

Original Article

Open Access

Distributed dynamic stochastic approximation algorithm over time-varying networks



Kewei Fu^{1,2}, Han-Fu Chen^{1,2} and Wenxiao Zhao^{1,2*}

Abstract

In this paper, a distributed stochastic approximation algorithm is proposed to track the dynamic root of a sum of time-varying regression functions over a network. Each agent updates its estimate by using the local observation, the dynamic information of the global root, and information received from its neighbors. Compared with similar works in optimization area, we allow the observation to be noise-corrupted, and the noise condition is much weaker. Furthermore, instead of the upper bound of the estimate error, we present the asymptotic convergence result of the algorithm. The consensus and convergence of the estimates are established. Finally, the algorithm is applied to a distributed target tracking problem and the numerical example is presented to demonstrate the performance of the algorithm.

Keywords: Distributed algorithm, Dynamic stochastic approximation algorithm, Time-varying network

1 Introduction

In recent years, there is a huge increase in the scale of data in many real-life application problems. The traditional centralized computing method faces many challenges and sometimes is entirely infeasible for large-scale problems. As a result, distributed algorithms over multi-agent systems have received much attention from researchers of diverse areas, including consensus problem [1–4], resource allocation (RA) [5, 6], multi-unmanned aerial vehicle (MUAV) control [7], and distributed target tracking [8] etc. Distributed algorithms are usually associated with a network of agents where each agent has limited computation and communication ability. The agents are required to cooperatively achieve a global objective by using their local observations and information transmitted from their neighbors. Compared with centralized

approaches, distributed algorithms have the advantage of robustness on network link failure, privacy protection, and reduction on communication and computation cost.

One of the important branches of distributed algorithms is the distributed optimization problem. It seeks the minimizer of the global function which is written as a sum of the local functions of the agents. In particular, the distributed optimization for time-invariant cost function has become a mature discipline with many results, see [9–11] and references within. On the other hand, optimization problems with time-varying cost function have attracted much attention due to its appearance in various applications, for example, signal processing [12] and online optimization [13, 14]. The main challenge of time-varying optimization lies in the fact that the minimizer of the time-varying cost function is changing with time. Since the traditional optimization algorithms can only move the estimates towards the minimizer of the cost function of the current time, they cannot track the movement of the minimizer in the dynamic environment. To cope with this issue, two different strategies have been developed.

*Correspondence: wzxhao@amss.ac.cn

¹Key Laboratory of Systems and Control, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, People's Republic of China

²School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China

The first one is running method [15, 16], where the algorithms sample the time-varying cost function at a fixed frequency, and perform the traditional optimization algorithms on the sampled function in between the sample time. The second one is prediction-correction method [17, 18], where the algorithms only optimize the cost function of the current time at each steps. But an additional information of the dynamic of the moving minimizer is required, such as the second derivative of the cost function [18], or an additional constrain on the dynamic of the minimizer [14]. While we consider a different problem in this paper, the method we utilize is much similar to the second method.

It can be noticed that the optimization problem can often be transformed into a root-seeking problem, since the optimization of a differentiable convex function is equivalent to seeking the root of its gradient, and for the case where the gradients are unavailable we can use the finite time difference to estimate the gradient [19]. So it's natural to consider the distributed root-tracking problem. Distributed stochastic approximation for time-invariant regression functions has been studied by many researchers as a solution for distributed root-seeking problem [20–22]. Inspired by the work of dynamic stochastic approximation [23, 24], in this paper, we propose a distributed stochastic approximation algorithm for tracking the changing root of a sum of time-varying regression functions over a network. Each agent is aimed at tracking the changing root of the global function, but it can only access a noise-corrupted local observation and the information transmitted from its neighbor. In addition, the noise-corrupted dynamics of the roots of the global regression function is assumed to be known to all agents.

In this paper, the distributed root-tracking problem for time-varying regression function is considered. First, motivated by the truncation technique given in [22], a distributed stochastic approximation algorithm with expanding truncations is introduced. The key difference is that the observation of the local function in this algorithm is a noise-corrupted one, while an exact gradient information is often required in the optimization algorithms mentioned above. Second, under the assumption that the noise-corrupted dynamics of the global roots is known to all agents, the convergence conditions of the algorithm are introduced. Third, it is proved that the estimates generated by the distributed algorithm are of both consensus and convergence with probability one. Finally, we apply this algorithm to a distributed target tracking problem. The numerical example is given demonstrating the performance of the algorithm.

The rest of the paper is organized as follows. The problem formulation and the distributed stochastic approximation algorithm are given in Section 2. The convergence

conditions and results are presented in Section 3. To help the proof of the convergence result, two auxiliary sequences are defined and analysed in Section 4. The proof of the main result is given in Section 5. In Section 6, a distributed target tracking problem is solved by the algorithm and the numerical example is demonstrated. Some concluding remarks are addressed in Section 7.

2 Problem formulation and distributed root-tracking algorithm

2.1 Problem formulation

Consider a network system consisting of N agents. The interaction relationship among agents is described by a time-varying digraph $\mathcal{G}(k) = \{\mathcal{V}, \mathcal{E}(k)\}$, where k is the time index, $\mathcal{V} = \{1, \dots, N\}$ is the agent set, and $\mathcal{E}(k) \subset \mathcal{V} \times \mathcal{V}$ is the edge set. By $(i, j) \in \mathcal{E}(k)$ we mean that agent j can receive information from agent i at time k . Assume $(i, i) \in \mathcal{E}(k)$ for $\forall k = 1, 2, \dots$. Denote the neighbor of agent i at time k by $N_i(k) = \{j \in \mathcal{V} : (j, i) \in \mathcal{E}(k)\}$. The adjacency matrix associated with the graph is denoted by $W(k) = [w_{ij}(k)]_{i,j=1}^N$, where $w_{ij}(k) > 0$ if and only if $(j, i) \in \mathcal{E}(k)$, otherwise $w_{ij}(k) = 0$.

A time-independent digraph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ is called *strongly connected* if for any $i, j \in \mathcal{V}$, there exists a directed path from i to j . A directed path is a sequence of edges $(i, i_1), (i_1, i_2), \dots, (i_{p-1}, j)$ in the digraph with distinct agents $i_k \in \mathcal{V}$, $0 \leq k \leq p-1$, where p is called the length of this path. A nonnegative matrix A is called *doubly stochastic* if $A\mathbf{1} = \mathbf{1}$ and $\mathbf{1}^T A = \mathbf{1}^T$.

The time-varying global regression function is given by

$$f_k(\cdot) = \frac{1}{N} \sum_{i=1}^N f_{i,k}(\cdot), \quad (1)$$

where $f_{i,k}(\cdot) : \mathbb{R}^l \rightarrow \mathbb{R}^l$ is the local function associated with agent i . Denote by θ_k the root of the sum function $f_k(\cdot)$ at time k , i.e., $f_k(\theta_k) = 0$, $k = 1, 2, \dots$

Further, assume that the dynamics of the root θ_k is governed by

$$\theta_{k+1} = g_k(\theta_k) + \xi_{k+1}, \quad k \geq 0, \quad (2)$$

where the function $g_k(\cdot) : \mathbb{R}^l \rightarrow \mathbb{R}^l$ is known for all agents, and $\{\xi_k\}$ is the sequence of dynamic noises. As we can see in Section 6, this assumption is reasonable in some real-life application problems and have been studied before in [14, 25, 26].

For each agent i , the distributed root-tracking problem is to track the dynamic root of the time-varying global function by using its noise-corrupted observation of local function $f_{i,k}(\cdot)$, the dynamic information of the root $g_k(\cdot)$, and the information obtained from its adjacent neighbors.

2.2 Algorithm

We now introduce the distributed root-tracking algorithm as follows:

$$x'_{i,k+1} = \left\{ \sum_{j \in N_i(k)} w_{ij}(k) g_k(x_{j,k}) + a_k O_{i,k+1} \right\} \mathbb{I}_{[\sigma_{i,k} = \hat{\sigma}_{i,k}]} + h_k(x^*) \mathbb{I}_{[\sigma_{i,k} < \hat{\sigma}_{i,k}]}, \quad (3)$$

$$x_{i,k+1} = x'_{i,k+1} \mathbb{I}_{[\|x'_{i,k+1} - h_k(x^*)\| \leq M \hat{\sigma}_{i,k}]} + h_k(x^*) \mathbb{I}_{[\|x'_{i,k+1} - h_k(x^*)\| > M \hat{\sigma}_{i,k}]}, \quad (4)$$

$$\sigma_{i,k+1} = \hat{\sigma}_{i,k} + \mathbb{I}_{[\|x'_{i,k+1} - h_k(x^*)\| > M \hat{\sigma}_{i,k}]}, \quad (5)$$

$$\sigma_{i,0} = 0, \quad \hat{\sigma}_{i,k} = \max_{j \in N_i(k)} \sigma_{j,k}, \quad (6)$$

$$O_{i,k+1} = f_{i,k+1}(g_k(x_{i,k})) + \epsilon_{i,k+1}, \quad (7)$$

where 1) $x_{i,k} \in \mathbb{R}^l$ is the estimate of θ_k given by the agent i at time k , 2) $O_{i,k+1}$ defined by (7) is the local observation of agent i , 3) $\{a_k\}_{k \geq 0}$ is the sequence of the step-sizes used by all agents, 4) x^* is a fixed vector in \mathbb{R}^l known to all agents, 5) $\{M_k\}_{k \geq 0}$ is a sequence of positive numbers increasingly diverging to infinity with $M_0 \geq \|x^*\|$, 6) $\sigma_{i,k}$ is the truncation number of agent i up-to-time k , and 7) $h_k(\cdot)$ is a function defined as below

$$h_1(x) = g_1(x), \quad h_k(x) = g_k(h_{k-1}(x)), \quad \text{for } k = 2, 3, \dots \quad (8)$$

Let us explain the algorithm. 1) For agent i , the estimate $x_{i,k}$ is the estimate of θ_k . Since the dynamics of $\{\theta_k\}$ is governed by (2), in order to make sure the estimate track the dynamic root, the update at time $k+1$ utilize $g_k(x_{i,k})$ instead of $x_{i,k}$ as it was shown in (3). 2) For agent i , the truncation happens when one of the following cases hold true: a) $\sigma_{i,k} < \hat{\sigma}_{i,k}$, which means that there is at least one neighbor whose truncation number is larger than that of agent i . b) $\|x'_{i,k+1} - h_k(x^*)\| > M \hat{\sigma}_{i,k}$, which means that the distance between the intermediate value $x'_{i,k+1}$ and $h_k(x^*)$ is larger than the truncation bound. When truncation happens, the estimate $x_{i,k}$ is pulled to $h_{k-1}(x^*)$. 3) It can be seen that the truncation may not happen at the same time for different agents in the network. So for agent i , the update (5) makes sure that the truncation number of i is not smaller than the largest truncation number of its neighbors, i.e. $\hat{\sigma}_{i,k}$. As to be shown in Lemma 4, this technique guarantees that the difference between truncation numbers of different agents is bounded, which helps the algorithm converge. 4) The truncation mechanism makes sure that the estimates $x_{i,k}$ won't be too far away from $h_{k-1}(x^*)$. As to be shown in Lemma 1, we can prove that the distance between $\{h_{k-1}(x^*)\}$ and the dynamic root $\{\theta_k\}$ is bounded. So it is reasonable to choose this truncation condition.

Remark 1 In our previous work [27], we proposed the distributed root-tracking algorithm without the expanding truncation. To make sure the algorithm converge, we assumed the dynamic root $\{\theta_k\}$ and the estimate of all agents $\{x_{i,k}\}$ are bounded sequences in [27]. With the introduction of the expanding truncation mechanism, this assumption is removed in this paper.

3 Assumptions and convergence result

3.1 Assumptions

Let us list the assumptions to be used in the paper.

A1 $a_k > 0, a_k \rightarrow 0, \sum_{k=1}^{\infty} a_k = \infty$.

A2 There exists a continuously differentiable function $v(\cdot) : \mathbb{R}^l \rightarrow \mathbb{R}$ such that $v(x) \neq 0$ for $\forall x \neq 0$, $v(0) = 0$ and for any $0 < r_1 < r_2 < \infty$

$$\sup_k \sup_{r_1 \leq \|x - \theta_k\| \leq r_2} f_k^T(x) v_x(x - \theta_k) < -a,$$

where a is a positive constant possibly depending on r_1, r_2 . A constant $r > \eta$ exists such that

$$\sup_{\|y\| \leq \eta} v(y) < \sup_{\|x\|=r} v(x),$$

where η is an unknown constant specified later in Lemma 1.

A3 The class of functions $\{f_{i,k}(\cdot)\}_{k \geq 0}$ is equi-continuous for $i = 1, \dots, N$, i.e., for any fixed i and any $\epsilon > 0$, there exists $\delta > 0$ such that

$$\|f_{i,k}(x) - f_{i,k}(y)\| \leq \epsilon \quad \forall k, \quad \text{whenever } \|x - y\| \leq \delta,$$

where δ only depends on ϵ . Furthermore, for $\forall c > 0$, there exists a constant $\alpha(c)$ such that

$$\|f_{i,k}(\theta_k + v)\| < \alpha(c) \quad \text{for } \forall v \text{ with } \|v\| \leq c, \quad \forall i \in \mathcal{V}, k = 1, 2, \dots$$

A4 a) The adjacent matrices $W(k) \forall k \geq 0$ are doubly stochastic;

b) There exists a constant $0 < \kappa < 1$ such that

$$w_{ij}(k) \geq \kappa \quad \forall j \in N_i(k) \quad \forall i \in \mathcal{V} \quad \forall k \geq 0,$$

c) The digraph $\mathcal{G}_{\infty} = \{\mathcal{V}, \mathcal{E}_{\infty}\}$ is strongly connected, where

$$\mathcal{E}_{\infty} = \{(j, i) : (j, i) \in \mathcal{E}(k) \text{ for infinitely many indices } k\},$$

d) There exists a positive integer B such that

$$(j, i) \in \mathcal{E}(k) \cup \mathcal{E}(k+1) \cup \dots \cup \mathcal{E}(k+B-1)$$

for all $(j, i) \in \mathcal{E}_{\infty}$ and any $k \geq 0$.

A5 For any $i \in \mathcal{V}$, the noise sequence $\{\epsilon_{i,k+1}\}_{k \geq 0}$ is such that

$$\lim_{T \rightarrow 0} \limsup_{k \rightarrow \infty} \frac{1}{T} \left\| \sum_{m=n_k}^{m(n_k, t_k)} a_m \epsilon_{i,m} \right\| = 0, \quad \forall t_k \in [0, T]$$

where $m(k, T) \triangleq \max \{m : \sum_{i=k}^m a_i \leq T\}$ and $\{n_k\}$ denotes the indices of any convergent subsequence $\{x_{i,n_k} - \theta_{n_k}\}$.

A6 $g_k(\cdot) : \mathbb{R}^l \rightarrow \mathbb{R}^l$ is equi-continuous with respect to k and is such that $\|d_k(x)\| \leq \gamma_k \|x - \theta_k\|$ $\forall x, k = 1, 2, \dots$, where

$$d_k(x) \triangleq g_k(x) - g_k(\theta_k) - (x - \theta_k),$$

$\gamma_k = o(a_k)$, and $\sum_{k=1}^{\infty} \gamma_k < \infty$.

A7 $\|\xi_k\| = o(a_k)$ and $\sum_{k=1}^{\infty} \|\xi_k\| < \infty$.

A1 and A2 are the standard assumptions for stochastic approximation. A3 implies the local boundedness of the functions $f_{i,k}(\cdot)$. Notice that the upper bound $\alpha(c)$ in A3 should be uniform with respect to k .

A4 describes the information exchanging among agents. We refer to [9] for the detailed explanation. Set $\Phi(k, k + 1) \triangleq \mathbf{I}_N$ and

$$\Phi(k, s) \triangleq W(k) \cdots W(s) \quad \forall k \geq s.$$

By Proposition 1 in [9] it follows that there exist constants $c > 0$ and $0 < \rho < 1$ such that

$$\|\Phi(k, s) - \frac{1}{N} \mathbf{1}\mathbf{1}^T\| \leq c\rho^{k-s+1} \quad \forall k \geq s. \tag{9}$$

Notice that in A5 b), the noise condition is required only along the indices of any convergent subsequence $\{x_{i,n_k} - \theta_{n_k}\}$. As to be seen in the next section, this makes the convergence analysis much easier compared with requiring the noise condition to hold along the whole sequence.

In A6, $d_k(x)$ measures the difference between the estimation error $x_{i,k} - \theta_k$ and the prediction error $g_k(x_{i,k}) - g_k(\theta_k)$. This assumption implies that the dynamic of the root, i.e. $g_k(\cdot)$, will tend to be a linear function as time k goes to infinity. For example, if the dynamics of the changing roots is $g_k(x) = x + c$, then A6 holds with $\gamma_k = 0$.

3.2 Main result

Set $\text{col}\{x_1, \dots, x_m\} \triangleq (x_1^T, \dots, x_m^T)^T$, and define

$$X_k \triangleq \text{col}\{x_{1,k}, \dots, x_{N,k}\},$$

$$\Theta_k \triangleq \mathbf{1} \otimes \theta_k \in \mathbb{R}^{Nl},$$

$$\epsilon_k \triangleq \text{col}\{\epsilon_{1,k}, \dots, \epsilon_{N,k}\},$$

$$\Xi_k \triangleq \mathbf{1} \otimes \xi_k \in \mathbb{R}^{Nl},$$

$$G_k(X_k) \triangleq \text{col}\{g_k(x_{1,k}), \dots, g_k(x_{N,k})\},$$

$$F_{k+1}(X_k) \triangleq \text{col}\{f_{1,k+1}(x_{1,k}), \dots, f_{N,k+1}(x_{N,k})\},$$

$$D_k(X_k) \triangleq \text{col}\{d_k(x_{1,k}), \dots, d_k(x_{N,k})\}.$$

Further, we denote the disagreement vector of X_k by $X_{\perp,k} \triangleq D_{\perp} X_k$ with $D_{\perp} \triangleq (\mathbf{I}_N - \frac{\mathbf{1}\mathbf{1}^T}{N}) \otimes \mathbf{I}_l$. Define $x_k \triangleq$

$\frac{1}{N} \sum_{i=1}^N x_{i,k}$, the average of all agents' estimates at time k . Define $\Delta_{i,k} \triangleq x_{i,k} - \theta_k$, $\Lambda_k \triangleq X_k - \Theta_k$, and $\Delta_k \triangleq x_k - \theta_k$.

Theorem 1 Let $\{x_{i,k}\}$ be the estimates produced by (3)–(7) with an arbitrary initial value $x_{i,0}$. Assume A1–A4 and A6 hold. If for a fixed sample ω , A5 holds for all agents, and A7 holds, then for this ω , the following assertion takes place:

i) There exists a positive integer k_0 depending on ω such that

$$x_{i,k+1} = \sum_{j \in N_i(k)} w_{ij}(k) g_j(x_{j,k}) + a_k O_{i,k+1}. \tag{10}$$

or in the compact form

$$X_{k+1} = (W(k) \otimes \mathbf{I}_l) G_k(X_k) + a_k (F_{k+1}(G(X_k)) + \epsilon_{k+1}) \tag{11}$$

for any $k \geq k_0$;

ii)

$$\lim_{k \rightarrow \infty} X_{\perp,k} = \mathbf{0}, \quad \lim_{k \rightarrow \infty} \Lambda_k = \mathbf{0}.$$

Theorem 1 i) shows that the truncation ceases after a finite number of steps. This implies that the difference between the estimate $x_{i,k}$ and $h_{k-1}(x^*)$ is bounded, which is desirable as to be shown in Lemma 1. Before we move on to the proof of Theorem 1, we first show that the truncation mechanism is reasonable.

Lemma 1 If A6 and A7 hold, the sequence $\{h_k(x^*) - \theta_{k+1}\}$ is bounded for any x^* .

Proof By A6 and A7 from (2) it follows that

$$\begin{aligned} \|h_k(x^*) - \theta_{k+1}\| &= \|g_k(h_{k-1}(x^*)) - g_k(\theta_k) - \xi_{k+1}\| \\ &= \|d_k(h_{k-1}(x^*)) + h_{k-1}(x^*) - \theta_k - \xi_{k+1}\| \\ &\leq (1 + \gamma_k) \|h_{k-1}(x^*) - \theta_k\| + \|\xi_{k+1}\| \\ &\leq \prod_{i=1}^k (1 + \gamma_i) \|g_1(x^*) - \theta_1\| + \sum_{i=1}^k \prod_{j=i+1}^k (1 + \gamma_j) \|\xi_{j+1}\| \\ &\leq \prod_{i=1}^{\infty} (1 + \gamma_i) \|g_1(x^*) - \theta_1\| + \sum_{i=1}^{\infty} \prod_{j=i+1}^{\infty} (1 + \gamma_j) \|\xi_{j+1}\| \\ &\triangleq \eta, \end{aligned}$$

where $\prod_{i=1}^{\infty} (1 + \gamma_i) < \infty$ is implied by $\sum_{j=1}^{\infty} \gamma_j < \infty$. \square

As we mentioned in Section 2, Lemma 1 shows that the distance between $\{h_{k-1}(x^*)\}$ and $\{\theta_k\}$ is bounded. Since we hope the estimate $\{x_{i,k}\}$ generated by the algorithm (3)–(7) track the root $\{\theta_k\}$, the truncation mechanism $\mathbb{I}_{[\|x'_{i,k+1} - h_k(x^*)\| \leq M\hat{\sigma}_{i,k}]}$ is intuitively reasonable.

4 Auxiliary sequences

The next two sections of this paper focus on the proof of Theorem 1. But prior to analyzing $\{x_{i,k}\}$, we need to introduce two auxiliary sequences $\{\tilde{x}_{i,k}\}$ and $\{\tilde{\epsilon}_{i,k}\}$ for each agent $i \in \mathcal{V}$. The motivation of constructing these two sequences comes from the character of distributed algorithm with expanding truncation. Recall the convergence analysis of stochastic approximation algorithm with expanding truncation (SAAWET) [28]. The key step is to show that the truncations cease after a finite number of steps, therefore, the boundedness of the estimates is established. If the number of truncations increases unboundedly, then the estimate is pulled back to x^* infinitely many times. This produces a convergent subsequence from the estimate sequence. Then a contradiction can be shown analysing the property along this subsequence, which proves the boundedness of the estimates.

Although the problem in this paper is different from the one in [28] since the regression function is time-varying in this paper, we use the same approach to prove the boundedness of the estimates. Notice the distributed algorithm with expanding truncation (3)–(7). When $\lim_{k \rightarrow \infty} \sigma_{i,k} = \infty \forall i \in \mathcal{V}$, the estimate $x_{i,k}$ is pulled back to $h_{k-1}(x^*)$ infinitely times. By lemma 1, we know that $\|h_{k-1}(x^*) - \theta_k\|$ is bounded. So $\{\Delta_{i,k}\}$ contains a convergent subsequence. However, $\{\Delta_k\}$ may still not contain any convergent subsequences. This is because truncation may occur at different times for different $i \in \mathcal{V}$. Therefore, the analysis approach used for SAAWET cannot directly be applied to the algorithm (3)–(7).

To overcome this difficulty, we introduce the auxiliary sequences $\{\tilde{x}_{i,k}\}$ and $\{\tilde{\epsilon}_{i,k}\}$. As to be shown, the auxiliary sequences $\{\tilde{x}_{i,k}\}$ satisfies the recursions (19)–(21), for which the number of truncation at time k for all agents is the same and the estimates $\tilde{x}_{i,k}$ for all the agents are pulled back to $h_{k-1}(x^*)$ when $\sigma_k > \sigma_{k-1}$. The auxiliary noise $\{\tilde{\epsilon}_{i,k}\}$ satisfies a condition similar to A5 b). These make the analysis for (19)–(21) feasible.

It is shown below that the important feature of the auxiliary sequences consists in that $\{\tilde{x}_{i,k}\}$ and $\{x_{i,k}\}$ coincide in a finite number of steps, which means that the convergence of these two sequences is equivalent.

Denote by $\tau_{i,m} \triangleq \inf\{k : \sigma_{i,k} = m\}$ the smallest time when the truncation number of agent i has reached m , by $\tau_m \triangleq \min_{i \in \mathcal{V}} \tau_{i,m}$ the smallest time when at least one of agents has its truncation number reached m , and by

$$\sigma_k \triangleq \max_{i \in \mathcal{V}} \sigma_{i,k} \quad (12)$$

the largest truncation number among all agents at time k . Set $\tilde{\tau}_{i,m} \triangleq \tau_{i,m} \wedge \tau_{m+1}$, where $a \wedge b = \min\{a, b\}$.

For any $i \in \mathcal{V}$, define the auxiliary sequences $\{\tilde{x}_{i,k}\}_{k \geq 0}$ and $\{\tilde{\epsilon}_{i,k}\}_{k \geq 0}$ as follows:

$$\begin{aligned} \tilde{x}_{i,k} &\triangleq h_{k-1}(x^*), \tilde{\epsilon}_{i,k+1} \\ &\triangleq -f_{i,k+1}(h_k(x^*)), \forall k : \tau_m \leq k < \tilde{\tau}_{i,m}, \end{aligned} \quad (13)$$

$$\begin{aligned} \tilde{x}_{i,k} &\triangleq x_{i,k}, \tilde{\epsilon}_{i,k+1} \\ &\triangleq \epsilon_{i,k+1}, \forall k : \tilde{\tau}_{i,m} \leq k < \tau_{m+1}, \end{aligned} \quad (14)$$

where m is an integer.

Note that for the considered sample ω there exists a unique integer $m \geq 0$ corresponding to an integer $k \geq 0$ such that $\tau_m \leq k < \tau_{m+1}$. By definition $\tilde{\tau}_{i,m} \leq \tau_{m+1} \forall i \in \mathcal{V}$. So, $\{\tilde{x}_{i,k}\}_{k \geq 0}$ and $\{\tilde{\epsilon}_{i,k}\}_{k \geq 0}$ are uniquely determined by the sequences $\{x_{i,k}\}_{k \geq 0}$ and $\{\epsilon_{i,k}\}_{k \geq 0}$.

Lemma 2 For $k \in [\tau_m, \tau_{m+1})$, the following assertions hold:

$$\begin{aligned} i) \tilde{x}_{i,k} &= h_{k-1}(x^*), \tilde{\epsilon}_{i,k+1} \\ &= -f_{i,k+1}(h_k(x^*)), \text{ if } \sigma_{i,k} < m; \end{aligned} \quad (15)$$

$$ii) \tilde{x}_{i,k} = x_{i,k}, \tilde{\epsilon}_{i,k+1} = \epsilon_{i,k+1}, \text{ if } \sigma_{i,k} = m; \quad (16)$$

$$iii) \tilde{x}_{j,k} = h_{k-1}(x^*), \text{ if } \sigma_{j,k-1} < m; \quad (17)$$

$$iv) \tilde{x}_{j,k+1} = h_k(x^*), \forall j \in \mathcal{V}, \text{ if } \sigma_{k+1} = m + 1. \quad (18)$$

Proof i) Since $\sigma_{i,k} < m$, by the definition of $\tau_{i,m}$ and the fact that $k \in [\tau_m, \tau_{m+1})$, we know $\tau_{i,m} > k$. Thus, $\tilde{\tau}_{i,m} = \tau_{i,m} \wedge \tau_{m+1} > k$. Hence, we conclude (15) from (13).

ii) Since $\sigma_{i,k} = m$, by definition we have $\tau_{i,m} \leq k$. Hence $\tilde{\tau}_{i,m} = \tau_{i,m} \wedge \tau_{m+1} = \tau_{i,m} \leq k$. So by (14) we conclude (16).

iii) By $\tau_m \leq k < \tau_{m+1}$ we know $\sigma_{j,k} \leq m$. We consider two cases: $\sigma_{j,k} = m$, and $\sigma_{j,k} < m$. 1) For $\sigma_{j,k} = m$, since $\sigma_{j,k-1} < m$, we know that truncation happens at time k for agent j . Truncation happens only when one of the following cases hold true: a) $\sigma_{j,k-1} < \hat{\sigma}_{j,k-1}$. For this case, by (3) we have $x'_{j,k} = h_{k-1}(x^*)$, hence by (4) we have $x_{j,k} = h_{k-1}(x^*)$; b) $\|x'_{j,k} - h_{k-1}(x^*)\| > M_{\hat{\sigma}_{j,k-1}}$. For this case, by (4) we have $x_{j,k} = h_{k-1}(x^*)$. In conclusion, when $\sigma_{j,k} = m$ holds true, we have $x_{j,k} = h_{k-1}(x^*)$. Furthermore, from (16) it follows that $\tilde{x}_{j,k} = x_{j,k}$ if $\sigma_{j,k} = m$. So, we have $\tilde{x}_{j,k} = h_{k-1}(x^*)$. 2) For $\sigma_{j,k} < m$, from (15) we have $\tilde{x}_{j,k} = h_{k-1}(x^*)$.

iv) From $k \in [\tau_m, \tau_{m+1})$ we know $\sigma_k = m$. Hence from $\sigma_{k+1} = m + 1$ by definition we have $\tau_{m+1} = k + 1$, and $k+1 \in [\tau_{m+1}, \tau_{m+2})$. By $\sigma_k = m$ we see $\sigma_{j,k} < m+1 \forall j \in \mathcal{V}$. Then we derive (18) by (17). \square

Lemma 3 The auxiliary sequences $\{\tilde{x}_{i,k}\}$, $\{\tilde{\epsilon}_{i,k}\}$ defined by (13)(14) satisfy the following recursions:

$$\hat{x}_{i,k+1} = \sum_{j \in N_i(k)} w_{ij}(k) g_k(\tilde{x}_{j,k}) + a_k (f_{i,k+1}(g_k(\tilde{x}_{i,k})) + \tilde{\epsilon}_{i,k+1}) \quad (19)$$

$$\tilde{x}_{i,k+1} = \hat{x}_{i,k+1} \mathbb{I}_{\left[\|\hat{x}_{j,k+1} - h_k(x^*)\| \leq M_{\sigma_k}, \forall j \in \mathcal{V} \right]} + h_k(x^*) \mathbb{I}_{\left[\exists j \in \mathcal{V} \|\hat{x}_{j,k+1} - h_k(x^*)\| > M_{\sigma_k} \right]} \quad (20)$$

$$\sigma_{k+1} = \sigma_k + \mathbb{I}_{\left[\exists j \in \mathcal{V} \|\hat{x}_{j,k+1} - h_k(x^*)\| > M_{\sigma_k} \right]}, \sigma_0 = 0. \quad (21)$$

Proof We prove this by induction.

First we prove (19)–(21) for $k = 0$. Since $0 \in [\tau_0, \tau_1)$ and $\sigma_0 = 0 \forall i \in \mathcal{V}$, by (16) we have $\tilde{x}_{i,0} = x_{i,0}$, $\tilde{\epsilon}_{i,1} = \epsilon_{i,1} \forall i \in \mathcal{V}$. Then by $\hat{\sigma}_{i,0} = \sigma_{i,0} = 0 \forall i \in \mathcal{V}$, from (3)(19) we see

$$\hat{x}_{i,1} = x'_{i,1}, \forall i \in \mathcal{V}. \quad (22)$$

Now we prove that $\tilde{x}_{i,1}$ and σ_1 generated by (19)–(21) are consistent with the definition (12)(13)(14). We consider two cases:

i) There is no truncation at time $k = 1$, i.e., $\sigma_{i,1} = 0 \forall i \in \mathcal{V}$. Since $\sigma_{i,0} = 0$, by (5) we know that $\|x'_{i,1} - h_0(x^*)\| \leq M_0$. Then we have $x_{i,1} = x'_{i,1}$ by (4), and $\tilde{x}_{i,1} = \hat{x}_{i,1}$, $\sigma_1 = 0$ by (20)(21). Combining these with (22) it is shown that $\tilde{x}_{i,1} = x_{i,1} \forall i \in \mathcal{V}$, which is consistent with (14) since $\tilde{\tau}_{i,0} \leq 1 < \tau_1$. By (12) we see $\sigma_1 = \max_{i \in \mathcal{V}} \sigma_{i,1} = 0$, which is consistent with the one derived from (21).

ii) There is a truncation at $k = 1$ for agent i_0 , i.e., $\sigma_{i_0,1} = 1$. Then by (4)(5) we have $x_{i_0,1} = h_0(x^*)$, $\|x'_{i_0,1} - h_0(x^*)\| > M_0$. Hence $\|\hat{x}_{i_0,1} - h_0(x^*)\| > M_0$ by (22). From (20)(21) we have $\hat{x}_{i,1} = h_0(x^*) \forall i \in \mathcal{V}$ and $\sigma_1 = 1$. By (12) from $\sigma_{i_0,1} = 1$ we derive $\sigma_1 = 1$. Since $0 \in [\tau_0, \tau_1)$ and $\sigma_1 = 1$, by (18) we have $\tilde{x}_{i,1} = h_0(x^*) \forall i \in \mathcal{V}$. Thus, $\tilde{x}_{i,1}$ and σ_1 defined by (13)(14)(12) are consistent with those generated by (19)–(21).

In summary, we prove the lemma for $k = 0$.

By induction, we assume (19)–(21) hold for $k = 0, 1, \dots, p$. At a fixed sample ω for a given integer p there exists a unique integer m such that $\tau_m \leq p < \tau_{m+1}$. Now we aim to show that (19)–(21) hold for $k = p + 1$. Before this, we first express $\hat{x}_{i,p+1} \forall i \in \mathcal{V}$ produced by (19) for the following two cases:

Case 1: $\sigma_{i,p} < m$. Since $p \in [\tau_m, \tau_{m+1})$, by (15) we see

$$\tilde{x}_{i,p} = h_{p-1}(x^*), \quad \tilde{\epsilon}_{i,p+1} = -f_{i,p+1}(h_p(x^*)). \quad (23)$$

From $\sigma_{i,p} < m$ it follows that $\sigma_{j,p-1} < m \forall j \in N_i(p)$ by (5). Then by (17) we have $\tilde{x}_{j,p} = h_{p-1}(x^*) \forall j \in N_i(p)$, which combining with (19)(23) shows

$$\hat{x}_{i,p+1} = h_p(x^*) \quad \forall i : \sigma_{i,p} < m \quad (24)$$

Case 2: $\sigma_{i,p} = m$. By $\tau_m \leq p < \tau_{m+1}$ we have $\sigma_{j,p} \leq m \forall j \in \mathcal{V}$ and hence by (6) we derive

$$\hat{\sigma}_{i,p} = m, \quad \forall i : \sigma_{i,p} = m \quad (25)$$

Then by (3)

$$x'_{i,p+1} = \sum_{j \in N_i(p)} w_{ij}(p) g_p(x_{j,p}) + a_p (f_{i,p+1}(g_p(x_{i,p})) + \epsilon_{i,p+1}). \quad (26)$$

From $\sigma_{i,p} = m$ and $p \in [\tau_m, \tau_{m+1})$, by (16) it can be shown that

$$\tilde{x}_{i,p} = x_{i,p}, \quad \tilde{\epsilon}_{i,p+1} = \epsilon_{i,p+1}. \quad (27)$$

Substituting (27) into (19), by (26) we know that

$$\hat{x}_{i,p+1} = x'_{i,p+1} \quad \forall i : \sigma_{i,p} = m. \quad (28)$$

So we have expressed $\hat{x}_{i,p+1} \forall i \in \mathcal{V}$ produced by (19) for the two cases above.

Since $\tau_m \leq p < \tau_{m+1}$, we have $\sigma_p < m + 1$ and hence $\sigma_{p+1} \leq m + 1$. From $\tau_m \leq p$ it follows that $\sigma_p = m$ and $\sigma_{p+1} \geq m$, hence $m \leq \sigma_{p+1} \leq m + 1$.

Now we show that $\tilde{x}_{i,p+1}$ and σ_{p+1} generated by (19)–(21) are consistent with their definitions (12)(13)(14). We prove this for two cases $\sigma_{p+1} = m + 1$ and $\sigma_{p+1} = m$.

Case 1: $\sigma_{p+1} = m + 1$. We first show

$$\sigma_{i,p+1} \leq m, \text{ if } \sigma_{i,p} < m \quad (29)$$

for the following two cases: 1) $\sigma_{i,p} < m$ and $\sigma_{j,p} < m \forall j \in N_i(p)$. For this case by (6) we have $\hat{\sigma}_{i,p} < m$, and hence $\sigma_{i,p+1} \leq \hat{\sigma}_{i,p} + 1 \leq m$ by (5). 2) $\sigma_{i,p} < m$ and $\sigma_{j,p} = m$ for some $j \in N_i(p)$. For this case we derive $\hat{\sigma}_{i,p} = m$, $x'_{i,p+1} = h_p(x^*)$ by (3)(6). So, by (5) we derive $\sigma_{i,p+1} = \hat{\sigma}_{i,p} = m$. Thus, $\sigma_{i,p+1} \leq m$ when $\sigma_{i,p} < m$. Hence (29) holds. Furthermore, this means that

$$\sigma_{i,p+1} = m + 1 \text{ only if } \sigma_{i,p} = m. \quad (30)$$

Since we are considering the case where $\sigma_{p+1} = m + 1$, by definition we know that there exists some agent $i_0 \in \mathcal{V}$ such that $\sigma_{i_0,p+1} = m + 1$. Then $\sigma_{i_0,p} = m$ by (30), and hence $\hat{\sigma}_{i_0,p} = m$ from (25). Then from $\sigma_{i_0,p+1} = m + 1$ by (5) we know that $\|x'_{i_0,p+1} - h_p(x^*)\| > M_m$. So from (20)(21) we derive $\tilde{x}_{i_0,p+1} = h_p(x^*) \forall i \in \mathcal{V}$ and $\sigma_{p+1} = m + 1$, which is consistent with the σ_{p+1} defined by (12). Since $\sigma_{p+1} = m + 1$ and $p \in [\tau_m, \tau_{m+1})$, by (18) we see that $\tilde{x}_{i,p+1} = h_p(x^*) \forall i \in \mathcal{V}$, which is consistent with that generated by (19)–(21).

Case 2: $\sigma_{p+1} = m$. In this case $\sigma_{i,p+1} \leq m \forall i \in \mathcal{V}$. By (25), from (4)(5) we see that

$$\|x'_{i,p+1} - h_p(x^*)\| \leq M_m, \quad x_{i,p+1} = x'_{i,p+1} \quad \forall i : \sigma_{i,p} = m. \quad (31)$$

So, by (28) we derive

$$\|\hat{x}_{i,p+1} - h_p(x^*)\| \leq M_m, \quad \forall i : \sigma_{i,p} = m. \quad (32)$$

From (24) we have $\|\hat{x}_{i,p+1} - h_p(x^*)\| = 0 \leq M_m \forall i : \sigma_{i,p} < m$, which incorporating with (32) yields $\|\hat{x}_{i,p+1} - h_p(x^*)\| \leq M_m \forall i \in \mathcal{V}$. Then from (20) it follows

$$\tilde{x}_{i,p+1} = \hat{x}_{i,p+1} \forall i \in \mathcal{V}, \quad \sigma_{p+1} = m. \quad (33)$$

which means that σ_{p+1} is consistent with that defined by (12).

It remains to show that $\tilde{x}_{i,p+1}$ generated by (19)–(21) is consistent with that defined by (13)(14). We consider two cases: 1) $\sigma_{i,p} = m$. For this case, by (25)(31)(32) we see $\tilde{x}_{i,p+1} = x_{i,p+1} \forall i : \sigma_{i,p} = m$. By $\sigma_{p+1} = m$ we see $p \in [\tau_m, \tau_{m+1})$, and hence $\tilde{x}_{i,p+1} = x_{i,p+1}$ by (16). So the consistency assertion holds for any i with $\sigma_{i,p} = m$. 2) $\sigma_{i,p} < m$. For this case, from $\sigma_{p+1} = m$ we see $p+1 \in [\tau_m, \tau_{m+1})$, and hence by $\sigma_{i,p} < m$ and (17) we know $\tilde{x}_{i,p+1}$ defined by (13)(14) is equal to $h_p(x^*)$. By (24)(33) we derive $\tilde{x}_{i,p+1} = h_p(x^*)$. So the consistency assertion holds for i with $\sigma_{i,p} < m$ too.

In summary, $\tilde{x}_{i,p+1}$ and σ_{p+1} generated by (19)–(21) are consistent with their definitions (12)(13)(14). So the induction is complete. \square

Lemma 4 Assume A4 holds. Then

i)

$$\sigma_{j,k} + Bd_{ij} \geq \sigma_{i,k} \quad \forall j \in \mathcal{V} \forall k > 0, \quad (34)$$

where d_{ij} is the length of the shortest directed path from i to j in \mathcal{G}_∞ , and B is the positive integer given in A4 d).

ii)

$$\tilde{\tau}_{j,m} \leq \tau_m + BD \quad \forall j \in \mathcal{V} \text{ for } m > 1, \quad (35)$$

where $D \triangleq \max_{i,j \in \mathcal{V}} d_{ij}$.

Proof i) Since \mathcal{G}_∞ is strongly connected by A4 c), for any $j \in \mathcal{V}$ there exists a sequence of nodes $i_1, i_2, \dots, i_{d_{ij}-1}$ such that $(i, i_1) \in \mathcal{E}_\infty, (i_1, i_2) \in \mathcal{E}_\infty, \dots, (i_{d_{ij}-1}, j) \in \mathcal{E}_\infty$.

Noticing that $(i, i_1) \in \mathcal{E}_\infty$, by A4 d) we have

$$(i, i_1) \in \mathcal{E}(k) \cup \mathcal{E}(k+1) \cup \dots \cup \mathcal{E}(k+B-1).$$

Therefore, there exists a positive integer $k' \in [k, k+B-1]$ such that $(i, i_1) \in \mathcal{E}(k')$. So, $i \in N_{i_1}(k')$, and hence by (6) and (5) we have

$$\sigma_{i_1, k+B} \geq \sigma_{i_1, k'+1} \geq \hat{\sigma}_{i_1, k'} \geq \sigma_{i, k'} \geq \sigma_{i, k}.$$

Repeat this procedure, we can obtain $\sigma_{i_2, k+2B} \geq \sigma_{i_1, k+B} \geq \sigma_{i, k}$. Finally we can reach (34).

ii) For some $m \geq 1$, let $\tau_m = k_1$. Then there exists an i such that $\tau_{i,m} = k_1$. By (34) we have $\sigma_{j, k_1+Bd_{ij}} \geq \sigma_{i, k_1} = m \forall j \in \mathcal{V}$.

For the case where $\sigma_{j, k_1+Bd_{ij}} = m \forall j \in \mathcal{V}$, we have $\tau_{j,m} \leq k_1 + Bd_{ij} \forall j \in \mathcal{V}$. By noticing $\tau_m = k_1$, by the definition of $\tilde{\tau}_{j,m}$ we have (35):

$$\tilde{\tau}_{j,m} \leq \tau_{j,m} \leq \tau_m + Bd_{ij} \leq \tau_m + BD \quad j \in \mathcal{V}.$$

For the case where $\sigma_{j, k_1+Bd_{ij}} > m$ for some $j \in \mathcal{V}$, we have $\tau_{m+1} \leq k_1 + Bd_{ij}$ for some $j \in \mathcal{V}$, and hence $\tau_{m+1} \leq \tau_m + BD$. Again, we obtain (35):

$$\tilde{\tau}_{j,m} \leq \tau_{m+1} \leq \tau_m + BD \quad j \in \mathcal{V}.$$

\square

Corollary 1 If $\sigma_k \xrightarrow[k \rightarrow \infty]{\infty}$, then $\lim_{k \rightarrow \infty} \sigma_{i,k} = \infty \forall i \in \mathcal{V}$.

This corollary can be easily obtained from (34).

Lemma 5 Assume A5 holds at the sample path ω under consideration, A1, A3, A6, and A7 hold. Then for this ω

$$\lim_{T \rightarrow \infty} \limsup_{k \rightarrow \infty} \frac{1}{T} \left\| \sum_{s=n_k}^{m(n_k, t_k) \wedge (\tau_{\sigma_{n_k}+1}-1)} a_s \tilde{\epsilon}_{s+1} \right\| = 0 \quad \forall t_k \in [0, T] \quad (36)$$

along indices $\{n_k\}$ whenever $\{\tilde{\Lambda}_{n_k}\}$ converges at ω , where $\tilde{\epsilon}_k \triangleq \text{col}\{\tilde{\epsilon}_{1,k}, \dots, \tilde{\epsilon}_{N,k}\}$, $\tilde{X}_k \triangleq \text{col}\{\tilde{x}_{1,k}, \dots, \tilde{x}_{N,k}\}$, and $\tilde{\Lambda}_k \triangleq \tilde{X}_k - \Theta_k$.

Proof It suffices to show

$$\lim_{T \rightarrow \infty} \limsup_{k \rightarrow \infty} \frac{1}{T} \left\| \sum_{s=n_k}^{m(n_k, t_k) \wedge (\tau_{\sigma_{n_k}+1}-1)} a_s \tilde{\epsilon}_{i, s+1} \right\| = 0 \quad \forall t_k \in [0, T] \text{ for sufficiently large } K > 0 \quad (37)$$

along indices $\{n_k\}$ whenever $\{\tilde{x}_{i, n_k} - \theta_{n_k}\}$ converges at sample ω where A5 b) holds for agent i .

We consider two cases:

Case 1: $\lim_{k \rightarrow \infty} \sigma_k = \sigma < \infty$. From definition we can obtain

$$\tau_{\sigma+1} = \infty \text{ when } \lim_{k \rightarrow \infty} \sigma_k = \sigma. \quad (38)$$

By (35) we have $\tilde{\tau}_{i, \sigma} \leq \tau_\sigma + BD$, hence by (14)(38)

$$\tilde{x}_{i, k} = x_{i, k}, \quad \tilde{\epsilon}_{i, k+1} = \epsilon_{i, k+1} \quad \forall k \geq \tau_\sigma + BD. \quad (39)$$

So,

$$\left\| \sum_{s=k}^{m(k, t) \wedge (\tau_{\sigma_k+1}-1)} a_s \tilde{\epsilon}_{i, s+1} \right\| = \left\| \sum_{s=k}^{m(k, t)} a_s \epsilon_{i, s+1} \right\|$$

for any $t > 0$ and any sufficiently large k . Then by A5 b) we prove (37).

Case 2: $\lim_{k \rightarrow \infty} \sigma_k = \infty$. In this case we prove (37) for three separate cases:

i): $\tilde{\tau}_{i, \sigma_{n_p}} \leq n_p$. For this case, $[n_p, \tau_{\sigma_{n_p}+1}) \subset [\tau_{i, \sigma_{n_p}}, \tau_{\sigma_{n_p}+1})$. So by (14) we have

$$\tilde{x}_{i, s} = x_{i, s}, \quad \tilde{\epsilon}_{i, s+1} = \epsilon_{i, s+1} \quad \forall n_p \leq s \leq \tau_{\sigma_{n_p}+1}. \quad (40)$$

Thus, for any $t_p \in [0, T]$

$$\begin{aligned} & \left\| \sum_{s=n_p}^{m(n_p, t_p) \wedge (\tau_{\sigma_{n_p+1}} - 1)} a_s \tilde{\epsilon}_{i, s+1} \right\| \\ &= \left\| \sum_{s=n_p}^{m(n_p, t_p) \wedge (\tau_{\sigma_{n_p+1}} - 1)} a_s \epsilon_{i, s+1} \right\|. \end{aligned} \tag{41}$$

Notice that by (40), $\tilde{x}_{i, n_p} = x_{i, n_p}$, and the indices $\{n_p\}$ is taken when $\{\tilde{x}_{i, n_p} - \theta_{n_p}\}$ is a convergent subsequence. So $\{x_{i, n_p} - \theta_{n_p}\}$ is a convergent subsequence as well. It can be seen $\sum_{s=n_p}^{m(n_p, t_p) \wedge (\tau_{\sigma_{n_p+1}} - 1)} a_s \leq \sum_{s=n_p}^{m(n_p, t_p)} a_s \leq t_p \leq T$. Hence, from (41) and A5 b) we conclude (37).

ii): $\tilde{\tau}_{i, \sigma_{n_p}} > n_p$ and $\tilde{\tau}_{i, \sigma_{n_k}} = \tau_{\sigma_{n_k+1}}$. By definition of τ_k and σ_k we have $\tau_{\sigma_k} \leq k$, and hence $\tau_{\sigma_{n_p}} \leq n_p$. Then $[n_p, \tau_{\sigma_{n_p+1}}) \subset [n_p, \tau_{i, \sigma_{n_p}})$, and hence by (13) we have

$$\begin{aligned} \tilde{x}_{i, s} &= h_{s-1}(x^*), \quad \tilde{\epsilon}_{i, s+1} = -f_{i, s+1}(h_s(x^*)) \\ & \quad \forall s : n_p \leq s < \tau_{\sigma_{n_p+1}}. \end{aligned} \tag{42}$$

From $\tilde{\tau}_{i, \sigma_{n_p}} = \tau_{\sigma_{n_p+1}}$ by (35) we know that $\tau_{\sigma_{n_p+1}} \leq \tau_{\sigma_{n_p}} + BD \leq n_p + BD$. Then for any $t_p \in [0, T]$, utilizing A1, A3 and Lemma 1 we have

$$\begin{aligned} & \left\| \sum_{s=n_p}^{m(n_p, t_p) \wedge (\tau_{\sigma_{n_p+1}} - 1)} a_s \tilde{\epsilon}_{i, s+1} \right\| \\ & \leq \sum_{s=n_p}^{n_p+BD} a_s \|f_{i, s+1}(h_s(x^*))\| \\ & = \sum_{s=n_p}^{n_p+BD} a_s \|f_{i, s+1}(\theta_{s+1} + h_s(x^*) - \theta_{s+1})\| \\ & \leq \sum_{s=n_p}^{n_p+BD} a_s \alpha(\eta) \\ & \leq BD \cdot a_{n_p} \cdot \alpha(\eta) \xrightarrow{p \rightarrow \infty} 0, \end{aligned} \tag{43}$$

and hence (37) holds for this case.

iii): $\tilde{\tau}_{i, \sigma_{n_p}} > n_p$ and $\tilde{\tau}_{i, \sigma_{n_p}} < \tau_{\sigma_{n_p+1}}$. For this case from definition we know that $\tilde{\tau}_{i, \sigma_{n_p}} = \tau_{i, \sigma_{n_p}}$. So by (35) we have $\tau_{i, \sigma_{n_p}} \leq \tau_{\sigma_{n_p}} + BD$. Noticing $\tau_{\sigma_{n_p}} \leq n_p$, we conclude that

$$\tau_{\sigma_{n_p}} \leq n_p < \tilde{\tau}_{i, \sigma_{n_p}} = \tau_{i, \sigma_{n_p}} \leq n_p + BD \tag{44}$$

So, $[n_p, \tau_{i, \sigma_{n_p}}) \subset [n_p, \tau_{\sigma_{n_p+1}})$. From this and $\tilde{\tau}_{i, \sigma_{n_p}} = \tau_{i, \sigma_{n_p}}$, by (13)(14) we derive

$$\begin{aligned} \tilde{x}_{i, s} &= h_{s-1}(x^*), \quad \tilde{\epsilon}_{i, s+1} = -f_{i, s+1}(h_s(x^*)), \quad \forall n_p \leq s < \tau_{i, \sigma_{n_p}}, \\ \tilde{x}_{i, s} &= x_{i, s}, \quad \tilde{\epsilon}_{i, s} = \epsilon_{i, s}, \quad \forall \tau_{i, \sigma_{n_p}} \leq s < \tau_{\sigma_{n_p+1}}. \end{aligned}$$

Consequently, for any $t_p \in [0, T]$

$$\begin{aligned} & \left\| \sum_{s=n_p}^{m(n_p, t_p) \wedge (\tau_{\sigma_{n_p+1}} - 1)} a_s \tilde{\epsilon}_{i, s+1} \right\| \\ & \leq \left\| \sum_{s=n_p}^{m(n_p, t_p) \wedge (\tau_{\sigma_{n_p+1}} - 1)} a_s f_{i, s+1}(h_s(x^*)) \mathbb{I}_{[n_p \leq s < \tau_{i, \sigma_{n_p}}]} \right\| \\ & \quad + \left\| \sum_{s=\tau_{i, \sigma_{n_p}}}^{m(n_p, t_p) \wedge (\tau_{\sigma_{n_p+1}} - 1)} a_s \epsilon_{i, s+1} \right\| \end{aligned} \tag{45}$$

Analyze the first term at the right hand of (45):

$$\begin{aligned} & \left\| \sum_{s=n_p}^{m(n_p, t_p) \wedge (\tau_{\sigma_{n_p+1}} - 1)} a_s f_{i, s+1}(h_s(x^*)) \mathbb{I}_{[n_p \leq s < \tau_{i, \sigma_{n_p}}]} \right\| \\ & \leq \sum_{s=n_p}^{\tau_{i, \sigma_{n_p}}} a_s \|f_{i, s+1}(h_s(x^*))\| \\ & \leq \sum_{s=n_p}^{n_p+BD} a_s \alpha(\eta) \xrightarrow{p \rightarrow \infty} 0 \end{aligned}$$

From the definition of $\tau_{i, k}$, the truncation number of agent i at time $\tau_{i, \sigma_{n_p}}$ is σ_{n_p} while it's smaller than σ_{n_p} at time $\tau_{i, \sigma_{n_p}} - 1$. So by the algorithm (3)-(6) we know $x_{i, \tau_{i, \sigma_{n_p}}} = h_{\tau_{i, \sigma_{n_p}} - 1}(x^*)$. Notice the second term at the right hand of (45). If we can show $\{h_{\tau_{i, \sigma_{n_p}} - 1}(x^*) - \theta_{\tau_{i, \sigma_{n_p}}}\}$ is convergent, combining it with the fact $\sum_{s=n_p}^{m(n_p, t_p) \wedge (\tau_{\sigma_{n_p+1}} - 1)} a_s \leq \sum_{s=n_p}^{m(n_p, t_p)} a_s \leq t_p$, from A5 we can conclude that the second term at the right hand of (45) tends to zero as $p \rightarrow \infty$.

We show that $\{h_k(x^*) - \theta_{k+1}\}$ is a convergent sequence by proving that the sequence is a Cauchy sequence. For two different integer $j > i > 0$, we see

$$\begin{aligned} & \|h_j(x^*) - \theta_{j+1} - h_i(x^*) + \theta_{i+1}\| \\ & \leq \|h_j(x^*) - \theta_{j+1} - h_{j-1}(x^*) + \theta_j \\ & \quad + h_{j-1}(x^*) - \theta_j - h_{j-2}(x^*) + \theta_{j-1} + \dots + h_{i+1}(x^*) \\ & \quad - \theta_{i+2} \\ & \quad - h_i(x^*) + \theta_{i+1}\| \\ & = \|g_j(h_{j-1}(x^*)) - g_j(\theta_j) - h_{j-1}(x^*) + \theta_j - \xi_{j+1} + \dots\| \\ & \leq \sum_{l=i+1}^j \|d_l(h_{l-1}(x^*))\| + \sum_{l=i+2}^{j+1} \|\xi_l\| \\ & \leq \sum_{l=i+1}^j \gamma_l \cdot \eta + \sum_{l=i+2}^{j+1} \|\xi_l\|, \end{aligned}$$

where the last inequality comes from A6 and Lemma 1. By A6 and A7 we know that $\sum_{k=1}^{\infty} \gamma_k < \infty$ and $\sum_{k=1}^{\infty} \xi_k < \infty$

∞ . Thus, for $\forall \epsilon > 0$, we can find a sufficiently large $N > 0$ such that $\|h_j(x^*) - \theta_{j+1} - h_i(x^*) + \theta_{i+1}\| < \epsilon$ for $\forall i, j > N$, which means that $\{h_k(x^*) - \theta_{k+1}\}$ is a Cauchy sequence. Furthermore, we can prove that (37) holds for case iii).

Since one of case i), ii), iii) must take place for the case $\lim_{k \rightarrow \infty} \sigma_k = \infty$, we can conclude that (37) holds in Case 2.

Combining Case 1 and Case 2, we conclude (36). \square

Corollary 2 In (38) we show that $\tau_{\sigma+1} = \infty$ when $\lim_{k \rightarrow \infty} \sigma_k = \sigma < \infty$. So, if $\lim_{k \rightarrow \infty} \sigma_k = \sigma < \infty$, by (20) we know that $\{\tilde{x}_{i,k}\}$ and $\{x_{i,k}\}$, $\{\tilde{\epsilon}_{i,k}\}$ and $\{\epsilon_{i,k}\}$ coincide in a finite number of steps.

5 Proof of the main result

Define:

$$\Psi(k, s) \triangleq [D_{\perp}(W(k) \otimes I_l)] [D_{\perp}(W(k-1) \otimes I_l)] \cdots [D_{\perp}(W(s) \otimes I_l)] \quad \forall k \geq s,$$

and

$$\Psi(k-1, k) \triangleq I_{Nl}.$$

Since $W(k)$ are doubly stochastic, by the property of Kronecker product $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$ we know that for $\forall k \geq s-1$

$$\Psi(k, s) = (\Phi(k, s) - \frac{1}{N} \mathbf{1}\mathbf{1}^T) \otimes I_l, \tag{46}$$

$$\Psi(k, s) D_{\perp} = (\Phi(k, s) - \frac{1}{N} \mathbf{1}\mathbf{1}^T) \otimes I_l. \tag{47}$$

The following lemma characterizes the closeness of the auxiliary sequence $\{\tilde{\Lambda}_k\}_{k \geq 1}$ along its convergent subsequence $\{\tilde{\Lambda}_{n_k}\}$.

Lemma 6 Assume A1, A3, A4, and A6 hold. Further, for a fixed sample ω , assume A5 and A7 hold for all the agents. Let $\{\tilde{\Lambda}_{n_k}\}$ be a convergent subsequence of $\{\tilde{\Lambda}_k\}$ for ω . $\tilde{\Lambda}_{n_k} \xrightarrow[k \rightarrow \infty]{} \tilde{\Lambda}$. Then for this ω there is a $T > 0$ such that for all sufficiently large k and any $T_k \in [0, T]$

$$\tilde{X}_{m+1} = (W(m) \otimes I_l) G_m(\tilde{X}_m) + a_m (F_{m+1}(G(\tilde{X}_m)) + \tilde{\epsilon}_{m+1}) \tag{48}$$

for any $m = n_k, \dots, m(n_k, T_k)$, and

$$\|\tilde{\Lambda}_{m+1} - \tilde{\Lambda}_{n_k}\| \leq c_1 T_k + M'_0 \tag{49}$$

$$\|\tilde{\Delta}_{m+1} - \tilde{\Delta}_{n_k}\| \leq c_2 T_k, \quad \forall n_k \leq m \leq m(n_k, T_k), \tag{50}$$

where $\tilde{x}_k \triangleq \frac{1}{N} \sum_{k=1}^N \tilde{x}_{i,k}$, $\tilde{\Delta}_k \triangleq \tilde{x}_k - \theta_k$, and $\tilde{\Delta}_{i,k} \triangleq \tilde{x}_{i,k} - \theta_k$.

Proof Consider a fixed sample path ω where A5 and A7 hold.

Let $C > \|\tilde{\Lambda}\|$. There exists an integer $k_C > 0$ such that

$$\begin{aligned} \|\tilde{\Lambda}_{n_k}\| &\leq C, \quad \gamma_k < a_k, \\ \|\xi_{k+1}\| &< a_k, \quad a_k < 1, \quad \forall k \geq k_C \end{aligned} \tag{51}$$

From Lemma 5 we know that there exist constants $T_1 > 0$ and $k_0 > k_C$ such that

$$\left\| \sum_{s=n_k}^{m(n_k, T_k) \wedge (\tau_{\sigma_{n_k+1}-1})} a_s \tilde{\epsilon}_{s+1} \right\| \leq T_0 \tag{52}$$

$$\forall T_k \in [0, T_k], \quad \forall T_0 \in [0, T_1], \quad \forall k \geq k_0.$$

Define

$$M'_0 \triangleq C(c\rho + 2) + 1, \tag{53}$$

$$c_1 \triangleq \sqrt{N} \cdot c_2 + 2 + \frac{c(1+\rho)}{1-\rho}, \tag{54}$$

$$c_2 \triangleq M'_0 + C + 2 + \alpha(2M'_0 + 2C + 3) + \frac{1}{\sqrt{N}}, \tag{55}$$

where c and ρ are given by (9). Select T such that

$$0 < T \leq T_1, \quad c_1 T < 1. \tag{56}$$

For any $k \geq k_0$ and any $T_k \in [0, T]$ define

$$\begin{aligned} s_k &\triangleq \sup\{s \geq n_k : \|\tilde{\Lambda}_j - \tilde{\Lambda}_{n_k}\| \\ &\leq c_1 T_k + M'_0 \quad \forall n_k \leq j \leq s\} \end{aligned} \tag{57}$$

So from (51) and (56) it follows that

$$\|\tilde{\Lambda}_j\| \leq M'_0 + C + 1, \quad \forall n_k \leq j \leq s_k. \tag{58}$$

We intend to prove $s_k > m(n_k, T_k)$. Assume the converse that for sufficiently large $k \geq k_0$ and any $T_k \in [0, T]$

$$s_k \leq m(n_k, T_k). \tag{59}$$

We first show that there exists a positive integer $k_1 > k_0$ such that for any $k \geq k_1$

$$s_k < \tau_{\sigma_{n_k+1}}, \quad \forall k \geq k_1, \quad \forall T_k \in [0, T]. \tag{60}$$

We prove (60) for two cases: $\lim_{k \rightarrow \infty} \sigma_k = \infty$ and $\lim_{k \rightarrow \infty} \sigma_k = \sigma < \infty$.

i) $\lim_{k \rightarrow \infty} \sigma_k = \infty$: From (58) we know that $\|\tilde{x}_{i,n_k} - \theta_{n_k}\| \leq M'_0 + C + 1 \quad \forall i \in \mathcal{V}$. First, we prove that for sufficiently large k , truncation does not happen at time $n_k + 1$. For any $i \in \mathcal{V}$, we consider the following two cases:

a) \tilde{x}_{i,n_k} and $\tilde{\epsilon}_{i,n_k+1}$ take value as (13): From (19) we have

$$\begin{aligned} \hat{x}_{i,n_k+1} &= \sum_{j \in N_i(n_k)} w_{ij}(n_k) g_{n_k}(\tilde{x}_{j,n_k}) \\ &= \sum_{j \in N_i(n_k)} w_{ij}(n_k) (g_{n_k}(\tilde{x}_{j,n_k}) - g_{n_k}(\theta_{n_k}) - (\tilde{x}_{j,n_k} - \theta_{n_k})) \\ &\quad + \sum_{j \in N_i(n_k)} w_{ij}(n_k) (\theta_{n_k+1} - \xi_{n_k+1} + (\tilde{x}_{j,n_k} - \theta_{n_k})). \end{aligned}$$

Since A4 indicates that $W(n_k)$ is doubly stochastic, by A6, (51) and direct calculation we have the following inequalities

$$\begin{aligned} \|\hat{x}_{i,n_k+1} - \theta_{n_k+1}\| &\leq (\gamma_{n_k} + 1)(M'_0 + C + 1) + \|\xi_{n_k+1}\| \\ &\leq 2M'_0 + 2C + 3, \end{aligned}$$

and hence by Lemma 1 we know $\|\hat{x}_{i,n_k+1} - h_{n_k}(x^*)\| \leq \eta + 2M'_0 + 2C + 3$.

b) \tilde{x}_{i,n_k} and $\tilde{\epsilon}_{i,n_k+1}$ take value as (14): From (19) we have

$$\begin{aligned} \hat{x}_{i,n_k+1} &= \sum_{j \in N_i(n_k)} w_{ij}(n_k) \left(g_{n_k}(\tilde{x}_{j,n_k}) - g_{n_k}(\theta_{n_k}) - (\tilde{x}_{j,n_k} - \theta_{n_k}) \right) \\ &\quad + \sum_{j \in N_i(n_k)} w_{ij}(n_k) \left(\theta_{n_k+1} - \xi_{n_k+1} + (\tilde{x}_{j,n_k} - \theta_{n_k}) \right) \\ &\quad + a_{n_k} f_{i,n_k+1} \left(\theta_{n_k+1} + g_{n_k}(\tilde{x}_{i,n_k}) - g_{n_k}(\theta_{n_k}) \right. \\ &\quad \left. - (\tilde{x}_{i,n_k} - \theta_{n_k}) + (\tilde{x}_{i,n_k} - \theta_{n_k}) - \xi_{n_k+1} \right) \\ &\quad + a_{n_k} \epsilon_{i,n_k+1}. \end{aligned}$$

By A5 a) we know that $a_{n_k} \epsilon_{i,n_k+1} < 1$ for sufficiently large k . Then, by A3, A4, A6 and (51), we have the following inequalities

$$\begin{aligned} \|\hat{x}_{i,n_k+1} - \theta_{n_k+1}\| &\leq 2M'_0 + 2C + 3 + a_{n_k} \alpha (2M'_0 + 2C + 3) \\ &\quad + a_{n_k} \epsilon_{i,n_k+1} \\ &\leq 2M'_0 + 2C + 4 + \alpha (2M'_0 + 2C + 3) \\ &\triangleq M_1, \end{aligned}$$

and hence by Lemma 1 we know $\|\hat{x}_{i,n_k+1} - h_{n_k}(x^*)\| \leq \eta + M_1$.

So we show that when $\|\tilde{\Lambda}_{n_k}\| \leq M'_0 + C + 1$, we have $\|\hat{x}_{i,n_k+1} - h_{n_k}(x^*)\| \leq \eta + M_1$. Since $\{M_k\}$ is a sequence of positive number increasingly diverging to infinity, there exists a positive integer $k_1 > k_0$ such that $M_{\sigma_{n_k}} > \eta + M_1$ for all $k \geq k_1$. Thus, we prove that truncation does not happen at time $n_k + 1$.

Notice (58) holds for $j : n_k \leq j \leq s_k$. So, similar to the proof above, we can prove that truncation does not happen for time $n_k + 1, \dots, s_k + 1$. Then we conclude $s_k < \tau_{\sigma_{n_k}+1}$.

ii) $\lim_{k \rightarrow \infty} \sigma_k = \sigma < \infty$: For this case there exists a positive integer $k_1 > k_0$ such that $\sigma_{n_k} = \sigma$ for all $k \geq k_1$, and hence $\tau_{\sigma_{n_k}+1} = \infty$ by definition. Then $m(n_k, T_k) < \tau_{\sigma_{n_k}+1}$, hence by (59) we know $s_k < \tau_{\sigma_{n_k}+1}$. So (60) is proven.

By (56) we see $T_k \in [0, T] \subset [0, T_1]$, then from (52) we know that for sufficiently large $k > k_1$ and any $T_k \in [0, T]$

$$\left\| \sum_{s=n_k}^{m(n_k, t_k) \wedge (\tau_{\sigma_{n_k}+1})} a_s \tilde{\epsilon}_{s+1} \right\| \leq T_k \quad \forall t_k \in [0, T_k]. \quad (61)$$

By setting $t_k = \sum_{m=n_k}^s a_m$ for some $s \in [n_k, s_k]$, from (59) we see $\sum_{m=n_k}^s a_m \leq \sum_{m=n_k}^{s_k} a_m \leq T_k$. Noticing $m(n_k, t_k) = s$, from (60) we derive $m(n_k, t_k) \wedge (\tau_{\sigma_{n_k}+1} -$

$1) = s$. So by (61) we know that

$$\left\| \sum_{m=n_k}^s a_m \tilde{\epsilon}_{m+1} \right\| \leq T_k \quad \forall s : n_k \leq s \leq s_k \quad (62)$$

for sufficiently large $k \geq k_1$ and any $T_k \in [0, T]$.

Now we consider the following recursive algorithm starting from n_k :

$$\begin{aligned} Z_{m+1} &= (W(m) \otimes \mathbf{I}_l) G_m(Z_m) + a_m (F_{m+1}(G_m(Z_m)) + \tilde{\epsilon}_{m+1}), \\ Z_{n_k} &= \tilde{X}_{n_k}. \end{aligned} \quad (63)$$

where $Z_k \triangleq \text{col}\{z_{i,k}, \dots, z_{N,k}\}$. By (60) we know that (48) holds for $m = n_k, \dots, s_k - 1$ for $\forall k \geq k_1 \forall T_k \in [0, T]$. Then we derive

$$Z_m = \tilde{X}_m \quad \forall m : n_k \leq m \leq s_k \quad (64)$$

Set $z_k = \frac{\mathbf{1}^T \otimes \mathbf{I}_l}{N} Z_k$, $\hat{\Delta}_{i,k} \triangleq z_{i,k} - \theta_k$, $\hat{\Delta}_k \triangleq z_k - \theta_k$, and $\hat{\Lambda}_k \triangleq Z_k - \Theta_k$. By multiplying both sides of (63) with $\frac{\mathbf{1}^T}{N} (\mathbf{1}^T \otimes \mathbf{I}_l)$, from $\mathbf{1}^T W(m) = \mathbf{1}^T$ and $(A \otimes B)(C \otimes D) = AB \otimes CD$ we derive

$$z_{s+1} = \frac{1}{N} \sum_{i=1}^N g_s(z_{i,s}) + \frac{\mathbf{1}^T \otimes \mathbf{I}_l}{N} a_s (F_{s+1}(G_s(Z_s)) + \tilde{\epsilon}_{s+1})$$

and hence

$$\begin{aligned} z_{s+1} &= \frac{1}{N} \sum_{i=1}^N g_s(z_{i,s}) - \frac{1}{N} \sum_{i=1}^N g_s(\theta_s) - \xi_{s+1} + \theta_{s+1} \\ &\quad + \frac{\mathbf{1}^T \otimes \mathbf{I}_l}{N} a_s F_{s+1}(G_s(Z_s) - G_s(\Theta_s) - \Xi_{s+1} + \Theta_{s+1}) \\ &\quad + \frac{\mathbf{1}^T \otimes \mathbf{I}_l}{N} a_s \tilde{\epsilon}_{s+1} \\ &= \frac{1}{N} \sum_{i=1}^N d_s(z_{i,s}) + \hat{\Delta}_s - \xi_{s+1} + \theta_{s+1} \\ &\quad + \frac{\mathbf{1}^T \otimes \mathbf{I}_l}{N} a_s F_{s+1}(\Theta_{s+1} + D_s(Z_s) + \hat{\Lambda}_s - \Xi_{s+1}) \\ &\quad + \frac{\mathbf{1}^T \otimes \mathbf{I}_l}{N} a_s \tilde{\epsilon}_{s+1}. \end{aligned}$$

So

$$\begin{aligned}
 \|\widehat{\Delta}_{s+1} - \widehat{\Delta}_{n_k}\| &\leq \left\| \sum_{j=n_k}^s \frac{1}{N} \sum_{i=1}^N d_j(z_{i,j}) \right\| + \left\| \sum_{j=n_k}^s \xi_{j+1} \right\| \\
 &+ \left\| \frac{\mathbf{1}^T \otimes \mathbf{I}_l}{N} \sum_{j=n_k}^s a_j F_{j+1}(\Theta_{j+1} + D_j(Z_j) + \widehat{\Lambda}_j - \Xi_{j+1}) \right\| \\
 &+ \left\| \frac{\mathbf{1}^T \otimes \mathbf{I}_l}{N} \sum_{j=n_k}^s a_j \tilde{\epsilon}_{j+1} \right\| \\
 &\leq \frac{1}{N} \sum_{j=n_k}^s \sum_{i=1}^N \gamma_j \|z_{i,j} - \theta_j\| + \left\| \sum_{j=n_k}^s \xi_{j+1} \right\| \\
 &+ \frac{1}{N} \sum_{j=n_k}^s a_j \sum_{i=1}^N \|f_{i,j+1}(\theta_{j+1} + d_j(z_{i,j}) + \widehat{\Delta}_{i,j} - \xi_{j+1})\| \\
 &+ \frac{1}{\sqrt{N}} \left\| \sum_{j=n_k}^s a_j \tilde{\epsilon}_{j+1} \right\| \\
 &\leq \sum_{j=n_k}^s \gamma_j \cdot (M'_0 + C + 1) + \left\| \sum_{j=n_k}^s \xi_{j+1} \right\| \\
 &+ \sum_{j=n_k}^s a_j \alpha \left((1 + \gamma_j)(M'_0 + C + 1) + \|\xi_{j+1}\| \right) \\
 &+ \frac{1}{\sqrt{N}} T_k \\
 &\leq \left(M'_0 + C + 1 + 1 + \alpha(2M'_0 + 2C + 3) + \frac{1}{\sqrt{N}} \right) T_k \\
 &= c_2 T_k, \quad \forall s: n_k \leq s \leq s_k, \tag{65}
 \end{aligned}$$

where the second inequality comes from A6, the third inequality comes from (58) A3 and A5, the fourth inequality comes from (51) (59), and the last inequality comes from (59).

Denote by $Z_{\perp,s} = D_{\perp} Z_s$ the disagreement vector of Z_s . By multiplying both sides of (63) with D_{\perp} we have

$$\begin{aligned}
 Z_{\perp,m+1} &= D_{\perp}(W(m) \otimes \mathbf{I}_l)G_m(Z_m) + \\
 &a_m D_{\perp} \left(F_{m+1}(G_m(Z_m)) + \tilde{\epsilon}_{m+1} \right)
 \end{aligned}$$

Notice that $D_{\perp}(W(m) \otimes \mathbf{I}_l) = D_{\perp}(W(m) \otimes \mathbf{I}_l)D_{\perp}$, so we have

$$\begin{aligned}
 Z_{\perp,m+1} &= D_{\perp}(W(m) \otimes \mathbf{I}_l)D_{\perp}Z_m \\
 &+ D_{\perp}(W(m) \otimes \mathbf{I}_l)(G_m(Z_m) - Z_m) \\
 &+ a_m D_{\perp} \left(F_{m+1}(G_m(Z_m)) + \tilde{\epsilon}_{m+1} \right)
 \end{aligned}$$

By definition we know that $D_{\perp}G_m(\Theta_m) = D_{\perp}\Theta_m = 0$, hence we have

$$\begin{aligned}
 Z_{\perp,m+1} &= D_{\perp}(W(m) \otimes \mathbf{I}_l)Z_{\perp,m} \\
 &+ D_{\perp}(W(m) \otimes \mathbf{I}_l)(G_m(Z_m) - G_m(\Theta_m) - Z_m + \Theta_m) \\
 &+ a_m D_{\perp} \left(F_{m+1}(G_m(Z_m)) + \tilde{\epsilon}_{m+1} \right) \\
 &= D_{\perp}(W(m) \otimes \mathbf{I}_l)Z_{\perp,m} + D_{\perp}(W(m) \otimes \mathbf{I}_l)D_m(Z_m) \\
 &+ a_m D_{\perp} \left(F_{m+1}(G_m(Z_m)) + \tilde{\epsilon}_{m+1} \right).
 \end{aligned}$$

So inductively

$$\begin{aligned}
 Z_{\perp,s+1} &= \Psi(s, n_k)Z_{n_k} + \sum_{m=n_k}^s \Psi(s, n_k)D_m(Z_m) \\
 &+ \sum_{m=n_k}^s \Psi(s, m+1)D_{\perp}a_m F_{m+1}(G_m(Z_m)) \\
 &+ \sum_{m=n_k}^s \Psi(s, m+1)D_{\perp}a_m \tilde{\epsilon}_{m+1},
 \end{aligned}$$

by (46)(47) we have

$$\begin{aligned}
 Z_{\perp,s+1} &= \left[(\Phi(s, n_k) - \frac{1}{N}\mathbf{1}\mathbf{1}^T) \otimes \mathbf{I}_l \right] (Z_{n_k} - \Theta_{n_k}) \\
 &+ \sum_{m=n_k}^s \left[(\Phi(s, m) - \frac{1}{N}\mathbf{1}\mathbf{1}^T) \otimes \mathbf{I}_l \right] D_m(Z_m) \\
 &+ \sum_{m=n_k}^s a_m \left[(\Phi(s, m+1) - \frac{1}{N}\mathbf{1}\mathbf{1}^T) \otimes \mathbf{I}_l \right] F_{m+1}(G_m(Z_m)) \\
 &+ \sum_{m=n_k}^s a_m \left[(\Phi(s, m+1) - \frac{1}{N}\mathbf{1}\mathbf{1}^T) \otimes \mathbf{I}_l \right] \tilde{\epsilon}_{m+1}. \tag{66}
 \end{aligned}$$

From (9), (51), (58), (64), A3, and A6 we can derive

$$\begin{aligned}
 \|Z_{\perp,s+1}\| &\leq Cc\rho^{s+1-n_k} + \sum_{m=n_k}^s a_m \|Z_m - \Theta_m\| c\rho^{s+1-m} \\
 &+ \sum_{m=n_k}^s a_m \alpha(2M'_0 + 2C + 3)c\rho^{s-m+2} \\
 &+ \left\| \sum_{m=n_k}^s a_m \left[(\Phi(s, m+1) - \frac{1}{N}\mathbf{1}\mathbf{1}^T) \otimes \mathbf{I}_l \right] \tilde{\epsilon}_{m+1} \right\|. \tag{67}
 \end{aligned}$$

Set $\Gamma_n \triangleq \sum_{m=1}^n a_m \tilde{\epsilon}_{m+1}$, by (62) we know that $\|\Gamma_s - \Gamma_{n_k-1}\| \leq T_k \quad \forall s : n_k \leq s \leq s_k$. Notice

$$\begin{aligned} & \sum_{m=n_k}^s a_m (\Phi(s, m+1) \otimes \mathbf{I}_l) \tilde{\epsilon}_{m+1} \\ &= \sum_{m=n_k}^s (\Phi(s, m+1) \otimes \mathbf{I}_l) (\Gamma_m - \Gamma_{m-1}) \\ &= \sum_{m=n_k}^s (\Phi(s, m+1) \otimes \mathbf{I}_l) (\Gamma_m - \Gamma_{n_k-1}) \\ &+ \sum_{m=n_k}^s (\Phi(s, m+1) \otimes \mathbf{I}_l) (\Gamma_{m-1} - \Gamma_{n_k-1}). \end{aligned}$$

So, summing by parts with (9) we have

$$\begin{aligned} & \left\| \sum_{m=n_k}^s a_m (\Phi(s, m+1) \otimes \mathbf{I}_l) \tilde{\epsilon}_{m+1} \right\| \\ & \leq \|\Gamma_s - \Gamma_{n_k}\| \\ & + \sum_{m=n_k}^s \|\Phi(s, m+1) - \Phi(s, m+2)\| \cdot \|\Gamma_m - \Gamma_{n_k-1}\| \\ & \leq T_k + \sum_{m=n_k}^s (c\rho^{s-m} + c\rho^{s-m-1}) \cdot T_k \\ & \leq T_k + \frac{c(\rho+1)}{1-\rho} T_k, \end{aligned}$$

which incorporating with (62) produces

$$\begin{aligned} & \left\| \sum_{m=n_k}^s a_m \left[(\Phi(s, m+1) - \frac{1}{N} \mathbf{1}\mathbf{1}^T) \otimes \mathbf{I}_l \right] \tilde{\epsilon}_{m+1} \right\| \\ & \leq \left(2 + \frac{c(\rho+1)}{1-\rho} \right) T_k \quad \forall s : n_k \leq s \leq s_k \end{aligned} \tag{68}$$

for sufficiently large $k \geq k_1$ and any $T_k \in [0, T]$.

Notice $\sum_{m=n_k}^s a_m \rho^{s-m} \leq \frac{1}{1-\rho} \sup_{m \geq n_k} a_m$ from $a_m \xrightarrow{m \rightarrow \infty} 0$. Combine this with (67)(68) we have

$$\begin{aligned} \|Z_{\perp, s+1}\| & \leq Cc\rho + (M'_0 + C + 1)C \frac{1}{1-\rho} \sup_{m \geq n_k} a_m \\ & + \alpha(2M'_0 + 2C + 3)C \frac{1}{1-\rho} \sup_{m \geq n_k} a_m + \left(2 + \frac{c(\rho+1)}{1-\rho} \right) T_k \\ & \leq Cc\rho + 1 + \left(2 + \frac{c(\rho+1)}{1-\rho} \right) T_k \quad \forall s : n_k \leq s \leq s_k \end{aligned} \tag{69}$$

for sufficiently large $k \geq k_1$ and any $T_k \in [0, T]$. Notice that $\hat{\Lambda}_s = Z_{\perp, s} + (\mathbf{1} \otimes \mathbf{I}_l) \hat{\Delta}_s$. We derive

$$\begin{aligned} \|\hat{\Lambda}_{s+1} - \hat{\Lambda}_{n_k}\| &= \|(\mathbf{1} \otimes \mathbf{I}_l) \hat{\Delta}_{s+1} + Z_{\perp, s+1} - Z_{\perp, n_k} - (\mathbf{1} \otimes \mathbf{I}_l) \hat{\Delta}_{n_k}\| \\ &\leq \|Z_{\perp, s+1}\| + \|Z_{\perp, n_k}\| + \sqrt{N} \|\hat{\Delta}_{s+1} - \hat{\Delta}_{n_k}\|. \end{aligned}$$

Since $\|Z_{\perp, n_k}\| \leq 2\|\hat{\Lambda}_{n_k}\| \leq 2C$, from (65)(67) it follows that for sufficiently large $k \geq k_1$ and any $T_k \in [0, T]$

$$\begin{aligned} & \|\hat{\Lambda}_{s+1} - \hat{\Lambda}_{n_k}\| \\ & \leq Cc\rho + 1 + \left(2 + \frac{c(\rho+1)}{1-\rho} \right) T_k + 2C + \sqrt{N}c_2T_k \\ & \leq C(c\rho + 2) + 1 + \left(2 + \frac{c(\rho+1)}{1-\rho} + c_2\sqrt{N} \right) T_k \\ & = m'_0 + c_1T_k. \end{aligned} \tag{70}$$

Therefore, from (56)(51) we know that for sufficiently large $k \geq k_1$ and any $T_k \in [0, T]$

$$\|\hat{\Lambda}_{s_k+1}\| \leq \|\hat{\Lambda}_{n_k}\| + m'_0 + c_1T_k \leq M'_0 + 1 + C. \tag{71}$$

Now we look back at the recursive algorithm (19). We rewrite it in the compact form as follows

$$\begin{aligned} \hat{X}_{s_k+1} &= [W(s_k) \otimes \mathbf{I}_l] G_{s_k}(\tilde{X}_{s_k}) \\ & + a_{s_k} \left(F_{s_k+1}(G_{s_k}(\tilde{X}_{s_k})) + \tilde{\epsilon}_{s_k+1} \right), \end{aligned}$$

where $\hat{X}_k \triangleq \text{col}\{\hat{x}_{1,k}, \dots, \hat{x}_{N,k}\}$. Then by (63)(64), $\hat{X}_{s_k+1} = Z_{s_k+1}$. So by (71) it follows that

$$\|\hat{X}_{s_k+1} - \Theta_{s_k+1}\| \leq M'_0 + 1 + C. \tag{72}$$

We now show

$$\tilde{X}_{s_k+1} = \hat{X}_{s_k+1}, \quad s_k + 1 < \tau_{\sigma_k+1} \tag{73}$$

for sufficiently large $k \geq k_1$ and any $T_k \in [0, T]$. We consider the following two cases: $\lim_{k \rightarrow \infty} \sigma_k = \infty$ and $\lim_{k \rightarrow \infty} \sigma_k = \sigma < \infty$.

i) $\lim_{k \rightarrow \infty} \sigma_k = \infty$: Notice $M_{\sigma_k} > \eta + M'_0 + 1 + C$ when $k \geq k_1$. By (20)(21) we know that $\tilde{X}_{s_k+1} = \hat{X}_{s_k+1}$, $\sigma_{s_k+1} = \sigma_{s_k}$. So $s_k + 1 < \tau_{\sigma_{s_k+1}}$ by (60).

ii) $\lim_{k \rightarrow \infty} \sigma_k = \sigma < \infty$: For this case $\tau_{\sigma_{n_k+1}} = \infty$ for all $k \geq k_1$. By (60) we see $s_k + 1 < \tau_{\sigma_{n_k+1}}$. Then by $\sigma_{n_k} = \sigma$ we conclude $\sigma_{s_k+1} = \sigma_{s_k} = \sigma$, and hence by (20) we derive $\tilde{X}_{s_k+1} = \hat{X}_{s_k+1}$. Thus (73) holds.

From (73) we know that (48) holds for $m = s_k$ for sufficiently large $k \geq k_1$ and any $T_k \in [0, T]$. From $\hat{X}_{s_k+1} = Z_{s_k+1}$ by (73) we see $\tilde{X}_{s_k+1} = Z_{s_k+1}$. It follows that for sufficiently large $k \geq k_1$ and any $T_k \in [0, T]$

$$\|\tilde{\Lambda}_{s_k+1} - \tilde{\Lambda}_{n_k}\| \leq M'_0 + c_1T_k,$$

which contradicts with the definition of s_k . Thus (59) does not hold. So $s_k > m(n_k, T_k)$ and hence (49) holds.

Since $s_k > m(n_k, T_k)$, we know $\{\tilde{\Lambda}_s : n_k \leq s \leq m(n_k, T_k) + 1\}$ is bounded. Similar to proving (60) we can be shown that $m(n_k, T_k) < \tau_{\sigma_k+1}$. So (48) holds for $m = n_k, \dots, m(n_k, T_k)$. Similar to (65) we can prove

$$\|\tilde{\Delta}_{s+1} - \tilde{\Delta}_{n_k}\| \leq c_2T$$

for sufficiently large k and any $T_k \in [0, T]$. Hence, (50) holds.

In conclusion, the proof of Lemma 6 is complete. \square

By multiplying both sides of (48) with $\frac{1}{N}(\mathbf{1}^T \otimes \mathbf{I}_l)$, we have

$$\begin{aligned} \tilde{x}_{m+1} &= \frac{1}{N} \sum_{i=1}^N g_m(\tilde{x}_{i,m}) + a_m \frac{1}{N} \sum_{i=1}^N f_{i,m+1}(g_m(\tilde{x}_{i,m})) \\ &+ a_m \frac{1}{N} \sum_{i=1}^N \tilde{\epsilon}_{i,m+1} \\ &= g_m(\tilde{x}_m) + a_m f_{m+1}(g_m(\tilde{x}_m)) \\ &+ a_m \frac{1}{N} \frac{1}{a_m} \sum_{i=1}^N (g_m(\tilde{x}_{i,m}) - g_m(\tilde{x}_m)) \\ &+ a_m \frac{1}{N} \sum_{i=1}^N (f_{i,m+1}(g_m(\tilde{x}_{i,m})) - f_{i,m+1}(g_m(\tilde{x}_m))) \\ &+ a_m \frac{1}{N} \sum_{i=1}^N \tilde{\epsilon}_{i,m+1}. \end{aligned} \tag{74}$$

Setting

$$\begin{aligned} \zeta_{k+1}^{(1)} &= \frac{1}{N} \frac{1}{a_k} \sum_{i=1}^N (g_k(\tilde{x}_{i,k}) - g_k(\tilde{x}_k)), \\ \zeta_{k+1}^{(2)} &= \frac{1}{N} \sum_{i=1}^N (f_{i,k+1}(g_k(\tilde{x}_{i,k})) - f_{i,k+1}(g_k(\tilde{x}_k))), \\ \zeta_{k+1}^{(3)} &= \frac{1}{N} \sum_{i=1}^N \tilde{\epsilon}_{i,k+1}, \\ \zeta_{k+1} &= \zeta_{k+1}^{(1)} + \zeta_{k+1}^{(2)} + \zeta_{k+1}^{(3)} \end{aligned}$$

We can rewrite (74) as

$$\tilde{x}_{m+1} = g_m(\tilde{x}_m) + a_m f_{m+1}(g_m(\tilde{x}_m)) + a_m \zeta_{m+1}. \tag{75}$$

The following lemma gives the noise property of the sequence $\{\zeta_{k+1}\}$.

Lemma 7 Assume all the conditions in Lemma 6 hold. $\{\tilde{\Lambda}_{n_k}\}$ is a convergent subsequence with limit $\tilde{\Lambda}$ at the considered sample ω . Then for this ω

$$\lim_{T \rightarrow 0} \limsup_{k \rightarrow \infty} \frac{1}{T} \left\| \sum_{s=n_k}^{m(n_k, T_k)} a_s \zeta_{s+1} \right\| = 0 \quad T_k \in [0, T]. \tag{76}$$

Proof In the proof of Lemma 6 it has been pointed out that there exists a $T \in (0, 1)$ such that $m(n_k, T) < \tau_{\sigma_{n_k+1}}$ for sufficiently large k . So $\left\| \sum_{s=n_k}^{m(n_k, T_k) \wedge (\tau_{\sigma_{n_k+1}} - 1)} a_s \tilde{\epsilon}_{s+1} \right\| = \left\| \sum_{s=n_k}^{m(n_k, T_k)} a_s \tilde{\epsilon}_{s+1} \right\|$. Thus, by Lemma 5 we can immediately derive that

$$\lim_{T \rightarrow 0} \limsup_{k \rightarrow \infty} \frac{1}{T} \left\| \sum_{s=n_k}^{m(n_k, T_k)} a_s \zeta_{s+1}^{(3)} \right\| = 0 \quad T_k \in [0, T].$$

Now we need to show that $\zeta_{k+1}^{(i)}$ also satisfies the property above for $i = 1, 2$. First we consider $\zeta_{k+1}^{(1)}$. We see that

$$\begin{aligned} &\left\| \sum_{s=n_k}^{m(n_k, T_k)} a_s \zeta_{s+1}^{(1)} \right\| \\ &= \left\| \sum_{s=n_k}^{m(n_k, T_k)} \frac{1}{N} \sum_{i=1}^N (g_s(\tilde{x}_{i,s}) - g_s(\tilde{x}_s)) \right\| \\ &= \left\| \frac{1}{N} \sum_{s=n_k}^{m(n_k, T_k)} \sum_{i=1}^N (g_s(\tilde{x}_{i,s}) - g_s(\theta_s) - g_s(\tilde{x}_s) + g_s(\theta_s)) \right\| \\ &= \frac{1}{N} \left\| \sum_{s=n_k}^{m(n_k, T_k)} \sum_{i=1}^N (d_s(\tilde{x}_{i,s}) + \tilde{\Delta}_{i,s} - d_s(\tilde{x}_s) - \tilde{\Delta}_s) \right\| \\ &= \frac{1}{N} \left\| \sum_{s=n_k}^{m(n_k, T_k)} \sum_{i=1}^N (d_s(\tilde{x}_{i,s}) - d_s(\tilde{x}_s)) \right\| \\ &\leq \frac{1}{N} \sum_{s=n_k}^{m(n_k, T_k)} \sum_{i=1}^N \|d_s(\tilde{x}_{i,s})\| + \sum_{s=n_k}^{m(n_k, T_k)} \|d_s(\tilde{x}_s)\| \\ &\leq \frac{1}{N} \sum_{s=n_k}^{m(n_k, T_k)} \sum_{i=1}^N \gamma_s \|\tilde{\Delta}_{i,s}\| + \sum_{s=n_k}^{m(n_k, T_k)} \gamma_s \|\tilde{\Delta}_s\|, \end{aligned} \tag{77}$$

where the last inequality comes from A6.

Since $\lim_{k \rightarrow \infty} \tilde{\Lambda}_{n_k} = \tilde{\Lambda}$, by setting $\tilde{\Delta} \triangleq \frac{\mathbf{1}^T \otimes \mathbf{I}_l}{N} \tilde{\Lambda}$ we see that $\lim_{k \rightarrow \infty} \tilde{\Delta}_{n_k} = \tilde{\Delta}$. So by (49)(50) we conclude that $\{\|\tilde{\Delta}_{i,s}\|, s : n_k \leq s \leq m(n_k, T_k)\}$ and $\{\|\tilde{\Delta}_s\|, s : n_k \leq s \leq m(n_k, T_k)\}$ are bounded. Without the loss of generality we denote the bound of these two sequence by A . Then from (77) with A6 it follows

$$\left\| \sum_{s=n_k}^{m(n_k, T_k)} a_s \zeta_{s+1}^{(1)} \right\| \leq 2A \sum_{s=n_k}^{m(n_k, T_k)} \gamma_s \xrightarrow[k \rightarrow \infty]{} 0.$$

So we conclude

$$\lim_{T \rightarrow 0} \limsup_{k \rightarrow \infty} \frac{1}{T} \left\| \sum_{s=n_k}^{m(n_k, T_k)} a_s \zeta_{s+1}^{(1)} \right\| = 0 \quad T_k \in [0, T].$$

Finally, we consider the case $i = 2$. Notice

$$\begin{aligned} &\left\| \sum_{s=n_k}^{m(n_k, T_k)} a_s \zeta_{s+1}^{(2)} \right\| \\ &= \left\| \sum_{s=n_k}^{m(n_k, T_k)} \frac{1}{N} \sum_{i=1}^N a_s (f_{i,s+1}(g_s(\tilde{x}_{i,s})) - f_{i,s+1}(g_s(\tilde{x}_s))) \right\| \end{aligned}$$

Similar to the proof in Lemma 6 we know that there exist constants $c_3, c_4, c_5 > 0$ such that for sufficiently large k

$$\|\tilde{X}_{\perp, s+1}\| \leq c_3 \rho^{s+1-n_k} + c_4 \sup_{m \geq n_k} a_m + c_5 T \tag{78}$$

hold for $\forall s : n_k \leq s \leq m(n_k, T)$. From A1,A6, and A7 we can assume that for sufficiently large k

$$a_k < 1, \gamma_k \leq a_k, \|\xi_{k+1}\| \leq a_k. \tag{79}$$

Since $0 < \rho < 1$, there exists a positive integer m' such that $\rho^{m'} < T$. Then $\sum_{m=n_k}^{n_k+m'} a_m \xrightarrow[k \rightarrow \infty]{} 0$. Thus, we have $n_k + m' < m(n_k, T)$ for sufficiently large k . So from (78) it follows

$$\begin{aligned} \|\tilde{X}_{\perp, s+1}\| &\leq o(1) + (c_3 + c_5)T \\ \forall s : n_k + m' \leq s \leq m(n_k, T). \end{aligned} \tag{80}$$

where $o(1) \rightarrow 0$ as $k \rightarrow \infty$. Hence for $n_k + m' < m(n_k, T)$ and sufficiently large k

$$\|\tilde{x}_{i,s} - \tilde{x}_s\| = o(1) + \delta(T)$$

where $\delta(T) \rightarrow 0$ as $T \rightarrow 0$.

So

$$\begin{aligned} &\left\| \sum_{s=n_k}^{m(n_k, T_k)} a_s \zeta_{s+1}^{(2)} \right\| \\ &\leq \left\| \sum_{s=n_k}^{n_k+m'} \frac{1}{N} \sum_{i=1}^N a_s (f_{i,s+1}(g_s(\tilde{x}_{i,s})) - f_{i,s+1}(g_s(\tilde{x}_s))) \right\| \\ &+ \left\| \sum_{s=n_k+m'}^{m(n_k, T)} \frac{1}{N} \sum_{i=1}^N a_s (f_{i,s+1}(g_s(\tilde{x}_{i,s})) - f_{i,s+1}(g_s(\tilde{x}_s))) \right\| \\ &\leq \sum_{s=n_k}^{n_k+m'} \frac{1}{N} \sum_{i=1}^N a_s \left(\|f_{i,s+1}(g_s(\tilde{x}_{i,s}))\| + \|f_{i,s+1}(g_s(\tilde{x}_s))\| \right) \\ &+ \sum_{s=n_k+m'}^{m(n_k, T)} a_s (o(1) + \delta(T)) \\ &\leq \sum_{s=n_k}^{n_k+m'} a_s \alpha(2A + 1) + T(o(1) + \delta(T)) \\ &\leq \alpha(2A + 1) \cdot m' \sup_{m \geq n_k} a_m + T(o(1) + \delta(T)). \end{aligned}$$

And hence $\limsup_{k \rightarrow \infty} \frac{1}{T} \left\| \sum_{s=n_k}^{m(n_k, T_k)} a_s \zeta_{s+1}^{(2)} \right\| = \delta(T)$, which implies that

$$\lim_{T \rightarrow 0} \limsup_{k \rightarrow \infty} \frac{1}{T} \left\| \sum_{s=n_k}^{m(n_k, T_k)} a_s \zeta_{s+1}^{(2)} \right\| = 0 \quad T_k \in [0, T].$$

So we complete the proof. \square

Lemma 8 Assume A1-A4, A6 hold, and that A5, A7 hold at the sample path ω under consideration. Then any nonempty interval $[\delta_1, \delta_2]$ with $0 \notin [\delta_1, \delta_2]$ cannot be crossed by $\{v(\tilde{\Delta}_{m_k}), \dots, v(\tilde{\Delta}_{l_k})\}$ infinitely many times with $\{\tilde{\Lambda}_{m_k}\}$ bounded. By "[δ_1, δ_2] being crossed by $\{v(\tilde{\Delta}_{m_k}), \dots, v(\tilde{\Delta}_{l_k})\}$ " it means that $v(\tilde{\Delta}_{m_k}) \leq \delta_1, v(\tilde{\Delta}_{l_k}) \geq \delta_2$ and $\delta_1 < v(\tilde{\Delta}_s) < \delta_2 \quad \forall s : m_k < s < l_k$.

Proof Assume the converse: for some interval $[\delta_1, \delta_2]$ with $0 \notin [\delta_1, \delta_2]$, there are infinitely many crossing $\{v(\tilde{\Delta}_{m_k}), \dots, v(\tilde{\Delta}_{l_k})\}$ with $\{\tilde{\Lambda}_{m_k}\}$ bounded.

By the boundedness of $\{\tilde{\Lambda}_{m_k}\}$ we can extract a convergent subsequence still denoted by $\{\tilde{\Lambda}_{m_k}\}$ with limit $\lim_{k \rightarrow \infty} \tilde{\Lambda}_{m_k} = \tilde{\Lambda}$. So, $\lim_{k \rightarrow \infty} \tilde{\Delta}_{m_k} = \tilde{\Delta}$ with $\tilde{\Delta} \triangleq \frac{1^T \otimes I_l}{N} \tilde{\Lambda}$. By (50), letting $T \rightarrow 0$ we have $\|\tilde{\Delta}_{m_k+1} - \tilde{\Delta}_{m_k}\| \leq c_2 T \rightarrow 0$. By the definition of crossing, $v(\tilde{\Delta}_{m_k}) \leq \delta_1 < v(\tilde{\Delta}_{m_k+1})$, we can obtain

$$v(\tilde{\Delta}_{m_k}) \xrightarrow[k \rightarrow \infty, T \rightarrow 0]{} \delta_1 = v(\tilde{\Delta}) > 0. \tag{81}$$

So by the assumption $v(x) = 0 \Leftrightarrow x = 0$ we know there exists a constant β such that $\|\tilde{\Delta}\| > \beta$. And hence by (50) we conclude

$$\|\tilde{\Delta}_j\| > \frac{\beta}{2}, \quad j : m_k \leq j \leq m(m_k, T) + 1 \tag{82}$$

for sufficiently small $T > 0$ and large k .

Setting $\tilde{\Delta}_k$ to be a vector in-between $\tilde{\Delta}_{m_k}$ and $\tilde{\Delta}_{m(m_k, T)}$. From (50) it follows that $\|\tilde{\Delta}_k\| \leq c_2 T + \|\tilde{\Delta}\| + 1$ for sufficiently large k . We consider the following Taylor's expansion

$$\begin{aligned} &v(\tilde{\Delta}_{m(m_k, T)}) - v(\tilde{\Delta}_{m_k}) \\ &= v_x^T(\tilde{\Delta}_k) \sum_{j=m_k}^{m(m_k, T)-1} (d_j(\tilde{x}_j) - \xi_{j+1} + a_j(f_{j+1}(g_j(\tilde{x}_j)) + \zeta_{j+1})) \\ &= v_x^T(\tilde{\Delta}_k) \left\{ \sum_{j=m_k}^{m(m_k, T)-1} (d_j(\tilde{x}_j) + \xi_{j+1}) + \sum_{j=m_k}^{m(m_k, T)-1} a_j \zeta_{j+1} \right\} \\ &+ \sum_{j=m_k}^{m(m_k, T)-1} a_j (v_x^T(\tilde{\Delta}_k) - v_x^T(g_j(\tilde{x}_j) - \theta_{j+1})) f_{j+1}(g_j(\tilde{x}_j)) \\ &+ \sum_{j=m_k}^{m(m_k, T)-1} a_j v_x^T(g_j(\tilde{x}_j) - \theta_{j+1}) f_{j+1}(g_j(\tilde{x}_j)) \end{aligned} \tag{83}$$

Similar to (51), we take sufficiently large k , then by A3, A6 and Lemma 7, there exists a constant c' such that

$$\begin{aligned} &\left\| a_j(f_{j+1}(g_j(\tilde{x}_j)) + \zeta_{j+1}) \right\| \\ &\leq \left\| a_j f_{j+1}(\theta_{j+1} + d_j(\tilde{x}_j) - \xi_{j+1} + \tilde{\Delta}_j) \right\| \\ &+ \left\| \sum_{l=m_k}^j a_l \zeta_{l+1} - \sum_{l=m_k}^{j-1} a_l \zeta_{l+1} \right\| \\ &\leq a_j \alpha (2\|\tilde{\Delta}\| + 1) + 2c' T < \frac{\beta}{4} \end{aligned} \tag{84}$$

for sufficiently large k and sufficiently small T , where $j : m_k \leq j \leq m(m_k, T) - 1$. It follows that

$$\begin{aligned} \|g_j(\tilde{x}_j) - \theta_{j+1}\| &= \|\tilde{\Delta}_{j+1} + g_j(\tilde{x}_j) - \tilde{x}_{j+1}\| \\ &\geq \frac{\beta}{2} - \frac{\beta}{4} = \frac{\beta}{4}, \end{aligned} \tag{85}$$

$$\begin{aligned} \|g_j(\tilde{x}_j) - \theta_{j+1}\| &= \|\tilde{x}_{j+1} - \theta_{j+1} - a_j f_{j+1}(g_j(\tilde{x}_j)) + \zeta_{j+1}\| \\ &< \frac{\beta}{4} + \|\tilde{\Delta}_{j+1}\| \leq \frac{\beta}{4} + c_2 T + \|\tilde{\Delta}\| + 1. \end{aligned} \tag{86}$$

Identifying r_1 and r_2 in A2 to $\frac{\beta}{4}$ and $\frac{\beta}{4} + \|\tilde{\Delta}_{j+1}\| \leq \frac{\beta}{4} + c_2 T + \|\tilde{\Delta}\|$, respectively, we can find $a > 0$ such that for $\forall j : m_k \leq j \leq m(m_k, T)$

$$v_x^T(g_j(\tilde{x}_j) - \theta_{j+1})f_{j+1}(g_j(\tilde{x}_j)) < -a. \tag{87}$$

Noticing that for $\forall j : m_k \leq j \leq m(m_k, T)$, $\|d_j(\tilde{x}_j)\| \leq \gamma_j \|\tilde{\Delta}_j\| \leq \gamma_j(c_2 T + \|\tilde{\Delta}\| + 1)$, by A6 and A7 we have

$$\lim_{k \rightarrow \infty} \sum_{j=m_k}^{m(m_k, T)-1} (d_j(\tilde{x}_j) - \xi_{j+1}) = 0. \tag{88}$$

By Lemma 7 it follows

$$\limsup_{k \rightarrow \infty} \left\| \sum_{j=m_k}^{m(m_k, T)-1} a_j \zeta_{j+1} \right\| = \delta(T). \tag{89}$$

Notice that for $\forall j : m_k \leq j \leq m(m_k, T)$

$$\begin{aligned} &\|\tilde{\Delta}_k - (g_j(\tilde{x}_j) - \theta_{j+1})\| \\ &\leq \|\tilde{\Delta}_k - \tilde{\Delta}_{m_k}\| + \|\tilde{\Delta}_j - \tilde{\Delta}_{m_k}\| + \|g_j(\tilde{x}_j) - \theta_{j+1} - \tilde{\Delta}_j\| \\ &\leq 2c_1 T + \|d_j(\tilde{x}_j) - \xi_{j+1}\| \\ &\leq 2c_1 T + \gamma_j(c_2 T + \|\tilde{\Delta}\| + 1) + \|\xi_{j+1}\| \xrightarrow[k \rightarrow \infty, T \rightarrow 0]{} 0. \end{aligned} \tag{90}$$

So by the continuity of $v(\cdot)$ we know

$$v_x^T(\tilde{\Delta}_k) - v_x^T(g_j(\tilde{x}_j) - \theta_{j+1}) \xrightarrow[k \rightarrow \infty, T \rightarrow 0]{} 0. \tag{91}$$

From A3, A6 and (51), we've already utilized this inequality before

$$\left\| f_{j+1}(g_j(\tilde{x}_j)) \right\| \leq \alpha(2\|\tilde{\Delta}\| + 1). \tag{92}$$

So, from (88)(89)(92) we can conclude that the first and second term of (83) is $o(T)$ as $k \rightarrow \infty, T \rightarrow 0$. Combining this with (87) it follows that for sufficiently large k and sufficiently small T by (83) we have

$$v(\tilde{\Delta}_{m(m_k, T)}) - v(\tilde{\Delta}_{m_k}) \leq -\frac{a}{2}T. \tag{93}$$

Let $k \rightarrow \infty$ we have

$$\limsup_{k \rightarrow \infty} v(\tilde{\Delta}_{m(m_k, T)}) \leq \delta_1 - \frac{a}{2}T. \tag{94}$$

Notice that by Lemma 6 we have

$$\lim_{T \rightarrow 0} \max_{m_k \leq m \leq m(m_k, T)} \left\| v(\tilde{\Delta}_m) - v(\tilde{\Delta}_{m_k}) \right\| = 0$$

which implies that $m(m_k, T) < l_k$ for sufficiently small T . Therefore, $v(\tilde{\Delta}_{m(m_k, T)}) \in [\delta_1, \delta_2]$ which contradicts with (94). So, the converse assumption is not true. The proof is completed. \square

Lemma 9 Assume all the assumptions required by Lemma 8 hold. Then there exists a positive integer σ such that

$$\lim_{k \rightarrow \infty} \sigma_k = \sigma < \infty. \tag{95}$$

Proof Assume the converse:

$$\lim_{k \rightarrow \infty} \sigma_k = \infty. \tag{96}$$

Then there exists a sequence of integer $\{n_k\}_{k \geq 0}$ such that $\sigma_{n_k} = k$ and $\sigma_{n_k-1} = k - 1$. By the algorithm (3)–(7) we know that $\tilde{x}_{i, n_k} = h_{n_k-1}(x^*) \forall i \in \mathcal{V}$. Therefore, from Lemma 1 we know that $\{\Lambda_{n_k}\}$ is a bounded sequence, and hence, it contains a convergent subsequence. For the sake of convenience, We denote the convergent subsequence still by $\{\tilde{\Lambda}_{n_k}\}$ with limit $\tilde{\Lambda}$.

Since $\{M_k\}_{k \geq 0}$ is a sequence of positive numbers increasingly diverging to infinity, there exists a positive integer k_0 such that

$$M_k \geq 2\sqrt{N}r + 2 + M'_1 \quad \forall k \geq k_0 \tag{97}$$

where r is given in A2 and

$$M'_1 = 2 + (2\sqrt{N}r + 2)(c\rho + 2). \tag{98}$$

Now we show that under the converse assumption, $\{\tilde{\Delta}_{n_k}\}$ starting from n_k will exit the ball r infinitely many times. Define

$$m_k \triangleq \inf\{s > n_k : \|\tilde{\Lambda}_s\| \geq 2\sqrt{N}r + 2 + m'_1\}, \tag{99}$$

$$l_k \triangleq \sup\{s < m_k : \|\tilde{\Lambda}_s\| \leq 2\sqrt{N}r + 2\}. \tag{100}$$

Notice that $\|\tilde{\Lambda}_{n_k}\| = \sqrt{N}\eta$ by Lemma 1 and $r > \eta$ from A2, we derive $\|\tilde{\Lambda}_{n_k}\| < \sqrt{N}r$. Hence from (99) (100) we have $n_k < l_k < m_k$. By the definition of l_k we know that $\{\tilde{\Lambda}_{l_k}\}$ is bounded, then there exists a convergent subsequence denoted still by $\{\tilde{\Lambda}_{l_k}\}$.

By Lemma 6 there exist constants $M'_0 > 0$ defined by (53) with $C = 2\sqrt{N}r + 2, c_1 > 0$ defined by (54), $c_2 > 0$ defined by (55), and $0 < T < 1$ with $c_1 T < 1$ such that

$$\|\tilde{\Lambda}_{m+1} - \tilde{\Lambda}_{l_k}\| \leq c_1 T + M'_0 \quad \forall m : l_k \leq m \leq m(l_k, T)$$

for sufficiently large $k \geq k_0$. Then for sufficiently large $k \geq k_0$ we have

$$\begin{aligned} \|\tilde{\Lambda}_{m+1}\| &\leq \|\tilde{\Lambda}_{l_k}\| + c_1 T + M'_0 \\ &\leq 2\sqrt{Nr} + 2 + 1 + 1 + (2\sqrt{Nr} + 2)(c\rho + 2) \\ &= 2\sqrt{Nr} + 2 + M'_1 \quad \forall m : l_k \leq m \leq m(l_k, T). \end{aligned} \tag{101}$$

Then $m(l_k, T) < n_{k+1}$ for sufficiently large $k \geq k_0$ by (97) and the definition of n_k .

From (101) by the definition of m_k (99), we conclude $m(l_k, T) + 1 < m_k$ for sufficiently large $k \geq k_0$. Then by (99) (100) we know that for sufficiently large $k \geq k_0$

$$2\sqrt{Nr} + 2 < \|\tilde{\Lambda}_{m+1}\| \leq 2\sqrt{Nr} + 2 + M'_1 \tag{102}$$

holds for $m : l_k \leq m \leq m(l_k, T)$.

Since $0 < \rho < 1$, there exists a positive integer m_0 such that $4c\rho^{m_0} < 1$. Then $\sum_{m=l_k}^{l_k+m_0} a_m \xrightarrow[k \rightarrow \infty]{} 0$ by A1, and hence $l_k + m_0 < m(l_k, T) < n_{k+1}$ for sufficiently large $k \geq k_0$. So, from (102) it can be seen that for sufficiently large $k \geq k_0$ we have

$$\|\tilde{\Lambda}_{l_k+m_0}\| > 2\sqrt{Nr} + 2 \tag{103}$$

Notice that $\{\tilde{\Lambda}_{m+1} : l_k \leq m \leq m(l_k, T)\}$ is bounded, similarly to (67) (69) we know that for sufficiently large $k \geq k_0$

$$\begin{aligned} \|\tilde{X}_{\perp, m+1}\| &\leq (2\sqrt{Nr} + 2)c\rho^{m+1-l_k} \\ &\quad + (M'_0 + C + 1)C \frac{1}{1-\rho} \sup_{m \geq n_k} a_m \\ &\quad + \alpha(2M'_0 + 2C + 3)C \frac{1}{1-\rho} \sup_{m \geq n_k} a_m \\ &\quad + \left(2 + \frac{c(\rho + 1)}{1-\rho}\right) T_k. \end{aligned}$$

Since $4c\rho^{m_0} < 1$, $c_1 T < 1$ and $a_k \xrightarrow[k \rightarrow \infty]{} 0$, it follows that

$$\|\tilde{X}_{\perp, l_k+m_0}\| \leq \frac{1}{2}(\sqrt{Nr} + 1) + \frac{1}{2} + 1 = \frac{\sqrt{Nr}}{2} + 2 \tag{104}$$

for sufficiently large $k \geq k_0$. By noticing $(\mathbf{1} \otimes \mathbf{I}_l)\tilde{\Delta}_{l_k+m_0} = \tilde{\Lambda}_{l_k+m_0} - \tilde{X}_{\perp, l_k+m_0}$, from (103) (104) we conclude that

$$\begin{aligned} \sqrt{N}\|\tilde{\Delta}_{l_k+m_0}\| &= \|\tilde{\Lambda}_{l_k+m_0} - \tilde{X}_{\perp, l_k+m_0}\| \\ &\geq \|\tilde{\Lambda}_{l_k+m_0}\| - \|\tilde{X}_{\perp, l_k+m_0}\| > \frac{3}{2}\sqrt{Nr}. \end{aligned} \tag{105}$$

Therefore, $\|\tilde{\Delta}_{l_k+m_0}\| > r$. So we prove that $\{\tilde{\Delta}_{n_k}\}$ starting from n_k will exit the ball r infinitely many times.

Since $\{\tilde{\Lambda}_{n_k}\}$ is convergent, we know that $\{\tilde{\Delta}_{y_k}\}$ is convergent and from Lemma 1 it follows that $\|\tilde{\Delta}_{n_k}\| \leq \eta$. And we know that $\{\tilde{\Delta}_{n_k}\}$ starting from n_k will exit the ball r infinitely many times. Therefore, there exists an interval $[\delta_1, \delta_2] \in (\sup_{|y| \leq \eta} \nu(y), \inf_{\|x\|=r} \nu(x))$ with $0 \notin [\delta_1, \delta_2]$,

where for any k , there is a sequence $\tilde{\Delta}_{s_k}, \dots, \tilde{\Delta}_{t_k}$ such that $n_k \leq s_k, \nu(\tilde{\Delta}_{s_k}) \leq \delta_1, \delta_1 < \nu(\tilde{\Delta}_j) < \delta_2$ for $\forall j : s_k < j < t_k$ and $\nu(\tilde{\Delta}_{t_k}) > \delta_2$. In other words, the values of $\nu(\cdot)$ at sequence $\{\tilde{\Delta}_{s_k}, \dots, \tilde{\Delta}_{t_k}\}$ cross the interval $[\delta_1, \delta_2]$ infinitely many times with $\|\Lambda_{m_k}\| < r$. This contradicts with Lemma 8. So the proof is done. $\lim_{k \rightarrow \infty} \sigma_k < \infty$ is indeed true. \square

Lemma 10 Assume all the assumptions required by Lemma 9 hold. Then

$$\lim_{k \rightarrow \infty} \sigma_{i,k} = \sigma < \infty \quad \forall i \in \mathcal{V}. \tag{106}$$

Proof From Lemma 9 it follows that

$$\sigma_{i,k} \leq \sigma \quad \forall i \in \mathcal{V}.$$

By Lemma 4 we know $\tilde{\tau}_{i,\sigma} = \tau_{i,\sigma} \leq BD + \tau_\sigma$. So, by definition we know that $\sigma_{i,k} \geq \sigma \quad \forall k \geq BD + \tau_\sigma$.

In conclusion, we have $\sigma_{i,k} = \sigma \quad \forall k \geq BD + \tau_\sigma \quad \forall i \in \mathcal{V}$. The proof is completed. \square

By the definition of the auxiliary sequence $\{\tilde{x}_{i,k}\}$, we can see that Lemma 10 indicates the fact that $\{\tilde{x}_{i,k}\}$ and $\{x_{i,k}\}$ coincide in a finite number of steps.

Proof of Theorem 1: By (95) and (106), there exists a positive integer σ depending on ω such that

$$\hat{\sigma}_{i,k} = \sigma_{i,k} = \sigma \quad \forall k \geq k_0 \triangleq BD + \tau_\sigma \quad \forall i \in \mathcal{V}, \tag{107}$$

and hence by (3)

$$x'_{i,k+1} = \sum_{j \in N_i(k)} w_{ij}(k)g_j(x_{j,k}) + a_k O_{i,k+1} \quad \forall k \geq k_0 \quad \forall i \in \mathcal{V}, \tag{108}$$

by (5) $\|x'_{i,k+1} - h_k(x^*)\| \leq M_\sigma$ and by (4) $x_{i,k+1} = x'_{i,k+1}$ for any $k \geq k_0$ and any $i \in \mathcal{V}$. So, we have proved the assertion i).

Multiply (11) by D_\perp from left, we derive

$$\begin{aligned} X_{\perp, k+1} &= D_\perp(W(k) \otimes \mathbf{I}_l)G_k(X_k) \\ &\quad + a_k D_\perp(F_{k+1}(G_k(X_k)) + \epsilon_{k+1}) \\ &= D_\perp(W(k) \otimes \mathbf{I}_l)D_\perp X_k \\ &\quad + D_\perp(W(k) \otimes \mathbf{I}_l)(G_k(X_k) - X_k) \\ &\quad + a_k D_\perp(F_{k+1}(G_k(X_k)) + \epsilon_{k+1}) \\ &= D_\perp(W(k) \otimes \mathbf{I}_l)X_{\perp, k} \\ &\quad + D_\perp(W(k) \otimes \mathbf{I}_l)(G_k(X_k) - G_k(\Theta_k) - X_k + \Theta_k) \\ &\quad + a_k D_\perp(F_{k+1}(G_k(X_k)) + \epsilon_{k+1}) \\ &= D_\perp(W(k) \otimes \mathbf{I}_l)X_{\perp, k} \\ &\quad + D_\perp(W(k) \otimes \mathbf{I}_l)D_k(X_k) \\ &\quad + a_k D_\perp(F_{k+1}(G_k(X_k)) + \epsilon_{k+1}), \end{aligned}$$

where the third equality comes from the fact that $(W(k) \otimes \mathbf{I}_l)G_k(X_k) = (W(k) \otimes \mathbf{I}_l)D_\perp$ and $D_\perp G_k(X_k) = D_\perp X_k = 0$. So, for any $k \geq k_0$ by induction we have

$$X_{\perp,k+1} = \Psi(k, k_0)(X_{k_0} - \Theta_{k_0}) + \sum_{m=k_0}^k \Psi(k, m)D_m(X_m) + \sum_{m=k_0}^k a_m \Psi(k, m+1)D_\perp(F_{m+1}(G_m(X_m)) + \epsilon_{m+1}).$$

From Lemma 10 we know the number of truncations is finite. So, $\{x_{i,k} - h_{k-1}(x^*)\}_{k \geq 1}$ is bounded. Furthermore, Lemma 1 shows that $\{h_{k-1}(x^*) - \theta_k\}_{k \geq 1}$ is bounded as well. Therefore, $\{x_{i,k} - \theta_k\}_{k \geq 1}$ is bounded. By assumption A6, A7 we can take sufficiently large $k_1 > k_0$ such that $\|\xi_k\| \leq 1$, $\gamma_k \leq 1 \quad \forall k \geq k_1$. So, for sufficiently large k , there exists constants $c_6, c_7, c_8, c'_8, c_9 > 0$ such that

$$\begin{aligned} \|X_{\perp,k+1}\| &\leq c_6 \rho^{k+1-k_0} + c_7 \sum_{m=k_0}^k \gamma_m \rho^{k+1-m} \\ &+ c_8 \sum_{m=k_0}^k \alpha(2\|\Lambda_m\| + 1) a_m \rho^{k-m+2} \\ &+ c_9 \sum_{m=k_0}^k a_m \rho^{k-m+2} \|\epsilon_{m+1}\| \\ &\leq c_6 \rho^{k+1-k_0} + c_7 \sum_{m=k_0}^k \gamma_m \rho^{k+1-m} \\ &+ c'_8 \sum_{m=k_0}^k a_m \rho^{k-m+2} \\ &+ c_9 \sum_{m=k_0}^k a_m \rho^{k-m+2} \|\epsilon_{m+1}\|. \end{aligned} \quad (109)$$

Notice that for any $\epsilon > 0$, there exists integer $k_2 > k_1$ such that $\gamma_k < \epsilon$. We can derive

$$\begin{aligned} \sum_{m=0}^k \gamma_m \rho^{k-m+1} &= \sum_{m=0}^{k_2} \gamma_m \rho^{k-m+1} + \sum_{m=k_2+1}^k \gamma_m \rho^{k-m+1} \\ &\leq \rho^{k-k_2+1} \sum_{m=0}^{k_2} \gamma_m + \epsilon \frac{1}{1-\rho} \xrightarrow[k \rightarrow \infty, \epsilon \rightarrow 0]{} 0. \end{aligned}$$

Therefore, the second and third term at the right-hand side of (109) tends to zero as $k \rightarrow \infty$. Similarly, the last term of (109) also tends to zero since $\lim_{k \rightarrow \infty} a_k \epsilon_{k+1} = 0$. The first term of (109) tends to zero as $k \rightarrow \infty$ as well since $0 < \rho < 1$. So, we conclude that

$$X_{\perp,k} \xrightarrow[k \rightarrow \infty]{} 0.$$

We now show the convergence of $\{v(\tilde{\Delta}_k)\}$. Since

$$v_1 \triangleq \liminf_{k \rightarrow \infty} v(\tilde{\Delta}_k) \leq \limsup_{k \rightarrow \infty} v(\tilde{\Delta}_k) \triangleq v_2,$$

we aim to prove $v_1 = v_2$. Assume the converse: $v_1 < v_2$. Then there exists an interval $[\delta_1, \delta_2] \in (v_1, v_2)$ such that $0 \notin [\delta_1, \delta_2]$. $v(\tilde{\Delta}_k)$ crosses the interval $[\delta_1, \delta_2]$ infinite many times. By Lemma 9 and (13)–(14) we know $\tilde{\Delta}_k$ is bounded, so, $\hat{\Lambda}_k$ is bounded as well. This contradicts with Lemma 8. Therefore, $\{v(\tilde{\Delta}_k)\}$ is convergent.

Finally, we show $\tilde{\Delta}_k \xrightarrow[k \rightarrow \infty]{} 0$. Assume the converse.

Then there exists a convergent subsequence $\{\tilde{\Delta}_{n_k}\}$ with limit $\tilde{\Delta} \neq 0$. Take $\beta > 0$ such that $\|\tilde{\Delta}\| > \beta$. From Lemma 6 we know for sufficiently large k and sufficiently small T we have

$$\|\tilde{\Delta}_j\| > \frac{\beta}{2}, \quad n_k \leq j \leq m(n_k, T).$$

Similar to the proof of Lemma 8, by Taylor's expansion there exists $a > 0$ such that

$$v(\tilde{\Delta}_{m(n_k, T)}) - v(\tilde{\Delta}_{n_k}) < -\frac{a}{2}T. \quad (110)$$

Since $\{v(\tilde{\Delta}_k)\}$ is convergent and by definition $m(n_k, T) \xrightarrow[k \rightarrow \infty]{} \infty$, Let $k \rightarrow \infty$ for both sides of (110), we derive $0 < -\frac{a}{2}T$ which is impossible. So $\tilde{\Delta}_k \xrightarrow[k \rightarrow \infty]{} 0$.

By Lemma 10 we know that after a finite number of steps we have $\hat{\Delta}_k = \Delta_k$, and hence, $\Delta_k \xrightarrow[k \rightarrow \infty]{} 0$. Combining $\Delta_k \xrightarrow[k \rightarrow \infty]{} 0$ with $X_{\perp,k} \xrightarrow[k \rightarrow \infty]{} 0$, we can conclude $\Lambda_k \xrightarrow[k \rightarrow \infty]{} 0$. \square

6 Numerical simulation

In this section, we apply the distributed algorithm to a distributed tracking problem and demonstrate the performance of the algorithm. Consider a maneuvering target in the 2-D plane. The state of the target θ_k at each time consists of four components $\theta_k = [\theta_k^1, \theta_k^2, \theta_k^3, \theta_k^4]^T$. They are horizontal position, horizontal velocity, vertical position, and vertical velocity, respectively. The dynamic model of the target is chosen to be a nearly constant velocity model [29], which means that the dynamic of the target is governed by:

$$\theta_{k+1} = A\theta_k + \xi_{k+1}, \quad (111)$$

where $\xi_{k+1} \in \mathbb{R}^4$ is noise, and A is defined as

$$A = \mathbf{I}_2 \otimes \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix} \quad (112)$$

with T being the sampling interval. It can be seen that when $\xi_{k+1} = 0$ the target follows a constant velocity movement. The goal of the network is to track this target by estimating the state θ_k .

Consider a sensor network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V} = \{1, \dots, N\}$, $N = 20$, and $\mathcal{E} = G(N, p_N)$ being the Poisson

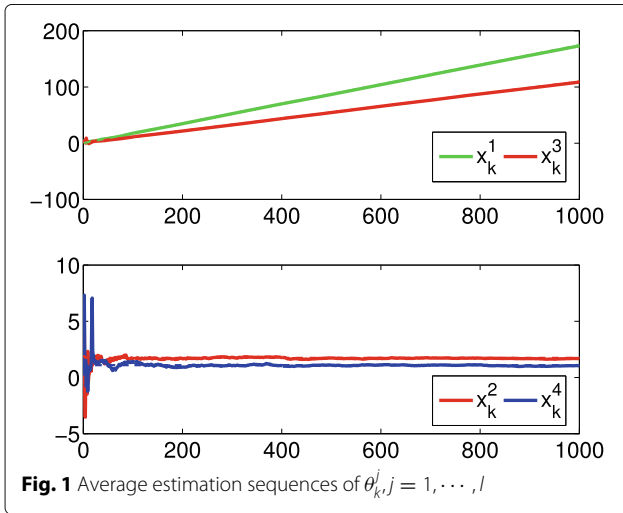


Fig. 1 Average estimation sequences of $\theta_k^j, j = 1, \dots, l$

random graph¹ with designing parameter $0 \leq p_N \leq 1$. We choose $p_N = 0.25$. Denote by N_i the neighbor set of agent i and by n_i the cardinality of N_i . Set $W(k) = [w_{ij}]_{i,j=1}^N \forall k \geq 1$ with $w_{ij} = \frac{1}{n_i}$ if agent j is in the set N_i . All agents aim to track the target state θ_k cooperatively. We assume for each agent, only one component of the target state can be observed with noise. To explain it in a mathematical model, the local function for agent i is defined as:

$$f_{i,k}(x) \triangleq e_{k_i} \theta_k - e_{k_i} x, \quad (113)$$

where e_k is a 4-dimensional square diagonal matrix with only the k th diagonal element being 1 and every other elements being 0, i.e.,

$$e_1 \triangleq \text{diag}(1, 0, 0, 0)$$

$$e_2 \triangleq \text{diag}(0, 1, 0, 0)$$

$$e_3 \triangleq \text{diag}(0, 0, 1, 0)$$

$$e_4 \triangleq \text{diag}(0, 0, 0, 1).$$

The selection of k_i will be explained later. Since the state θ_k is unknown to the agents, each agent can only get a noise-corrupted observation of this local function instead of the exact value. The global function can be written as:

$$f_k(x) \triangleq \frac{1}{N} \sum_{i=1}^N e_{k_i} \theta_k - e_{k_i} x. \quad (114)$$

It can be seen that while each agent can only estimate one component of θ_k with its own local function, θ_k is the unique root of the global function $f_k(x)$.

For our experiment, we take $\xi_k \triangleq \frac{1}{k^2} v_k$, where $\{v_k\}$ is a sequence of i.i.d. random variables uniformly distributed over $[-1, 1]$, and the step-sizes $a_k \triangleq \frac{20}{k}$. We let the sampling interval be $T \triangleq 0.1$ s, the truncation bound

$M_k \triangleq k + 80$, and $x^* \triangleq [1, 1, 1, 1]^T$. The initial value $x_{i,0}$ for all agents is chosen from the uniform distribution over $[-2, 2]$. Let the observation noise $\epsilon_{i,k}$ be the white Gaussian noise. As for the selection of k_i , for agent i , if $i \bmod 4 \neq 0$, then $k_i \triangleq i \bmod 4$, if $i \bmod 4 = 0$, then $k_i \triangleq 4$.

Denote by $\{x_{i,k}\}_{k \geq 1}, i \in \mathcal{V}$ the estimates given by (3)–(7) and by $\bar{x}_k = \frac{1}{N} \sum_{i=1}^N x_{i,k}$ the average of $x_{i,k}, i \in \mathcal{V}$. In Fig. 1, the dashed lines denote the state of the moving target and the solid lines the average estimates for entries $\{\theta_k^j, j = 1, \dots, 4\}_{k \geq 1}$ of $\{\theta_k\}_{k \geq 1}$. From the figure we can see that the estimate can track the moving target successfully.

7 Conclusion

The distributed root-tracking problem for a sum of time-varying regression functions over a network is considered in this paper. It is assumed that a noise-corrupted dynamic information of the roots is known to all agents in the network. Each agent updates its estimate by using the local observation, the dynamic information of the global root, and information received from its neighbors. A distributed stochastic approximation algorithm is proposed and the consensus and convergence of the estimates are established.

For future research, it is of interest to relax the conditions on the dynamic information of the global roots, and to consider the convergence results of the algorithm over an unbalanced network.

8 Notations

$\ \cdot\ , \ A\ $	L_2 norm of vector v , matrix A
\mathbf{I}_m	$m \times m$ identity matrix
$\mathbf{1}$	Vector or matrix with all entries equal to 1
$\mathbf{0}$	Vector or matrix with all entries equal to 0
X^T	Transpose of matrix X
$\text{col}\{x_1, \dots, x_m\}$	$\text{col}\{x_1, \dots, x_m\} \triangleq \{x_1^T, \dots, x_m^T\}^T$, stacks of vectors or matrices $\{x_1, \dots, x_m\}$
$\mathbb{I}_A(x)$	Indicator function, $\mathbb{I}_A(x) = 1$ if $x \in A$, $\mathbb{I}_A(x) = 0$ otherwise
\otimes	Kronecker product
D_{\perp}	$D_{\perp} \triangleq (\mathbf{I}_N - \frac{\mathbf{1}\mathbf{1}^T}{N}) \otimes \mathbf{I}_l$ with N the number of agents in the network and l the dimension of coefficient vector
$m(k, T)$	$m(k, T) \triangleq \max\{m : \sum_{i=k}^m a_i \leq T\}$
$\sigma_{i,k}$	Truncation number of agent i at time k
$\tau_{i,m}$	Smallest time when truncation number of agent i reaching m , i.e., $\tau_{i,m} \triangleq \inf\{k : \sigma_{i,k} = m\}$
τ_m	Smallest time when at least one of agents having its truncation number reached m , i.e., $\tau_m \triangleq \min_{i \in \mathcal{V}} \tau_{i,m}$
$\tilde{\tau}_{i,m}$	$\tilde{\tau}_{i,m} \triangleq \min\{\tau_{i,m}, \tau_{m+1}\}$
$\hat{\sigma}_{i,k}$	$\hat{\sigma}_{i,k} \triangleq \max_{j \in N_i(k)} \sigma_{j,k}$, where $N_i(k)$ is the set of neighbors of agent i at time k
σ_k	Largest truncation number of all agents at time k , i.e., $\sigma_k = \max_{i \in \mathcal{V}} \sigma_{i,k} = \max_{i \in \mathcal{V}} \hat{\sigma}_{i,k}$

¹For the details of Poisson random graph, we refer to [30].

Code availability

Not applicable.

Authors' contributions

All authors contributed to the study conception and design. Mathematical analysis was performed by Kewei Fu, Han-Fu Chen and Wenxiao Zhao. The first draft of the manuscript was written by Kewei Fu and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Authors' information

Kewei Fu received the B.S. degree from Shandong University, China in 2015. He got his Ph.D. degree from the Institute of Systems Science (ISS), Academy of Mathematics and Systems Science (AMSS), Chinese Academy of Sciences (CAS), China in 2020. His research interests include the stochastic approximation algorithm and the distributed optimization and estimation over networks.

Han-Fu Chen graduated from the Leningrad (St. Petersburg) State University, Russia. He is a Professor of the Key Laboratory of Systems and Control of the Chinese Academy of Sciences (CAS). His research interests are mainly in stochastic systems, including system identification, adaptive control, and stochastic approximation and its applications to systems, control, and signal processing. He authored and coauthored more than 210 journal papers and eight books. He served as an IFAC Council Member (2002-2005), President of the Chinese Association of Automation (1993-2002), and a Permanent member of the Council of the Chinese Mathematics Society (1991-1999). He is an IEEE Fellow, IFAC Fellow, a Member of TWAS, and a Member of CAS. Wenxiao Zhao received the B.Sc. degree from Shandong University, China in 2003 and the Ph.D. degree in operation research and cybernetics from the Institute of Systems Science (ISS), Academy of Mathematics and Systems Science (AMSS), Chinese Academy of Sciences (CAS), China, in 2008. He is currently a professor with AMSS, CAS. His research interests are mainly in machine learning, system identification and adaptive control, and stochastic optimization.

Funding

This work was supported by the National Key Research and Development Program of China under Grant 2018YFA0703800 and the National Natural Science Foundation of China under Grant 61822312. This work was also supported (in part) by the Strategic Priority Research Program of Chinese Academy of Sciences under Grant No. XDA27000000.

Availability of data and materials

Not applicable.

Declarations**Competing interests**

The authors declare no competing interests.

Received: 22 January 2021 Accepted: 2 March 2021

Published online: 17 August 2021

References

1. A. Jadbabaie, J. Lin, A. S. Morse, Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Trans. Autom. Control.* **48**(6), 988–1001 (2003)
2. W.ei. Ren, R. W. Beard, Consensus seeking in multiagent systems under dynamically changing interaction topologies. *IEEE Trans. Autom. Control.* **50**(5), 655–661 (2005)
3. R. Olfati-Saber, A. J. Fax, R. M. Murray, Consensus and cooperation in networked multi-agent systems. *Proc. IEEE.* **95**(1), 215–233 (2007)
4. W. Feng, H.-F. Chen, Output consensus of networked hamstein and wiener systems. *SIAM J. Control Optim.* **57**(2), 1230–1254 (2019)
5. P. Yi, Y. Hong, F. Liu, Initialization-free distributed algorithms for optimal resource allocation with feasibility constraints and application to economic dispatch of power systems. *Automatica.* **74**, 259–269 (2016)
6. A. Nedić, A. Olshevsky, W. Shi, in *2018 IEEE Conf. Decis. Control (CDC)*. Improved convergence rates for distributed resource allocation (IEEE, 2018), pp. 172–177. <https://doi.org/10.1109/cdc.2018.8619322>
7. Y. Kuriki, T. Namerikawa, in *2014 Amer. Control Conf.* Consensus-based cooperative formation control with collision avoidance for a multi-uav system (IEEE, 2014), pp. 2077–2082. <https://doi.org/10.1109/acc.2014.6858777>
8. S. Aeron, V. Saligrama, D. A. Castanon, Efficient sensor management policies for distributed target tracking in multihop sensor networks. *IEEE Trans. Sig. Process.* **56**(6), 2562–2574 (2008)
9. A. Nedic, A. Ozdaglar, Distributed subgradient methods for multi-agent optimization. *IEEE Trans. Autom. Control.* **54**(1), 48 (2009)
10. A. Nedic, A. Ozdaglar, P. A. Parrilo, Constrained consensus and optimization in multi-agent networks. *IEEE Trans. Autom. Control.* **55**(4), 922–938 (2010)
11. A. Simonetto, G. Leus, *Distributed asynchronous time-varying constrained optimization*. (IEEE, 2014), pp. 2142–2146. <https://doi.org/10.1109/acssc.2014.7094854>
12. F. Y. Jakubiec, A. R. D-map, Distributed maximum a posteriori probability estimation of dynamic systems. *IEEE Trans. Sig. Process.* **61**(2), 450–466 (2012)
13. X. Yi, X. Li, L. Xie, K. H. Johansson, Distributed online convex optimization with time-varying coupled inequality constraints. *IEEE Trans. Sig. Process.* **68**, 731–746 (2020)
14. S. Shahrampour, A. Jadbabaie, Distributed online optimization in dynamic environments using mirror descent. *IEEE Trans. Autom. Control.* **63**(3), 714–725 (2017)
15. D. Bajovic, J. Xavier, B. Sinopoli, J. M. F. Moura, et al., Distributed detection via gaussian running consensus: Large deviations asymptotic analysis. *IEEE Trans. Sig. Process.* **59**(9), 4381–4396 (2011)
16. M. Ye, H. Guoqiang, in *2015 54th IEEE Conference on Decision and Control (CDC)*. Distributed optimization for systems with time-varying quadratic objective functions (IEEE, 2015), pp. 3285–3290. <https://doi.org/10.1109/cdc.2015.7402713>
17. A. Simonetto, A. Koppel, A. Mokhtari, G. Leus, A. Ribeiro, Decentralized prediction-correction methods for networked time-varying convex optimization. *IEEE Trans. Autom. Control.* **62**(11), 5724–5738 (2017)
18. M. Fazlyab, S. Paternain, V. M. Preciado, A. Ribeiro, Prediction-correction interior-point method for time-varying convex optimization. *IEEE Trans. Autom. Control.* **63**(7), 1973–1986 (2017)
19. H.-F. Chen, T. E. Duncan, B. Pasik-Duncan, A kiefer-wolfowitz algorithm with randomized differences. *IEEE Trans. Autom. Control.* **44**(3), 442–453 (1999)
20. H. J. Kushner, G. Yin, Asymptotic properties of distributed and communicating stochastic approximation algorithms. *SIAM J. Control Optim.* **25**(5), 1266–1290 (1987)
21. P. Bianchi, G. Fort, W. Hachem, Performance of a distributed stochastic approximation algorithm. *IEEE Trans. Inform. Theory.* **59**(11), 7405–7418 (2013)
22. J. Lei, H.-F. Chen, Distributed stochastic approximation algorithm with expanding truncations. *IEEE Trans. Autom. Control.* **65**(2), 664–679 (2020)
23. V. Dupač, A dynamic stochastic approximation method. *Ann. Math. Stat.*, 1695–1702 (1965)
24. H.-F. Chen, K. Uosaki, Convergence analysis of dynamic stochastic approximation. *Syst. Control Lett.* **35**(5), 309–315 (1998)
25. D. Acemoglu, A. Nedic, A. Ozdaglar, in *2008 47th IEEE Conference on Decision and Control*. Convergence of rule-of-thumb learning rules in social networks (IEEE, 2008), pp. 1714–1720. <https://doi.org/10.1109/cdc.2008.4739167>
26. S. Shahrampour, S. Rakhlin, A. Jadbabaie, in *Advances in Neural Information Processing Systems*. Online learning of dynamic parameters in social networks, (2013). <https://doi.org/10.4018/978-1-4666-1815-2.ch006>
27. F. Kewei, H.-F. Chen, W. Zhao, in *2018 37th Chinese Control Conference (CCC)*. Distributed stochastic approximation algorithm for time-varying regression function over network (IEEE, 2018), pp. 1925–1930. <https://doi.org/10.23919/chicc.2018.8483554>
28. H.-F. Chen, *Stochastic Approximation and Its Applications*, vol. 64. (Springer Science & Business Media, 2002). <https://doi.org/10.1007/b101987>

29. X. Yuan, C. Han, Z. Duan, M. Lei, in *2005 7th International Conference on Information Fusion*. Comparison and choice of models in tracking target with coordinated turn motion, vol. 2 (IEEE, 2005), p. 6. <https://doi.org/10.1109/icf.2005.1592032>
30. M. O. Jackson, *Social and Economic Networks*. (Princeton university press, US, 2010)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.