= \frac{1}{2^2}\varsigma_1 + \frac{1}{2}\varsigma_2 + \varsigma_3, \\ \mu(t_i - 1) &= \frac{1}{2^{i-1}}\varsigma_1 + \frac{1}{2^{i-2}}\varsigma_2 + \dots + \varsigma_i \geq \varsigma_i \quad \forall i. \end{aligned}$$

Hence $\mu(t_i) \rightarrow \infty$ as $i \rightarrow \infty$, in violation (19). The contradiction obtained proves that $\sum_{t=0}^{\infty} \eta(t) < \infty$. •

Lemma 4 The system (16) is monotone: for any its two solutions $\{x'(t)\}_{t=0}^{\infty}$ and $\{x''(t)\}_{t=0}^{\infty}$, we have

$$x'(0) \leq x''(0) \Rightarrow x'(t) \leq x''(t) \quad \forall t = 0, 1, \dots$$

The proof of this lemma is straightforward from (16) and the first inequality in (2).

Corollary 1 For any solution $\{x(t)\}_{t=0}^{\infty}$ of (16), $x(0) \geq 0 \Rightarrow x(t) \geq 0 \forall t = 0, 1, 2, \dots$

Any solution with $x(0) > 0$ is said to be *positive*. By Corollary 1, $x(t) \geq 0 \forall t$ for them. Though it will not be used in the sequel, we note that in fact $x(t) > 0 \forall t$ for them due to the “self loop property” from Assumption 1.

Corollary 2 Suppose that $z(t) \rightarrow 0$ as $t \rightarrow \infty$ for some positive solution $\{z(t)\}_{t=0}^{\infty}$ of (16). Then the origin is globally asymptotically stable equilibrium of the system (16).

Proof: For any solution $\{x(t)\}_{t=0}^{\infty}$ of (16), there exist so large $\varsigma > 0$ and $\rho > 0$ that $-\varsigma z(0) \leq x(0) \leq \rho z(0)$ since $z(0) > 0$ here. So long as both $\{-\varsigma z(t)\}_{t=0}^{\infty}$ and $\{\rho z(t)\}_{t=0}^{\infty}$ are the solutions of (16), Lemma 4 guarantees that

$$-\varsigma z(t) \leq x(t) \leq \rho z(t) \quad \forall t = 0, 1, 2, \dots$$

It remains to invoke that $z(t) \rightarrow 0$ as $t \rightarrow \infty$. •

B. Some properties of the dynamic graph

Specifically, we address the graph $\mathcal{G}(t)$ from Assumption 1 and recall that its set of nodes \bar{N} does not vary with time. A *sequential dynamic path* in the graph on the time interval $\Delta := [t_1 : t_2]$ is a sequence $E(t_1), E(t_1 + 1), E(t_1 + 2), \dots, E(t_2)$ such that $E(t)$ is an edge in the graph $\mathcal{G}(t)$ for all $t \in \Delta$ and the head of $E(t)$ coincides with the tail of $E(t + 1)$ for all

$t \in [t_1 : t_2]$. The tail of the first edge $E(t_1)$ is called the *tail of the path*, the head of the last edge $E(t_2)$ is called the *head of the path*; the path is viewed as a method of consecutively hopping from the path’s tail to its head, with making only one hop at a time and, moreover, a hop that is feasible at the concerned time.

Now we invoke the “root” node r from Assumption 2.

Lemma 5 There exists an integer $M \geq 0$ such that whenever $t_2 - t_1 \geq M$, any node $i = 1, \dots, N$ can be reached from r by following a sequential dynamic path defined on $[t_1 : t_2]$.

Proof: According to i) in Assumption 2, any node $i = 1, \dots, N$ can be reached from r by following a path in any graph $\mathcal{G}[[kp, (k + 1)p]]$. Since its set of edges is the union of the edges of the graphs $\mathcal{G}(t)$ over $t \in [kp : (k + 1)p]$, this transition comes to following some edges of some of these $\mathcal{G}(t)$ ’s. This observation is similar in spirit to what is needed to prove, but, at the same time, is not identical to the desired statement due to the following reasons:

- There is no guarantee that the edges of $\mathcal{G}(t)$ ’s are followed in the correct temporal order
- There is no guarantee that the edges of the same $\mathcal{G}(t)$ are not employed several times over the path
- The observation deals with only special intervals, i.e., those of the form $[kp : (k + 1)p]$ with an integer k .

In fact, the rest of the proof is devoted to overcoming these logical deficits.

We start with a situation where t_1 is an integer multiple of p , i.e., $t_1 = pk_1$ with an integer k_1 . Let us consider $N! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot N$ intervals $[kp, (k + 1)p], k = k_1, k_1 + 1, \dots, k_1 + N! - 1$ and the associated graphs $\mathcal{G}_k := \mathcal{G}[[kp, (k + 1)p]]$. By i) in Assumption 2, any of them contains a path from r to i . Among such paths, there is a path of the minimal length, and we pick one of them \mathcal{P}_k . Due to the minimum length property, this path does not contain loops. Meanwhile, the set \mathfrak{P}_N of all paths among N nodes that start with r and have no loops contains $(N - 1)!$ elements. It follows that in the sequence $\mathcal{P}_{k_1}, \mathcal{P}_{k_1+1}, \dots, \mathcal{P}_{k_1+N!-1}$ of $N!$ elements drawn from this set, some element is encountered no less than N times. Let \mathcal{P} be such an element, and let

$$m_1 < m_2 < \dots < m_N \in [k_1 : k_1 + N!] \quad (20)$$

be indices such that $\mathcal{P}_k = \mathcal{P}$ for any k of the form $k = m_j, j = 1, \dots, N$.

Since the path \mathcal{P} does not contain loops, it is composed of no more than $N - 1$ edges $E_1, \dots, E_d, d \leq N - 1$ enumerated in accordance with their succession in the path. Since the number of these edges does not exceed the length of the sequence $\mathcal{P}_{m_1}, \dots, \mathcal{P}_{m_N}$, every edge E_i can be attached to its own member of this sequence: $E_i \in \mathcal{P}_{m_i}$ for all $i = 1, \dots, d$. Meanwhile, \mathcal{P}_{m_i} is a path in the graph $\mathcal{G}_{m_i} := \mathcal{G}[[m_i p, (m_i + 1)p]]$, whose edges are formed by uniting the edges of the graphs $\mathcal{G}(t)$ over $t \in [m_i p, (m_i + 1)p]$. So there exists $\tau_i \in [m_i p, (m_i + 1)p]$ such that E_i is an edge of the graph $\mathcal{G}(\tau_i)$. Inequalities (20) guarantee that

$$t_1 \leq \tau_1 < \tau_2 < \dots < \tau_d < t_1 + M, \quad \text{where } M := N!p.$$

Now we are in a position to construct a sequential dynamic path from r to i that departs at time t_1 and arrives at a pre-specified time $t_2 \geq t_1 + M$. From t_1 to τ_1 , this sequential path repeatedly follows the self-loop at node r . At time τ_1 , it follows the edge E_1 , thus jumping to its head h_1 . From τ_1 to τ_2 , let the path repeatedly follow the self-loop at node h_1 , and let it follow the edge E_2 at time τ_2 , thus jumping to its head h_2 . We continue likewise until the head $h_d = i$ of the last edge E_d is reached at time τ_d , and then follow the left-loop at the node i until time t_2 . Thus the required sequential path is constructed.

In the general case where t_1 is not necessarily an integer multiple of p , the sequential path first repeatedly follows the self-loop at node r until the first time that is an integer multiple of p , after which the above construction is put in use. This necessitates to alter the choice of M in the conclusion of the lemma: $M := M + p$. •

C. Final steps in the proof of Theorem 1

From now on, we focus on an arbitrarily chosen positive solution $\{x(t)\}_{t=0}^{\infty}$ of the system (16).

Lemma 6 The following relations hold:

$$m(t) \stackrel{\text{def}}{=} \max_{i=1, \dots, N} x_i(t) \geq 0 \quad \text{and} \quad m(t+1) \leq m(t) \quad (21)$$

Proof: The first relation is immediate from Corollary 1 and the definition of the positive solution, whereas the second one is justified as follows:

$$\begin{aligned} m(t+1) &= \max_{i=1, \dots, N} x_i(t+1) \\ &\stackrel{(16)}{=} \max_{i=1, \dots, N} \frac{1}{1 + (\delta_i(t))^2} \sum_{j=1}^N a_{ij}(t) x_j(t) \\ &\leq \max_{x_j(t) \leq m(t) \forall j} m(t) \sum_{j=1}^N a_{ij}(t) \stackrel{(2)}{=} m(t). \quad \bullet \end{aligned}$$

Corollary 3 There exists a finite and nonnegative limit

$$\bar{m} = \lim_{t \rightarrow \infty} m(t). \quad (22)$$

Corollary 4 It suffices to show that $\bar{m} = 0$ since then Theorem 1 is valid thanks to Corollary 2.

Lemma 7 Suppose that at a time $t = \tau$, the graph $\mathcal{G}(t)$ contains an edge from node j to node i . Let $m^\uparrow(\tau)$ be an upper bound on $m(\tau)$. Then

$$m^\uparrow(\tau) - x_i(\tau+1) \geq \underline{\alpha}[m^\uparrow(\tau) - x_j(\tau)],$$

where $\underline{\alpha}$ is taken from ii) in Assumption 1.

Proof: For $t = \tau$, we observe that

$$\begin{aligned} x_i(t+1) &\stackrel{(16)}{=} \frac{1}{1 + (\delta_i(t))^2} \sum_{q=1}^N a_{iq}(t) x_q(t) \leq \sum_{q=1}^N a_{iq}(t) x_q(t) \\ &= \sum_{q \neq j} a_{iq}(t) x_q(t) + a_{ij}(t) x_j(t) \stackrel{(21)}{\leq} m^\uparrow(t) \sum_{q=1}^N a_{iq}(t) + \\ &a_{ij}(t) [x_j(t) - m^\uparrow(t)] \stackrel{(2)}{=} m^\uparrow(t) - a_{ij}(t) [m^\uparrow(t) - x_j(t)] \\ &\leq m^\uparrow(t) - \underline{\alpha} [m^\uparrow(t) - x_j(t)]. \quad \bullet \end{aligned}$$

Corollary 5 Let a sequential dynamic path $E(t_1), E(t_1 + 1), \dots, E(t_2)$ starts with node j and ends with node i . Then

$$m(t_1) - x_i(t_2) \geq \underline{\alpha}^{t_2 - t_1} [m(t_1) - x_j(t_1)]. \quad (23)$$

Proof: We argue via induction on $t_2 = t_1, t_1 + 1, \dots$. For $t_2 = t_1$, (23) is evident. Suppose that (23) is true for some $t_2 \geq t_1$, and consider a dynamic path $E(t_1), \dots, E(t_2), E(t_2 + 1)$ that starts with node j and ends with node i . Let i_* stand for the head of $E(t_2)$. By the induction hypothesis,

$$m(t_1) - x_{i_*}(t_2) \geq \underline{\alpha}^{t_2 - t_1} [m(t_1) - x_j(t_1)]. \quad (24)$$

Meanwhile, $m(t_1)$ is an upper bound on $m(t_2)$ by the second relation from (21). By applying Lemma 7 to $\tau := t_2, j := i_*$, and $m^\uparrow(\tau) := m(t_1)$, we see that

$$\begin{aligned} m(t_1) - x_i(t_2 + 1) &\geq \underline{\alpha} [m(t_1) - x_{i_*}(t_2)] \\ &\stackrel{(24)}{\geq} \underline{\alpha}^{t_2 - t_1 + 1} [m(t_1) - x_j(t_1)]. \quad \bullet \end{aligned}$$

To proceed, we invoke the ‘‘root’’ node from Assumption 2 and the integer M from Lemma 5.

Lemma 8 The infinite sequence $\{\eta(t) := m(t) - x_r(t)\}_{t=0}^{\infty}$ is summable, i.e. belongs to ℓ_1 .

Proof: For any time instant τ and any node i , there is a sequential dynamic path that exists within the time interval $[\tau : \tau + M]$, starts with node r , and ends with node i by Lemma 5. By applying Corollary 5 to this path, $t_1 := \tau, t_2 := \tau + M$, and $j := r$, we see that

$$m(\tau) - x_i(\tau + M) \geq \underline{\alpha}^M [m(\tau) - x_r(\tau)].$$

Taking the maximum over $i = 1, \dots, N$ yields that

$$m(\tau) - m(\tau + M) \geq \underline{\alpha}^M [m(\tau) - x_r(\tau)].$$

By taking an arbitrary integer $\theta \geq 0$ and substituting here $\tau := \theta + qM$ with an integer $q \geq 0$, we infer that

$$\begin{aligned} \sum_{q=0}^{\infty} \eta(\theta + qM) &= \sum_{q=0}^{\infty} [m(\theta + qM) - x_r(\theta + qM)] \\ &\leq \underline{\alpha}^{-M} \sum_{q=0}^{\infty} [m(\theta + qM) - m(\theta + (q+1)M)] \\ &= \underline{\alpha}^{-M} \lim_{K \rightarrow \infty} \sum_{q=0}^K [m(\theta + qM) - m(\theta + (q+1)M)] \\ &= \underline{\alpha}^{-M} \left\{ m(\theta) - \lim_{K \rightarrow \infty} m(\theta + (K+1)M) \right\} \\ &\stackrel{(22)}{=} \underline{\alpha}^{-M} [m(\theta) - \bar{m}]; \\ \sum_{t=0}^{\infty} \eta(t) &= \sum_{\theta=0}^{\infty} \sum_{q=0}^{M-1} \eta(\theta + qM) \\ &\leq \underline{\alpha}^{-M} \left[\sum_{\theta=0}^{M-1} m(\theta) - M\bar{m} \right] < \infty. \quad \bullet \end{aligned}$$

Completion of the proof of Theorem 1: Thanks to Lemma 8, $\eta(t) := m(t) - x_r(t) \rightarrow 0$ as $t \rightarrow \infty$ and so $x_r(t) \rightarrow \bar{m}$ as $t \rightarrow \infty$ by (22). Meanwhile,

$$\begin{aligned} x_r(t+1) &\stackrel{(16)}{=} \frac{1}{1 + (\delta_r(t))^2} \sum_{q=1}^N a_{rq}(t)x_q(t) \\ &\stackrel{(21)}{\leq} \frac{1}{1 + (\delta_r(t))^2} \sum_{q=1}^N a_{rq}(t)m(t) \stackrel{(2)}{=} \frac{m(t) - x_r(t) + x_r(t)}{1 + (\delta_r(t))^2} \\ &= \frac{x_r(t)}{1 + (\delta_r(t))^2} + \frac{\eta(t)}{1 + (\delta_r(t))^2} \leq \frac{x_r(t)}{1 + (\delta_r(t))^2} + \eta(t). \end{aligned}$$

So by applying Lemma 2 to $\mu(t) := x_r(t)$, $\delta(t) := \delta_r(t)$ and invoking Lemma 8, we see that $x_r(t) \rightarrow 0$ as $t \rightarrow \infty$ and so $\bar{m} = 0$. Corollary 4 completes the proof. •