ASYNCROUS STOCHASTIC APPROXIMATION ALGORITHMS
FOR NETWORKED SYSTEMS: REGIME-SWITCHING
TOPOLOGIES AND MULTISCALE STRUCTURE

G. YIN†, QUAN YUAN†, AND LE YI WANG‡

Abstract. This work develops asynchronous stochastic approximation (SA) algorithms for
networked systems with multiagents and regime-switching topologies to achieve consensus control.
There are several distinct features of the algorithms: (1) In contrast to most of the existing consensus
algorithms, the participating agents compute and communicate in an asynchronous fashion without
using a global clock. (2) The agents compute and communicate at random times. (3) The regime-
switching process is modeled as a discrete-time Markov chain with a finite state space. (4) The
functions involved are allowed to vary with respect to time; hence, nonstationarity can be handled.
(5) Multiscale formulation enriches the applicability of the algorithms. In the setup, the switching
process contains a rate parameter \( \varepsilon > 0 \) in the transition probability matrix that characterizes
how frequently the topology switches. The algorithm uses a step-size \( \mu \) that defines how fast the
network states are updated. Depending on their relative values, three distinct scenarios emerge.
Under suitable conditions, it is shown that a continuous-time interpolation of the iterates converges
weakly either to a system of randomly switching ordinary differential equations modulated by a
continuous-time Markov chain or to a system of differential equations (an average with respect to
certain measure). In addition, a scaled sequence of tracking errors converges to either a switching
diffusion or a diffusion. Simulation results are presented to demonstrate these findings.

Key words. asynchronous algorithm, stochastic approximation, switching model, random com-
putation time, nonstationarity, consensus

AMS subject classifications. 60F17, 62L20, 65Y05, 93E10, 93E25

DOI. 10.1137/120871614

1. Introduction. This paper develops consensus algorithms under the asynchronous
communication and random computation environments using random switching
topologies. Consensus problems are related to many control applications that involve
coordination of multiple entities with only limited neighborhood information to reach
a global goal for the entire team. Since the mid 1990s, there have been increasing and
resurgent efforts devoted to the study of consensus controls of multiagent systems.
The goal is to achieve a common theme such as position, speed, load distribution,
etc. for the mobile agents. In [28], a discrete-time model of autonomous agents was
proposed, which can be viewed as points or particles moving in the plane with the
same speed but with different headings. Each agent updates its heading using a local
rule based on the average headings of its own and its neighbors. This is in fact a
special version of a model introduced in [24] for simulating animation of flocking and
schooling behaviors. Technically, the problems considered are related to the parallel
computation model considered in [26], which was substantially generalized in [12];
see also related works in [1, 4, 6, 16, 17, 18, 19, 22, 35]. During the past decades, a
host of researchers have devoted their efforts to the study of the consensus problems;

Received by the editors March 28, 2012; accepted for publication (in revised form) May 29,
2013; published electronically August 15, 2013. This research was supported in part by the National
Science Foundation under DMS-1207667.
http://www.siam.org/journals/mms/11-3/87161.html
†Department of Mathematics, Wayne State University, Detroit, MI 48202 (gyin@math.wayne.edu,
quanyuan@wayne.edu).
‡Department of Electrical and Computer Engineering, Wayne State University, Detroit, MI 48202
(lywang@wayne.edu).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
see [7, 8, 9, 10, 15, 20, 21, 23, 25, 29] and the many references therein. Many results obtained thus far are for simple dynamic systems with fixed or highly simplified time-varying topologies, whereas [10], [33], and [34] dealt with time-varying topologies under Markovian switching.

In practical implementations of consensus or coordinated control schemes, control actions are almost always done asynchronously, especially over a large network of subsystems. For instance, subsystems operate independently with different clocks until they communicate with their neighbors; communication channels operate according to priorities and hence transmit data at different paces and at different times; even for data packets transmitted at the same time from a node system, their pathways through different routes and hubs introduce different latencies and hence arrive at different times. This is especially true in mobile agents when obstacles from terrains create interruptions, packet losses, and delays so that consensus must be performed on delayed information, which is an asynchronous operation.

In this paper, our problems are formulated to capture two aspects of the asynchronism: (i) Asynchronous execution of state updates at the subsystems: Each subsystem has a randomized timer representing the internal processing time. A subsystem can update its state only when the timer ticks. After the state update, the timer is renewed and internal processing resumes until the next ticking time. (ii) Asynchronous neighborhood information exchange: When a subsystem’s timer ticks, the subsystem will observe the states of its neighboring subsystems at that time and adjust its own state accordingly. Since the neighboring subsystems update their states independently, the received state information will always be a delayed information, creating another layer of asynchronism. These concepts are illustrated in Figure 1.

![Asynchronous operations and communications of networked subsystems](image-url)
This asynchronous framework introduces fundamental challenges to constrained consensus control problems. In the field of consensus control, most works are on unconstrained consensus; namely, as long as the states of the subsystems achieve consensus there are no other constraints to be satisfied. However, practical systems often impose constraints on the states. For example, for power grids, all power produced will have to be equal to the total load at steady state, even though transient power imbalance is allowed due to storage capabilities on generators and/or capacitance and inductance on transmission lines. In a team formation for area surveillance, a team of mobile sensors needs to be confined to the region to be covered. In parallel or cloud computing, steady-state service demands and service capacity must be equal. In our previous work [33, 34], this constraint is satisfied by employing a “link control” strategy in which a reduction on a state value is always balanced by an equal amount of increase in its neighboring subsystems.\(^1\) Asynchronous operation renders such a control scheme impractical. Further complication stems from the time-varying nature of network topologies, which changes a subsystem’s neighbors randomly. Consequently, the interaction between the stochastic processes of subsystem timers and the governing Markov chain for the network topologies must be carefully studied. In this work, we employ a new control strategy in which the state constraint can be asymptotically satisfied even though the asynchronous operation leaves the constraint unmet during the transient period.

Dealing with large interconnected systems such as in a communication network with multiple servers, it is natural to consider the distributed, asynchronous stochastic approximation (SA) algorithms. If synchronous SA algorithms are used, a new iteration will not begin until the current iteration is finished in all subsystems. Since the dimension of a networked system can be very large, the waiting time on subsystems will cause serious time delays. In [26], an asynchronous algorithm was proposed where separate processors iterated on the same system vector (with possibly different noise processes and/or different dynamics) and shared information in an asynchronous way. In [12, 13, 40], SA algorithms for parallel and distributed processing were further developed. The main efforts were on the study of convergence and rates of convergence of such algorithms.

In this paper, we concentrate on consensus-type algorithms. Here, each component in a system (with a large number of mobile agents) can be handled by different agents, and the information can be shared by agents. To each single agent, the agent can start the next iteration using the newest information of iteration on other components without waiting for other agents to finish. So for each component, the time of each iteration and the number of iterations up to that moment are random. We note that the underlying problems introduce some new challenges, and our solutions carry a number of new features. When representing the algorithms as discrete-time dynamic systems, the dynamics of the systems switch randomly among a finite number of regimes and at random times. The modulating force of the switching process is modeled as a discrete-time Markov chain with a finite state space. In our setup, the transition probability matrix of the Markov chain includes a small parameter \(\varepsilon\).

\(^1\)The constrained consensus control problems are motivated by load sharing and resource allocation problems. When node \(i\) estimates the state of node \(j\) and decides to shift a resource of amount \(u_{ij}\) to node \(j\), this will be a reduction on node \(i\) and an increase of the equal amount to node \(j\). In this sense, both node \(i\) and node \(j\) are controlled. But the decision resides with node \(i\), and node \(j\) receives it passively. In synchronized operation, this will guarantee that the sum of all states is a constant at all time. In contrast, in asynchronous modes, state updates occur at different times, and, as a result, the sum of the states may not be a constant during the transient period.
Henceforth, this parameter will be called the transition frequency parameter since it represents how frequently the state transition will take place. The SA algorithm defines its updating speed by another small parameter \( \mu \), which will be called the adaptation step-size. The interplay of the two parameters introduces a multiscale system dynamics. It turns out that the difference between the parameters \( \varepsilon = O(\mu) \), \( \varepsilon \ll \mu \), and \( \mu \ll \varepsilon \) gives rise to qualitatively different behaviors with stark contrasts.

To summarize, there are several novel features of the algorithms proposed in this paper: (1) In contrast to most of the existing consensus algorithms, the participating agents compute and communicate in an asynchronous fashion. (2) Based on their local clocks, the agents compute and communicate at random times without using a global clock. (3) The regime-switching process is modeled as a discrete-time Markov chain with a finite state space. (4) The functions involved are allowed to vary with respect to time; hence, nonstationarity can be handled. (5) Multiscale formulation enriches the applicability of the algorithms.

The rest of the paper is arranged as follows. Section 2 begins with the formulation of a typical consensus control problem for networked systems under randomly switching topologies. It serves to demonstrate how this problem naturally leads to asynchronous SA algorithms under switching dynamics. Mathematical formulation of the problem is then presented accordingly. Section 3 focuses on the case \( \varepsilon = O(\mu) \) to introduce new techniques in establishing asymptotic behavior of the algorithms. Using weak convergence methods, convergence of the algorithm is obtained. The limit behavior of the scaled estimation errors is also analyzed. Section 4 extends the main techniques of section 3 to the cases of \( \varepsilon \ll \mu \) and \( \mu \ll \varepsilon \). It is shown that, depending on relative scales between the transition frequency and adaptation step-size, the asynchronous SA algorithms demonstrate fundamentally different asymptotic behaviors. Section 5 illustrates the main findings of this paper by simulation examples. Section 6 provides further remarks and discusses some open issues.

2. Formulation. Throughout this paper, \( |\cdot| \) denotes a Euclidean norm. A point \( x \) in a Euclidean space is a column vector; the \( i \)th component of \( x \) is denoted by \( x^i \), and \( \mathbb{1} \) denotes transpose. The notation \( O(y) \) denotes a function of \( y \) satisfying \( \sup_y |O(y)|/|y| < \infty \). Likewise, \( o(y) \) denotes a function of \( y \) satisfying \( |o(y)|/|y| \to 0 \), as \( y \to 0 \). In particular, \( O(1) \) denotes the boundedness and \( o(1) \) indicates convergence to 0. To facilitate the reading, we have placed some basic formulation for consensus control algorithms in an appendix. Our formulation extends beyond the traditional setup. In lieu of the simple formulation in the appendix, we allow certain nonadditive noises to be added. More importantly, our main effort is on asynchronous computation and communication schemes. In lieu of the constraint \( \mathbb{1}^T x_n = \eta r \) at each step, we only require such an equality to hold asymptotically. This generalizes the setup in the appendix. Suppose that the network topology is represented by a graph \( G \). In contrast to the standard setting, the graph depends on a discrete-time Markov chain so it is given by \( G = G(\tilde{\alpha}_n) \). In our setup, the graph can take \( m_0 \) possible values. The Markov chain is used to model, for example, capacity of the network, random environment, and other random factors such as interrupts, rerouting of communication channels, etc. Thus \( G(\tilde{\alpha}_n) = \sum_{\ell=1}^{m_0} G(\ell) I(\tilde{\alpha}_{n-1} = \ell) \). To illustrate, suppose that initially the Markov chain is at \( \tilde{\alpha}_0 = \iota \). Then the graph takes the value \( G(\iota) \). At a random instance \( \rho_1 \), the first jump of the Markov chain takes place so that \( \tilde{\alpha}_{\rho_1} = \ell \neq \iota \). Then the graph switches to \( G(\ell) \) and holds that value for a random duration until the next jump of the Markov chain takes place, etc.
To carry out the recursive computational task, we consider a class of asynchronous and distributed algorithms in the following setup. Suppose that the state \( x \in \mathbb{R}^r \) and that there are \( r \) processors participating in the computational task. For notational simplicity, we assume that each processor handles only one component. It is clear that this can be made substantially more general by allowing each processor to handle a vector of possibly different dimensions. However, the mathematical framework will be essentially the same albeit the complex notation. Suppose that for each \( i \leq r \), \( \{Y^n_i\} \) is a sequence of positive integer-valued random variables (assuming the random sequence to be positive integer-valued is for notational convenience) that are generally state and data dependent such that the \( n \)th iteration of processor \( i \) takes \( Y^n_{n-1} \) units of time. Define a sequence of “renewal-type” random computation times \( \tau^n_i \) as

\[
\tau^n_0 = 0, \quad \tau^n_{i+1} = \tau^n_i + Y^n_i.
\]

For each \( i \), the sequence \( \{Y^n_i\} \) is an interarrival time and \( \{\tau^n_i\} \) is the corresponding “renewal” time. It is well known that \( \tilde{\alpha}_n \) is strongly Markov, so \( \tilde{\alpha}_{\tau^n_i} \) is a Markov chain.

Using constant step-size \( \mu > 0 \), we consider the asynchronous algorithm

\[
x^n_{\tau^n_{i+1}} = x^n_{\tau^n_i} + \mu[M_{\tau^n_i}(\tilde{\alpha}_{\tau^n_i})x^n_{\tau^n_i}]^i + \mu[W_{\tau^n_i}(\tilde{\alpha}_{\tau^n_i})\xi^n_{\tau^n_i}]^i + \mu\tilde{W}^i_{\tau^n_i}(x^n_{\tau^n_i}, \tilde{\alpha}_{\tau^n_i}, \zeta^n_{\tau^n_i}), \quad i \leq r,
\]

where \( \xi^n_{\tau^n_i} \in \mathbb{R}^r \) and \( \zeta^n_{\tau^n_i} \in \mathbb{R}^r \) are the noise sequences incurred in the \((n+1)\)st iteration. Note that the functions involved are time dependent. We use the same idea as in the setup of a fixed configuration in the appendix, but allow more general structure. Note also that for each \( n \) and \( \alpha \in \mathcal{M} \), \( M_n(\alpha) \) is not a generator of a Markov chain as the fixed \( M \) discussed in the appendix. We allow the nonadditive noise to be used. When \( M_n(\alpha) = M \) and \( W_n(\alpha) = W \) are constant matrices being generators of continuous Markov chains for all \( n \) and all \( \alpha \in \mathcal{M} \), and \( W_n = 0 \), (2.2) reduces to the existing standard consensus algorithm with additive noise. The nonadditive portion is a general nonlinear function of the analog state \( x \), the Markov chain state \( \alpha \in \mathcal{M} \), the noise source \( \zeta \), as well as \( n \).

The noise sequences are “exogenous” in that (loosely speaking) the distribution of their future evolution, conditioned on their past, does not change if we also condition on the past of the state values. The distribution of the computation interval \( Y^n_i \) is allowed to depend on the state \( x^n_{\tau^n_i} \) and the noise \( \xi^n_{\tau^n_i} \) that is used during that \((n+1)\)st interval in the \( i \)th processor. We define

\[
N_i(n) = \sup\{j : \tau^n_j \leq n\}, \quad \Delta^n_i = n - \tau^n_i, \quad \xi^n_i = \xi^n_{\tau^n_i}, \quad \zeta^n_i = \zeta^n_{\tau^n_i}, \quad \text{for } n \in [\tau^n_j, \tau^n_{j+1}),
\]

\[
x_n = (x^n_{\tau^n_1}, \ldots, x^n_{\tau^n_{N_i(n)}}), \quad \text{where } \bar{x}_n \equiv (x^n_{\tau^n_{N_i(n)+1}}, \ldots, x^n_{\tau^n_{N_i(n)}})^t,
\]

and with a slight abuse of notation, denote \( \alpha_n = \tilde{\alpha}_{\tau^n_{N_i(n)}} \). Note that \( \Delta^n_i = 0 \) if \( n \) is a renewal time for processor \( i \) (\( \Delta^n_{i,0} \) is the time elapsed since the start of a new computation for processor \( i \)). The \( \bar{x}_n \) is an aggregate vector of dimension \( r \cdot r \) and \( \bar{x}_n \) is the state value used for the \( i \)th processor at real time \( n \). We can now write

\[
x^n_{\tau^n_{i+1}} = x^n_{\tau^n_i} + \mu[M_n(\alpha_n)\bar{x}_n]^iI^n_{n+1} + \mu[W_n(\alpha_n)\xi^n_i]^iI^n_{n+1} + \mu\tilde{W}^i_n(\bar{x}_n, \alpha_n, \zeta^n_i)I^n_{n+1},
\]

where \( I^n_0 = I(\Delta^n_i=0) \). If \( I^n_i = 1 \), then \( n \) is a random computation time for the \( i \)th processor. Note that the dependence of \( \tilde{\alpha}_{\tau^n_{N_i(n)}} \) is only through the random computation time \( \tau^n_j \) with \( j = N_i(n) \). Thus in the notation of \( \alpha_n \), we suppressed the \( i \).
dependence for notational simplicity in the subsequent calculation; this information is also reflected from $I_n'$. In what follows, for each $i$, the mixing measures defined in (A2), namely, $\phi$ and $\psi$, should be $i$ dependent. Nevertheless, to simplify the notation, instead of writing $\phi^i$ and $\psi^i$, we will use $\phi$ and $\psi$ throughout the rest of the paper. We assume that the following conditions hold:

(A1) The process $\hat{\alpha}_n$ is a discrete-time Markov chain with a finite state space $\mathcal{M} = \{1, \ldots, m_0\}$ and transition probability matrix

\[
P = P^e = I + \varepsilon Q,
\]

where $\varepsilon > 0$ is a small parameter, $I$ is an $m_0 \times m_0$ identity matrix, and $Q = (q_{ij}) \in \mathbb{R}^{m_0 \times m_0}$ is the generator of a continuous-time Markov chain, (i.e., $Q$ satisfies $q_{ij} \geq 0$ for $i \neq j$; $\sum_{j=1}^{m_0} q_{ij} = 0$ for each $i = 1, \ldots, m_0$) such that $Q$ is irreducible.

(A2) (a) (i) The function $\tilde{\phi}_n$ is uniformly bounded on $\mathcal{M}$, where $\tilde{\phi}_n(x, \psi) = \tilde{\phi}_n(x, \psi^{\psi})$.

(b) Each $\tilde{\phi}_n(x, \psi)$ is continuous in $x$ uniformly in each bounded $(x, \psi)$ set such that $E_{\tau_{j}} \psi^{\psi^{\psi}} = \tilde{\phi}^{\psi^{\psi}}(x, \tau_{j}, \psi)$. The conditional expectation on $\mathcal{F}_m = \{x_0, \tau_{j}, \psi^{\psi^{\psi}} j = 0, \ldots, m\}$ is uniformly bounded.

(c) For each $i \in \mathcal{M}$, $\{M_i(x)\}$ and $\{W_i(x)\}$ are uniformly bounded. For each $i \in \mathcal{M}$, there is an $\tilde{\mathcal{M}}(i)$ such that for each $m$, $\sum_{j=m}^{m+n-1} M_j(i)/n \rightarrow \tilde{\mathcal{M}}(i)$, where $\tilde{\mathcal{M}}(i)$ is an irreducible generator of a Markov chain for each $i \in \mathcal{M}$.

Remark 2.1. (a) Note that as a consequence of (A2), for any positive integer $m$ and fixed $i$,

\[
\frac{1}{n} \sum_{j=m}^{m+n-1} E_m^{i} \tilde{W}_j(x, \tau_{j}^{\psi^{\psi^{\psi}}}) \rightarrow 0 \text{ in probability},
\]

\[
\frac{1}{n} \sum_{j=m}^{m+n-1} E_m^{i} \tilde{E}_j(x, \tau_{j}^{\psi^{\psi^{\psi}}}) \rightarrow 0 \text{ in probability}.
\]
ASYNCHRONOUS SA ALGORITHMS FOR NETWORKED SYSTEMS 819

(b) It is reasonable to assume that $\pi^i(\cdot, \iota)$ is strictly positive. This is essentially a suitably scaled limit of the mean of $Y^i_n$. Under the standard renewal setup with independent and identically distributed (i.i.d.) interarrival $Y^i_n$ (independent of data), it is simply a positive constant, the mean of $Y^i_n$.

(c) Note that (2.4) is an SA-type algorithm, but more difficult to analyze because of the switching topologies. In the traditional setup of SA problems, the limit or the averaged system is an ordinary differential equation (ODE). Very often these limits are autonomous. Even if they are time inhomogeneous ODEs, these equations are nonrandom. In certain cases treated here, the limit is no longer an ODE, but a randomly varying ODE subject to switching. In the literature of SA, the rate of convergence study is normally associated with a limit stochastic differential equation (SDE). In our case, some of the limits are Markovian-switching SDEs (i.e., switching diffusions [39]).

In the next two sections, three possibilities concerning the relative sizes of $\varepsilon$ and $\mu$ are analyzed. This idea also appears in related treatments of LMS (least mean squares)-type algorithms under regime-switching dynamic systems; see [30, 31, 32].

In treating the three different cases, careful analysis is needed to examine convergence, stability, and related consensus issues.

3. Asymptotic properties: $\varepsilon = O(\mu)$. This section concentrates on the case $\varepsilon = O(\mu)$. For notational simplicity and concreteness, in what follows, we simply consider $\varepsilon = \mu$ in this section. More general cases can be considered; they do not add further technical difficulties. The results will be similar in spirit.

3.1. Basic properties. To proceed, we first present a moment estimate for the recursive algorithm (2.4). Throughout the paper, we use $K$ to denote a generic positive constant with the conventions $K + K = K$ and $KK = K$. We also use $K_T$ to denote a generic positive constant that depends on $T$ (whose value may change for different appearances).

Lemma 3.1. Under assumption (A1), for any $0 < T < \infty$ and each $i = 1, \ldots, r$,

$$
\sup_{0 \leq n \leq T/\varepsilon} E|x^i_n|^2 \leq K_T \exp(T) < \infty.
$$

Proof. Note that for any $0 < T < \infty$ and $0 \leq n \leq T/\mu$, by the Cauchy–Schwarz inequality,

$$
\mu^2 E \sum_{k=0}^n |M_k(\alpha_k)\bar{x}_k^{i_1}I_{k+1}|^2 \leq \mu^2 (n + 1) E \sum_{k=0}^n \left| M_k(\alpha_k) \right|^2 \left| \bar{x}_k^{i_1} I_{k+1} \right|^2
$$

$$
\leq K_T \mu \sum_{k=0}^n E \left| \bar{x}_k^{i_1} I_{k+1} \right|^2,
$$

where $K_T > 0$. Likewise,

$$
\mu^2 E \sum_{k=0}^n \left| \hat{W}_k(\bar{x}_k^{i_1}, \alpha_k, \zeta_k) \right|^2 \leq K_T \mu \sum_{k=0}^n \left| \bar{x}_k^{i_1} I_{k+1} \right|^2 + K_T,
$$

$$
\mu^2 E \sum_{k=0}^n \left| W_k(\alpha_k)\bar{\zeta}_k^{i_1}I_{k+1} \right|^2 \leq K(\mu n)^2 \leq K_T.
$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
Iterating on $E|x_{n+1}^i|^2$ with the use of (2.4) and using (3.1) and (3.2), we obtain

$$E|x_{n+1}^i|^2 \leq (E|x_0^i|^2 + K_T) + K_T \mu \sum_{k=0}^{n} E\left|\tilde{x}_k^i I_{k+1}\right|^2$$

(3.3)

$$\leq (E|x_0^i|^2 + K_T) + K_T \mu \sum_{k=0}^{n} E|x_k^i|^2 + O(\mu).$$

Then by Gronwall’s inequality,

$$E|x_{n+1}^i|^2 \leq K_T \exp(n\mu) \leq K_T \exp(\mu(T/\mu)) \leq K_T \exp(T).$$

Taking sup over $n$, the desired estimate follows. \[\square\]

### 3.2. Convergence.

This section is devoted to obtaining asymptotic properties of algorithm (2.4). Before proceeding further, we state a result on estimation error bounds. The proof of the assertion on probability distributions is essentially in that of Theorems 3.5 and 4.3 of [36], whereas the proof of weak convergence of $\alpha^\varepsilon(\cdot)$ can be found in [38]; see also [37]. Thus the proof is omitted.

**Lemma 3.2.** Under condition (A2), with $P^\varepsilon$ given by (2.5), denote the $n$-step transition probability by $(P^\varepsilon)^n$ and $p_n^\varepsilon = (P(\alpha_n = 1), \ldots, P(\alpha_n = m_0))$, and define $\alpha^\varepsilon(t) = \tilde{\alpha}_n$ for $t \in [n\varepsilon, n\varepsilon + \varepsilon]$. Then the following claims hold:

$$p_n^\varepsilon = p(t) + O(\varepsilon + e^{-k_0 t/\varepsilon}),$$

$$\sum_{t=0}^{n} p_n^\varepsilon = \Xi(\varepsilon n, \varepsilon n_0) + O(\varepsilon + e^{-k_0 (n-n_0)}),$$

where $p(t) \in \mathbb{R}^{1 \times m_0}$ and $\Xi(t, t_0) \in \mathbb{R}^{m_0 \times m_0}$ are the continuous-time probability vector and transition matrix satisfying

$$\frac{dp(t)}{dt} = p(t)Q, \quad P(0) = p_0,$$

$$\Xi(t, t_0) = \Xi(t, t_0)Q, \quad \Xi(t_0, t_0) = I,$$

with $t_0 = \varepsilon n_0$ and $t = \varepsilon n$.

Moreover, $\alpha^\varepsilon(\cdot)$ converges weakly to $\alpha(\cdot)$, a continuous-time Markov chain generated by $Q$.

Since we consider $\varepsilon = O(\mu)$, without loss of generality, we take $\varepsilon = \mu$ in what follows. The next lemma concerns the property of the algorithm as $\mu \to 0$ through an appropriate continuous-time interpolation. We define

$$x^\mu(t) = x_n, \quad \alpha^\mu(t) = \tilde{\alpha}_n \quad \text{for} \quad t \in [\mu n, \mu n + \mu).$$

Then $(x^\mu(\cdot), \alpha^\mu(\cdot)) \in D([0, T] : \mathbb{R}^r \times \mathcal{M})$, which is the space of functions that are defined on $[0, T]$ taking values in $\mathbb{R}^r \times \mathcal{M}$, and that are right continuous and have left limits endowed with the Skorohod topology [14, Chapter 7]. Before proceeding further, we first state a lemma that gives the uniqueness of the solution of (3.6).

**Lemma 3.3.** The switched ODE

$$\frac{dx^i(t)}{dt} = \frac{[\Pi(\alpha(t))x(t)]^i}{\pi^i(x(t), \alpha(t))}, \quad i = 1, \ldots, r,$$

(3.6)

has a unique solution for each initial condition $(x(0), \alpha(0))$ with $x(0) = (x_0^1, \ldots, x_0^r)'$. 
Proof. For any \( f(\cdot, \cdot) : \mathbb{R}^r \times \mathcal{M} \to \mathbb{R} \) satisfying, for each \( t \in \mathcal{M} \), \( f(\cdot, t) \in C_0^1 \) (space of continuously differentiable functions with compact support), \( \mathcal{L}_1 \) is defined as

\[
\mathcal{L}_1 f(x, t) = \sum_{i=1}^{r} \frac{\partial f(x, t)}{\partial x_i} \frac{[M(t)x]^i}{\pi^i(x, t)} + Qf(x, \cdot)(t), \quad t \in \mathcal{M},
\]

where

\[
Qf(x, \cdot)(t) = \sum_{\ell=1}^{m_0} q_{\ell} f(x, \ell).
\]

Let \( (x(t), \alpha(t)) \) be a solution of the martingale problem with operator \( \mathcal{L}_1 \) defined in (3.7). We proceed to show that the solution is unique in the sense of distribution. Define

\[
g(x, k) = \exp(\gamma' x + \gamma_0 k) \quad \text{for all } \gamma \in \mathbb{R}^r, \gamma_0 \in \mathbb{R}, k \in \mathcal{M}.
\]

Consider \( \psi_{jk}(t) = E[I_{\{\alpha(t)=j\}}g(x(t), k)], j, k \in \mathcal{M} \). It is readily seen that \( \psi_{jk}(t) \) is the characteristic function associated with \( (x(t), \alpha(t)) \). By virtue of the Dynkin's formula,

\[
\psi_{j_0k_0}(t) - \psi_{j_0k_0}(0) - \int_0^t \mathcal{L}_1 \psi_{j_0k_0}(s) ds = 0,
\]

where

\[
\mathcal{L}_1 \psi_{j_0k_0}(s) = \sum_{i=1}^{m_0} \gamma_i \frac{[M(k_0)x]^i}{\pi^i(x(s), \alpha(s))} \psi_{j_0k_0}(s) + \sum_{j=0}^{m_0} q_{j_0} \psi_{j_0k_0}(s).
\]

Let \( \psi(t) = (\psi_{i,\ell}(t) : i \leq m_0, \ell \leq m_0) \). Combining (3.8) and (3.9), we obtain

\[
\psi(t) = \psi(0) + \int_0^t G\psi(s) ds,
\]

where \( G \) is an \( m_0 \times m_0 \) matrix. Thus (3.10) is an ODE with an initial condition \( \psi(0) \).

As a result, it has a unique solution. \( \square \)

By Lemmas 3.1–3.3, we can obtain the following theorem.

**Theorem 3.4.** Under (A1) and (A2), \( \{x^\mu(\cdot), \alpha^\mu(\cdot)\} \) is tight in \( D([0, T] : \mathbb{R}^r \times \mathcal{M}) \). Assume that \( x_0 \) and \( \alpha_0 \) are independent of \( \mu \) and are nonrandom without loss of generality. Then \( \{x^\mu(\cdot), \alpha^\mu(\cdot)\} \) converges weakly to \( (x(\cdot), \alpha(\cdot)) \), which is a solution of (3.6) with initial condition \( (x_0, \alpha_0) \).

**Proof.** (a) **Tightness.** The tightness of \( \{\alpha^\mu(\cdot)\} \) can be proved as in the proof of [36, Theorem 4.3]. So we need only prove the tightness of \( \{x^\mu(\cdot)\} \); i.e., we need only prove that \( \{x^\mu_{i,\ell}(\cdot)\} \) is tight for each \( i \).

For any \( \delta > 0 \), let \( t > 0 \) and \( s > 0 \) such that \( s \leq \delta \), and let \( t, t + \delta \in [0, T] \). Note that

\[
x^\mu_{i,\ell}(t + s) - x^\mu_{i,\ell}(t) = \mu \sum_{k=t/\mu}^{(t+s)/\mu-1} \left[ M_k(\alpha_k)x^i_k \right] I^i_{k+1} + \mu \sum_{k=t/\mu}^{(t+s)/\mu-1} \left[ W_k(\alpha_k)\xi^i_k \right] I^i_{k+1} + \mu \sum_{k=t/\mu}^{(t+s)/\mu-1} \left[ \hat{W}_k(x^i_k, \alpha_k, \xi^i_k) \right] I^i_{k+1}.
\]
As above, hereafter we use the conventions that \( t/\mu \) and \((t+s)/\mu\) denote the corresponding integer parts, i.e., \( \lfloor t/\mu \rfloor \) and \( \lfloor (t+s)/\mu \rfloor \), respectively. For notational simplicity, in what follows we will not use the floor function notation unless it is necessary.

Since \( \tilde{\alpha}_k \) is a finite state Markov chain, by (A1) \( |M_k(\alpha_k)| \) and \( |W_k(\alpha_k)| \) (see the notation \( \alpha_k \) in (2.3)) are uniformly bounded. Using the Cauchy–Schwarz inequality as in (3.1) (with \( \sum_{k=0}^{n} \) replaced by \( \sum_{k=\lfloor t/\mu \rfloor}^{\lfloor (t+s)/\mu \rfloor - 1} \)), together with Lemma 3.1, we get

\[
E_t^{\mu} \left| x^{\mu,i}(t + s) - x^{\mu,i}(t) \right|^2 \leq K \mu^2 E_t^{\mu} \left[ \sum_{k=\lfloor t/\mu \rfloor}^{\lfloor (t+s)/\mu \rfloor - 1} |M_k(\alpha_k)\tilde{x}_k^i| I_{k+1}^i \right]^2 + \sum_{k=\lfloor t/\mu \rfloor}^{\lfloor (t+s)/\mu \rfloor - 1} |W_k(\alpha_k)\tilde{\xi}_k^i| I_{k+1}^i \]
\[
+ K \mu^2 E_t^{\mu} \left[ \sum_{k=\lfloor t/\mu \rfloor}^{\lfloor (t+s)/\mu \rfloor - 1} |\hat{W}_k(\tilde{x}_k^i, \alpha_k, \tilde{\xi}_k^i)| I_{k+1}^i \right]^2 \leq K \mu s \sup_{t/\mu \leq k \leq (t+s)/\mu - 1} E_t^{\mu} |x_k^i|^2 + K s^2 \leq K \delta^2,
\]

where \( E_t^{\mu} \) denotes the conditioning on \( \mathcal{F}_t^{\mu} = \sigma\{x_0^\mu, \tilde{\xi}_k^i, \tilde{\xi}_k^i : i \leq r, k < \lfloor t/\mu \rfloor \} \). In the above, we have used \( E_t^{\mu} |x_k^i|^2 < \infty \) for \( \lfloor t/\mu \rfloor \leq k < \lfloor (t+s)/\mu \rfloor \), which can be shown as in Lemma 3.1. As a result,

\[
\lim_{\delta \to 0} \lim_{\mu \to 0} \sup_{0 \leq s \leq \delta} E \left[ \sup_{0 \leq s \leq \delta} E_t^{\mu} \left| x^{\mu,i}(t + s) - x^{\mu,i}(t) \right|^2 \right] = 0.
\]

The tightness of \( \{x^{\mu,i}(\cdot)\} \) follows from [11, p. 47].

(b) Characterization of the limit. For notational simplicity, we shall not use a function \( f(\cdot, \cdot) \in C_0^2 \) in the usual martingale problem formulation for the following derivation, but work with the underlying sequences directly. It is convenient to proceed with a scaling argument to treat the random renewal times.

Define the process \( Z_{n,i}^{\mu,i}, Z^{\mu,i}(\cdot), \Psi^{\mu,i}_n \), and \( \Psi^{\mu,i}(\cdot) \) by

\[
Z_{n,i}^{\mu,i} = \mu \sum_{j=0}^{n-1} Y_j^i, \quad Z^{\mu,i}(t) = Z_{n,i}^{\mu,i} \text{ on } [n\mu, n\mu + \mu),
\]

\[
\Psi^{\mu,i}_n = \mu \sum_{k=0}^{n-1} [M_{\tau_k^i}(\tilde{\alpha}_{\tau_k^i}) x_{\tau_k^i}^i] + \mu \sum_{k=0}^{n-1} W_{\tau_k^i}(\tilde{\alpha}_{\tau_k^i}) \tilde{\xi}_{\tau_k^i}^i + \mu \sum_{k=0}^{n-1} \hat{W}_{\tau_k^i}(x_{\tau_k^i}^i, \tilde{\alpha}_{\tau_k^i}, \tilde{\xi}_{\tau_k^i}^i),
\]

\[
\Psi^{\mu,i}(t) = \Psi^{\mu,i}_n \text{ for } t \in [n\mu, n\mu + \mu),
\]

Use a method similar to that in (a), we can prove that \( Z^{\mu,i}(\cdot) \) and \( \Psi^{\mu,i}(\cdot) \) are tight. As a result, we obtain that \( \{x^{\mu,i}(\cdot), Z^{\mu,i}(\cdot), \Psi^{\mu,i}(\cdot)\} \) is tight in \( D([0, \infty) : \mathbb{R}^3) \) and all limits are uniformly Lipschitz continuous. We fix and work with a weakly convergent subsequence, also indexed by \( \mu \), and with the limit denoted by \( (x^i(\cdot), Z^i(\cdot), \Psi^i(\cdot)) \), for \( i \leq r \). Now we state a lemma.

**Lemma 3.5.** Under the conditions of Theorem 3.4, the limits of \( Z^{\mu,i}(\cdot) \) and \( \Psi^{\mu,i}(\cdot) \) satisfy

\[
Z^i(t) = \int_0^t \pi'(x(Z^i(s)), \alpha(Z^i(s)))ds
\]

and
\[
\Psi^i(t) = \int_0^t \left[ \overline{M}(\alpha(Z^i(u)))x(Z^i(u)) \right]^i du,
\]

respectively.

Proof of Lemma 3.5. For fixed \(i\), using an argument similar to that in [13, pp. 224–226], we can derive (3.12). The details are omitted.

Next we work on \(\Psi^\mu(t)\) and concentrate on the proof of (3.13). First, it is readily seen that for any \(t, s > 0\),
\[
\Psi^\mu(t + s) - \Psi^\mu(t)
\]
\[
= \mu \sum_{\ell = 1}^{m_\mu} \sum_{k = t/\mu}^{(t + s)/\mu} \left\{ [M_{\tau^\ell_k}(\ell) x_{\tau^\ell_k}]^i + W_{\tau^\ell_k}(\ell) \xi^i_{\ell_k} + \widehat{W}^i_{\tau^\ell_k}(x^i_{\tau^\ell_k}, \ell, \xi^i_{\ell_k}) \right\} I(\tilde{\alpha}_{\tau^\ell_k} = \ell).
\]

Next, pick out any bounded and continuous function \(h(\cdot)\), any positive integer \(\kappa\), and any \(t_j \leq t\) for \(j \leq \kappa\); the weak convergence and Skorohod representation imply that
\[
\lim_{\mu \to 0} Eh(x^\mu(t_j), \alpha^\mu(t_j) : j \leq \kappa) [\Psi^\mu(t + s) - \Psi^\mu(t)]
\]
\[
= Eh(x(t_j), \alpha(t_j) : j \leq \kappa) [\Psi(t + s) - \Psi(t)].
\]

Choose \(m_\mu\) so that \(m_\mu \to \infty\) as \(\mu \to 0\), but \(\mu m_\mu = \delta_\mu \to 0\). By the continuity of the linear function in the variable \(x\),
\[
\lim_{\mu} Eh(x^\mu(t_j), \alpha^\mu(t_j) : j \leq \kappa) \left[ \mu \sum_{\ell = 1}^{m_\mu} \frac{(t + s)/\mu}{k = t/\mu} \sum_{\ell = 1}^{m_\mu} \left[ M_{\tau^\ell_k}(\ell) x_{\tau^\ell_k} \right]^i I(\tilde{\alpha}_{\tau^\ell_k} = \ell) \right]
\]
\[
= \lim_{\mu} Eh(x^\mu(t_j), \alpha^\mu(t_j) : j \leq \kappa) \left[ \sum_{\ell = 1}^{m_\mu} \frac{t + s}{\delta_\mu} \sum_{k = t/\mu}^{(t + s)/\mu} \left[ M_{\tau^\ell_k}(\ell) x_{\tau^\ell_k} \right]^i I(\tilde{\alpha}_{\tau^\ell_k} = \ell) \right]
\]
\[
= \lim_{\mu} Eh(x^\mu(t_j), \alpha^\mu(t_j) : j \leq \kappa) \left[ \sum_{\ell = 1}^{m_\mu} \frac{t}{\delta_\mu} \sum_{k = t/\mu}^{(t + s)/\mu} \left[ M_{\tau^\ell_k}(\ell) x_{\tau^\ell_k} \right]^i I(\tilde{\alpha}_{\tau^\ell_k} = \ell) \right]
\]
\[
= \lim_{\mu} Eh(x^\mu(t_j), \alpha^\mu(t_j) : j \leq \kappa) \left[ \sum_{\ell = 1}^{m_\mu} \frac{t + s}{\delta_\mu} \sum_{k = t/\mu}^{(t + s)/\mu} \left[ \overline{M}(\ell) x_{\tau^\ell_k} \right]^i I(\tilde{\alpha}_{\tau^\ell_k} = \ell) \right]
\]
\[
+ \lim_{\mu} Eh(x^\mu(t_j), \alpha^\mu(t_j) : j \leq \kappa) \left[ \sum_{\ell = 1}^{m_\mu} \frac{t}{\delta_\mu} \sum_{k = t/\mu}^{(t + s)/\mu} \left[ (M_{\tau^\ell_k}(\ell) - \overline{M}(\ell)) x_{\tau^\ell_k} \right]^i \right] I(\tilde{\alpha}_{\tau^\ell_k} = \ell).
\]

Denoting \(P(\tilde{\alpha}_{\tau^\ell_k} = \ell|\tau^i_{\ell_k} = \ell) = p(\tau^i_{\ell_k}, \tau^i_{\ell_k})\) with \(\ell\) suppressed, inserting conditional expectation and using a partial summation, we obtain that
\[
\frac{1}{m_\mu} \sum_{k = t/\mu}^{m_\mu} (M_{\tau^\ell_k}(\ell) - \overline{M}(\ell)) p(\tau^i_{\ell_k}, \tau^i_{\ell_k})
\]
\[
= \frac{1}{m_\mu} \sum_{k = t/\mu}^{m_\mu} (M_{\tau^\ell_k}(\ell) - \overline{M}(\ell)) p(\tau^i_{\ell_k}, \tau^i_{\ell_k})
\]
\[
+ \frac{1}{m_\mu} \sum_{k = t/\mu}^{m_\mu} (M_{\tau^\ell_k}(\ell) - \overline{M}(\ell)) [p(\tau^i_{\ell_k}, \tau^i_{\ell_k}) - p(\tau^i_{\ell_k}, \tau^i_{\ell_k})].
\]
By virtue of the assumption (A2)(c), the term in the second line of (3.17) goes to 0 as $\mu \to 0$. Next, noting

$$(I + \mu Q)^{r_j + 1 - r_{i_{m_{\mu}}}} - (I + \mu Q)^{r_j - r_{i_{m_{\mu}}}} = O(\mu),$$

we have

$$[p(r_j^i, r_{i_{m_{\mu}}}) - p(r_{j+1}^i, r_{i_{m_{\mu}}})] = O(\mu).$$

As a result,

$$E \left[ h(x^\mu(t_j), \alpha^\mu(t_j) : j \leq \kappa)E_{x_{1_{m_{\mu}}}^\mu} \left[ \sum_{\ell = 1}^{m_0} \sum_{\delta_{\mu}} \frac{\delta_{\mu}}{m_{\mu}} \sum_{k = l_{m_{\mu}}} [M_{r_k}(\ell) - \overline{M}(\ell)x_{r_{1_{m_{\mu}}}^\mu}]^i \bigg| I(\tilde{\alpha}_{r_k} = \ell) \right] \right]$$

$$\leq E \left[ \sum_{\ell = 1}^{m_0} \sum_{\delta_{\mu}} \frac{\delta_{\mu}}{m_{\mu}} \sum_{k = l_{m_{\mu}}} [M_{r_k}(\ell) - \overline{M}(\ell)] E|_{x_{r_{1_{m_{\mu}}}^\mu}} | O(\mu) \right] \to 0 \text{ as } \mu \to 0.$$

Since $\mu m_{\mu} \to 0$ as $\mu \to 0$, when $\mu |m_{\mu}| \to u$, for all $l_{m_{\mu}} \leq k \leq l_{m_{\mu}} + m_{\mu}, \mu k \to u$ as well. Therefore, the detailed estimates above lead to

$$\lim \mu E h(x^\mu(t_j), \alpha^\mu(t_j) : j \leq \kappa) \left[ \mu \sum_{\ell = 1}^{m_0} \sum_{\delta_{\mu}} \frac{\delta_{\mu}}{m_{\mu}} \sum_{k = l_{m_{\mu}}} [M_{r_k}(\ell)x_{r_{1_{m_{\mu}}}^\mu}]^i I(\tilde{\alpha}_{r_k} = \ell) \right]$$

$$= \lim \mu E h(x^\mu(t_j), \alpha^\mu(t_j) : j \leq \kappa) \sum_{\ell = 1}^{m_0} \sum_{\delta_{\mu}} \frac{\delta_{\mu}}{m_{\mu}} \overline{M}(\ell)^i \sum_{k = l_{m_{\mu}}} [M_{r_k}(\ell)x_{r_{1_{m_{\mu}}}^\mu}]^i I(\tilde{\alpha}_{r_k} = \ell)$$

$$= \lim \mu E h(x^\mu(t_j), \alpha^\mu(t_j) : j \leq \kappa) \sum_{\ell = 1}^{m_0} \sum_{\delta_{\mu}} \frac{\delta_{\mu}}{m_{\mu}} \sum_{k = l_{m_{\mu}}} \overline{M}(\tilde{\alpha}_{r_k})^i x_{r_{1_{m_{\mu}}}^\mu}^i I(\tilde{\alpha}_{r_k} = \ell)$$

$$= \lim \mu E h(x^\mu(t_j), \alpha^\mu(t_j) : j \leq \kappa) \left( \int_{t_1}^{t_2} \overline{M}(\alpha^\mu(Z^i(u)))x^\mu(Z^i(u)) \right)^i du.$$

Next, using the independence of $\tilde{\alpha}_{r_k}$ with $\tilde{\xi}_n^i$ and [5, Corollary 2.4], a similar conditional expectation together with the mixing conditions given in (A2) (a) yields

$$\mu \sum_{\ell = 1}^{m_0} \sum_{\delta_{\mu}} \frac{\delta_{\mu}}{m_{\mu}} \sum_{k = l_{m_{\mu}}} \overline{M}(\alpha^\mu(Z^i(u)))x^\mu(Z^i(u)) \right)^i du.$$

(3.19)
Likewise, using the continuity of \( \hat{W}(\cdot, \ell, \zeta_i^i) \), the limit of
\[
\sum_{\ell=1}^{m_0} \mu \sum_{k=\ell / \mu}^{(t+s)/\mu} \hat{W}_{\tau_k^i}(x_{\tau_k^i}, \ell, \zeta_i^i) I_{[\tau_{\tau_k^i}^i = \ell]}
\]
is the same as that of
\[
\sum_{\ell=1}^{m_0} \delta_{\mu} \frac{1}{m_\mu} \sum_{k=\ell m_\mu}^{(t+s)/\mu} \hat{W}_{\tau_k^i}(x_{\tau_k^i}, \ell, \zeta_i^i) I_{[\tau_{\tau_k^i}^i = \ell]}.
\]
Then using the uniform mixing (see [2, p. 166]) given in (A2) to the last line above, we obtain
\[
\mu \sum_{l=1}^{l_{\mu}} \delta_{l_{\mu}} \frac{1}{m_{l_{\mu}}} \sum_{k=lm_{l_{\mu}}}^{lm_{l_{\mu}}+m_{l_{\mu}}-1} E|E_{\tau_{lm_{l_{\mu}}}^i} \hat{W}_{\tau_k^i}(x_{\tau_k^i}, \ell, \zeta_i^i)| P(\tau_{\tau_k^i}^i = \ell | \tau_{\tau_k^i}/\mu) \rightarrow 0 \text{ as } \mu \rightarrow 0.
\]
Thus,
\[
(3.20)
\]
\[Eh(x^\mu(t_j), \alpha^\mu(t_j) : j \leq \kappa) \sum_{l=1}^{l_{\mu}} \delta_{l_{\mu}} \frac{1}{m_{l_{\mu}}} \sum_{k=lm_{l_{\mu}}}^{lm_{l_{\mu}}+m_{l_{\mu}}-1} \hat{W}_{\tau_k^i}(x_{\tau_k^i}, \ell, \zeta_i^i) I_{[\tau_{\tau_k^i}^i = \ell]} \rightarrow 0.
\]
Using the estimates obtained thus far together with (3.14), we have proved that
\[
(3.21)
\]
\[Eh(x(t_j), \alpha(t_j) : j \leq \kappa) [\Psi^i(t+s) - \Psi^i(t) - \int_t^{t+s} [\mathcal{M}(\alpha(Z^i(u)))x(Z^i(u))]^i du] = 0.
\]
Therefore, the proof of the lemma is concluded.

Completion of the proof of Theorem 3.4. With \( Z^{-1} \) denoting the inverse of \( Z \), we have
\[
(3.22)
\]
\[x^{\mu,i}(t) - x^i(0) = \mu \sum_{k=0}^{N_i(t)/\mu - 1} \left\{ [M_k(\ell)\bar{x}^i_k]^i + [W_k(\ell)\xi_i^i]^i + \hat{W}_k(\bar{x}^i_k, \ell, \zeta_i^i) \right\} I_{[\tau_{\tau_k^i}^i = \ell]} I_{[\tau_{\tau_k^i}^i = \ell]}
\]
\[= \mu \sum_{k=0}^{(2^{\nu-1})-1(t)} \left\{ [M_k(\ell)\bar{x}^i_k]^i + [W_k(\ell)\xi_i^i]^i + \hat{W}_k(\bar{x}^i_k, \ell, \zeta_i^i) \right\} I_{[\tau_{\tau_k^i}^i = \ell]} I_{[\tau_{\tau_k^i}^i = \ell]}
\]
Lemma 3.5 then yields the limit process
\[
x^i(t) = x^i(0) + \int_0^{(Z^i)^{-1}(t)} [\mathcal{M}(\alpha(Z^i(u)))x(Z^i(u))]^i du.
\]
This in turn implies
\[
\dot{x}^i(t) = \frac{[\mathcal{M}(\alpha(Z^i((Z^i)^{-1}(t))))(x(Z^i((Z^i)^{-1}(t))))]^i (Z^i)^{-1}(t)}{\pi^i(x(t), \alpha(t))}
\]
as desired. Thus the theorem is proved.
3.3. Invariance theorem. To study the long-time behavior, we derive an invariant theorem for the switched system. Following the discussion in [39, Chapter 9], recall that a Borel measurable set $U \subset \mathbb{R}^r \times \mathcal{M}$ is invariant with respect to the process $(x(t), \alpha(t))$ if $P((x(t), \alpha(t)) \in U \text{ for all } t \geq 0) = 1 \text{ for any initial } (x, i) \in U$. That is, a process starting from $U$ will remain in $U$ with probability 1. We also need the notion of stability of sets in probability. They are defined naturally as follows:

- A closed and bounded set $K_c \subset \mathbb{R}^r$ is said to be stable in probability if for any $\delta > 0$ and $\rho > 0$, there is a $\delta_1 > 0$ such that starting from $(x, i)$, $P(\sup_{t \geq 0} d(x(t), K_c) < \rho) \geq 1 - \delta$ whenever $d(x, K_c) < \delta_1$;
- a closed and bounded set $K_c \subset \mathbb{R}^r$ is said to be asymptotically stable in probability if it is stable in probability, and $P(\lim_{t \to \infty} d(x(t), K_c) = 1) \to 1$, as $d(x, K_c) \to 0$.

In the above, we have used the usual distance function $d(x, D) = \inf(|x - y| : y \in D)$. We proceed to obtain the following result.

**Theorem 3.6.** Assume that for each $i \in \mathcal{M}$, $\mathcal{M}(i)$ is irreducible. Under the conditions of Theorem 3.4, the following assertions hold:

(i) The set $Z = \text{span}\{1\}$ is an invariant set.

(ii) The set $Z$ is asymptotically stable in probability.

**Proof.** To prove (i), we divide the time intervals according to the associated switch times. We begin with $(x(0), \alpha(0)) = (x_0, i)$. Following the dynamic system given in (3.6), let $\rho_1$ be the first switching time, i.e., $\rho_1 = \inf\{t : \alpha(t) = i_1 \neq i\}$. Note that $x(t) = x(t, \omega)$, where $\omega \in \Omega$ is the sample point. Then in the interval $[0, \rho_1)$, for almost all $\omega$, (3.6) is a system with constant matrix $\mathcal{M}(i)$. For all $t \in [0, \rho_1)$, from (3.6) we have that for any $T < \infty$ and $t \in [0, T]$,

$$x(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \frac{d^k x(0)}{dt^k} \quad \text{and} \quad \sup_{0 \leq t \leq T} \left| \frac{d^k x(0)}{dt^k} \right| \leq K < \infty.$$  

Because the matrix $\mathcal{M}(i)$ is irreducible, there is an eigenvalue 0, and the rest of the eigenvalues all have negative real parts.

If $x(0) \in Z$, then $x(0) = c \mathbb{1}$, $[\mathcal{M}(i)x(0)]^t = 0$, and, for each $i \leq r$,

$$\frac{dx^i}{dt}(0) = \frac{[\mathcal{M}(i)x(0)]^i}{\pi^i(x(0), i)} = 0,$$

and similarly, we obtain

$$\frac{d^k x^i}{dt^k}(0) = 0 \quad \text{for all} \quad k > 0, \; i \leq r.$$  

Therefore, $x(t) = x(0)$ for all $t \in [0, \rho_1)$. Thus $x(t) \in Z$ for all $t \in [0, \rho_1)$. Now, define $\rho_2 = \inf\{t \geq \rho_1 : \alpha(t) = i_2 \neq i_1\}$. By the continuity of $x(\cdot)$, $x(\rho_1) = x(\rho_2) \in Z$. Similarly as in the previous paragraph, we can show that for all $t \in [\rho_1, \rho_2)$, $x(t) \in Z$. Continue in this way. For any $T > 0$, consider $[0, T]$. Then $0 < \rho_1 < \rho_2 < \ldots < \rho_{N(T)} \leq T$, where $N(t)$ is the counting that counts the number of switchings in the interval $[0, T]$, and $\rho_n$ is defined recursively such that $\alpha(\rho_n) = i_n$ and $\rho_{n+1} = \inf\{t \geq \rho_n : \alpha(t) = i_{n+1} \neq i_n\}$. Suppose that we have, for all $t \leq \rho_n$, $x(t) \in Z$ with probability 1 (w.p.1). Using induction, we can show $x(t) \in Z$ for all $t \in [0, \rho_{N(T)})$. Finally, we work with the interval $[\rho_{N(T)}, T]$; this establishes the first assertion.
To prove (ii), define \( V(x) = x'x/2 \). Since \( V(x) \) is independent of the switching component, \( \sum_{\ell=1}^{m_0} q_{\ell} V(x) = 0 \). Thus, for each \( \ell \in \mathcal{M} \), because of the irreducibility of \( \overline{\mathcal{M}}(\ell) \),

\[
L_1 V(x) = \sum_{i=1}^{r} \frac{x^i[\overline{\mathcal{M}}(\ell)x]^i}{\pi^i(x, \ell)} < 0 \quad \text{for all } x \notin \mathcal{Z}.
\]

The rest of the proof of the stability in probability of the set \( \mathcal{Z} \) is similar in spirit to that of [39, Chapter 9]. We omit the details for brevity. \( \square \)

Denote \( x_c = \eta \mathbb{1} \). With the above proposition, we can further obtain the following result as a corollary of Theorem 3.6.

**Corollary 3.7.** Assume the conditions of Theorem 3.6. Then for any \( t_\mu \to \infty \) as \( \mu \to 0 \), \( x^\mu(\cdot + t_\mu) \) converges to the consensus solution \( \eta \mathbb{1} \) in probability. That is, for any \( \delta > 0 \), \( \lim_{\mu \to 0} P(|x^\mu(\cdot + t_\mu) - x_c| \geq \delta) = 0 \).

### 3.4. Normalized error sequences

This section is devoted to analyzing the rates of variations of a scaled sequence of errors. We begin with a result on upper bounds on estimation errors in the mean square sense.

**Theorem 3.8.** Assume the conditions of Theorem 3.4. Then there is an \( N_\mu \) such that \( E|x_n|^2 = O(1) \) for all \( n \geq N_\mu \).

**Proof.** We prove the assertion by means of perturbed Liapunov function methods. Redefine \( V(x) = (x - x_c)(x - x_c)/2 \). Note that \( V^c(x) = (\partial / \partial x^i) V(x) = (x_i - x_c) \) and \( V_{xx}(x) = I \) is the identity matrix. Using a Taylor expansion for \( V(x) \), we have

\[
E_n V(x_{n+1}) - V(x_n) = \mu \sum_{i=1}^{r} (x_n^i - x_c^i) \left\{ \left[ \overline{\mathcal{M}}(\alpha_n)(x_n - x_c) \right]^i + \left[ (M_n(\alpha_n) - \overline{\mathcal{M}}(\alpha_n))(x_n - x_c) \right]^i 
\right. 
+ (M_n(\alpha_n)(\overline{x}_n^i - x_n))^i + |W_n(\alpha_n)\xi_n^i| + \overline{W}_n^i(\overline{x}_n^i, \alpha_n, \xi_n^i)
\left. + [(M_n(\alpha_n) - \overline{\mathcal{M}}(\alpha_n)x_c)]^i I_{n+1}^i + O(\mu^2)(V(x_n) + 1) \right. 
\]

Note that for all \( x_n \notin \mathcal{Z} \), \( \sum_{\ell=1}^{m_0} x_n^i [\overline{\mathcal{M}}(\ell)x_n]^i I_{n+1}^i = -\lambda V(x_n) \) for some \( \lambda > 0 \). We will use \( \overline{W}_n^i(\overline{x}_n, \alpha_n, \xi_n^i) \) and \( |W_n(\alpha_n)\xi_n^i| \) interchangeably in what follows. For some \( 0 < \lambda_0 < 1 \), define

\[
V_1^\mu(x, n) = \mu \sum_{i=1}^{r} \sum_{\ell=1}^{m_0} \sum_{j=n}^{\infty} \lambda_0^{j-n} E_n (x^i - x_c^i) [\overline{\mathcal{M}}(\ell) - \overline{\mathcal{M}}(\ell)](x - x_c)^i I_{\{\alpha_j = \ell\}} I_{j+1},
\]

\[
V_2^\mu(x, n) = \mu \sum_{i=1}^{r} \sum_{\ell=1}^{m_0} \sum_{j=n}^{\infty} \lambda_0^{j-n} E_n (x^i - x_c^i) [M_j(\ell) - \overline{\mathcal{M}}(\ell)](x - x_c)^i I_{\{\alpha_j = \ell\}} I_{j+1},
\]

\[
V_3^\mu(x, n) = \mu \sum_{i=1}^{r} \sum_{\ell=1}^{m_0} \sum_{j=n}^{\infty} E_n (x^i - x_c^i) [W_j(\ell)\xi_j^i] I_{\{\alpha_j = \ell\}} I_{j+1},
\]

\[
V_4^\mu(x, n) = \mu \sum_{i=1}^{r} \sum_{\ell=1}^{m_0} \sum_{j=n}^{\infty} E_n (x^i - x_c^i) \overline{W}_j^i(x^i, \ell, \xi_j^i) I_{\{\alpha_j = \ell\}} I_{j+1},
\]

\[
V_5^\mu(x, n) = \mu \sum_{i=1}^{r} \sum_{\ell=1}^{m_0} \sum_{j=n}^{\infty} \lambda_0^{j-n} E_n (x^i - x_c^i) [(M(\ell) - \overline{\mathcal{M}}(\ell))x_c] (x - x_c)^i I_{\{\alpha_j = \ell\}} I_{j+1}.
\]

It is easily checked that

\[
|V_1^\mu(x, n)| = O(\mu)(V(x) + 1), \quad i = 1, \ldots, 5.
\]
Noting
\[
\mu \sum_{i=1}^{r} \sum_{\ell=1}^{m_0} \sum_{j=n+1}^{\infty} \lambda_0^{-(n+1)} E_n(x_{n+1}^i - x_c^i)(M_j(\ell) - \bar{M}(\ell))(x_{n+1} - x_c)^i I_{(\alpha_j = \ell)} I_{j+1}^i \\
- \mu \sum_{i=1}^{r} \sum_{\ell=1}^{m_0} \sum_{j=n+1}^{\infty} \lambda_j^{-(n+1)} E_n(x_{n}^i - x_c^i)(M_j(\ell) - \bar{M}(\ell))(x_{n} - x_c)^i I_{(\alpha_j = \ell)} I_{j+1}^i \\
= O(\mu^2)(V(x_n) + 1),
\]
we have
\[
E_n V_1^\mu(x_{n+1}, n+1) - V_1^\mu(x_n, n) = E_n[V_1^\mu(x_{n+1}, n+1) - V_1^\mu(x_n, n+1)] + [E_n V_1^\mu(x_{n+1}, n+1) - V_1^\mu(x_n, n)] \\
= -\mu \sum_{i=1}^{r} E_n(x_{n}^i - x_c^i)(M_j(\alpha_n) - \bar{M}(\alpha_n))[x_{n}^i] I_{n+1}^i + O(\mu^2)(V(x_n) + 1).
\]
Likewise, we obtain
\[
E_n V_2^\mu(x_{n+1}, n+1) - V_2^\mu(x_n, n) = -\mu \sum_{i=1}^{r} E_n(x_{n}^i - x_c^i)[W^i_n(\alpha_n)] I_{n+1}^i + O(\mu^2)(V(x_n) + 1),
\]
\[
E_n V_3^\mu(x_{n+1}, n+1) - V_3^\mu(x_n, n) = -\mu \sum_{i=1}^{r} E_n(x_{n}^i - x_c^i)[W^i_n(\alpha_n, \xi_n^i)] I_{n+1}^i + O(\mu^2)(V(x_n) + 1),
\]
\[
E_n V_4^\mu(x_{n+1}, n+1) - V_4^\mu(x_n, n) = -\mu \sum_{i=1}^{r} E_n(x_{n}^i - x_c^i)[W^i_n(\alpha_n, \zeta_n^i)] I_{n+1}^i + O(\mu^2)(V(x_n) + 1),
\]
\[
E_n V_5^\mu(x_{n+1}, n+1) - V_5^\mu(x_n, n) = -\mu \sum_{i=1}^{r} E_n(x_{n}^i - x_c^i)(M_j(\alpha_n) - \bar{M}(\alpha_n))[x_{n}^i] I_{n+1}^i + O(\mu^2)(V(x_n) + 1).
\]
Define \( V^\mu(x, n) = V(x) + \sum_{i=1}^{5} V_i^\mu(x, n) \). Using (3.23), (3.26), and (3.27), we arrive at
\[
E_n V^\mu(x_{n+1}, n+1) - V^\mu(x_n, n) \leq -\lambda V(x_n) + O(\mu^2)(V(x_n) + 1).
\]
Using (3.25), replacing \( V(x_n) \) in the last line above by \( V^\mu(x_n, n) \), taking expectation, and iterating on the resulting inequality, we arrive at
\[
EV^\mu(x_{n+1}, n+1) \leq (1 - \lambda\mu)^n V^\mu(x_0, 0) + O(\mu^2) \sum_{k=0}^{n} (1 - \lambda\mu)^k \\
+ O(\mu^2) \sum_{k=0}^{n} (1 - \lambda\mu)^{n-k} V^\mu(x_k, k).
\]
Note that there is an \( N_\mu \) such that for all \( n \geq N_\mu \), we can make \((1 - \lambda\mu)^n EV^\mu(x_0, 0) \leq O(\mu)\). In addition, \( \sum_{k=0}^{n} (1 - \lambda\mu)^{n-k} = O(\mu^n) \) for all \( n \leq O(1/\mu) \). Using the estimates in the above paragraph, an application of the Gronwall’s inequality yields
that \( EV(x_{n+1}, n + 1) \leq O(\mu) \). Using (3.25) again in the estimate above, we obtain \( EV(x_n) \leq O(\mu) \). The desired result thus follows. \( \square \)

Define
\[
U_n = \frac{x_n - x_c}{\sqrt{\mu}} \quad \text{and} \quad \tilde{U}_n^i = \frac{x_n^i - x_c}{\sqrt{\mu}}.
\]

We assume that the following assumption holds:

\textbf{(A3)}

(i) For each \( \ell \in \mathcal{M} \) and each \( \zeta \), \( \hat{W}_n(\cdot, \ell, \zeta) \) has continuous partial derivatives with respect to \( x \) up to the second order and \( \hat{W}_{n,xx}(\cdot, \ell, \zeta) \) is uniformly bounded, where \( \hat{W}_{n,x}(\cdot) \) and \( \hat{W}_{n,xx}(\cdot) \) denote the first and second partial derivatives with respect to \( x \). The \( \{\hat{W}_n(x_c, \ell, \tilde{\zeta}^i)\} \) and \( \{\hat{W}_{n,x}(x_c, \ell, \tilde{\zeta}^i_{n,\ell})\} \) are bounded and uniform mixing sequences with the mixing measure satisfying \( \sum_k \psi_{i}(k) < \infty \). (ii) For a sequence of indicator functions \( \{\chi_j(A)\} \) where \( A \) is any measurable set with respect to \( \{\alpha_k, Y_k\} : i \leq r, k \leq j \}, \sum_{j=m}^{m+n-1} \{M_j(\ell) - \overline{M}(\ell)\}x_c/\sqrt{\mu} \rightarrow 0 \) in probability uniformly in \( m \). (iii) The averaging conditions in (A2) hold with fixed \( m \) replaced by \( N_t(N_\mu) \). (iv) The sets \( \{\tilde{\zeta}^i_n\} \) and \( \{\tilde{\zeta}^i_j\} \) for \( i = 1, \ldots, r \) are mutually independent.

Note that (A3) (i) implies that we can locally linearize \( \hat{W}_n(\cdot) \) around \( x_c \),

\[
\frac{1}{n} \sum_{j=m}^{m+n-1} E_n \hat{W}_{j,x}(x_c, \ell, \tilde{\zeta}^i) \rightarrow 0 \quad \text{in probability},
\]

\[
\sum_{k=n}^{\infty} E_n |\hat{W}_k(x_c, \ell, \tilde{\zeta}^i)| < \infty.
\]

Condition (A3) (ii) is a technical condition similar to [14, (A1.5), p. 318]. Recall that \( \varepsilon = \mu \). It is a requirement on the rates of local average for the sequence \( \{M_j(\ell)\} \). It is readily verified that

\[
U_n^i = U_n^i + \mu[M_n(\alpha_n)\hat{U}_n^i]\Gamma_{n+1}^i + \mu[\hat{W}_{n,x}(x_c, \alpha_n, \tilde{\zeta}^i_{n,\ell})\hat{U}_n^i]\Gamma_{n+1}^i
\]

\[
+ \sqrt{\mu}[W_n(\alpha_n)\tilde{\zeta}^i_{n,\ell}]\Gamma_{n+1}^i + \sqrt{\mu}[\hat{W}_n(x_c, \alpha_n, \tilde{\zeta}^i_{n,\ell})]\Gamma_{n+1}^i
\]

\[
+ \sqrt{\mu}[M_n(\alpha_n) - \overline{M}(\alpha_n)x_c]\Gamma_{n+1}^i + O(\mu^{3/2})O(\hat{U}_n^i)^2).
\]

To proceed, define \( U^\mu(t) = U_n \) for any \( t \in [\mu(n - N_\mu), \mu(n - N_\mu) + \mu] \). Under suitable conditions, we show that \( \{U^\mu(\cdot)\} \) converges weakly to a switching diffusion process. First note that by (A3) (ii),

\[
\frac{1}{\sqrt{\mu}} \sum_{j=t/\mu}^{(t+\delta)/\mu - 1} \left\{ [M_j(\alpha_j) - \overline{M}(\alpha_j)]x_c \right\} \Gamma_{j+1}^i
\]

\[
= \frac{1}{\sqrt{\mu}} \sum_{j=t/\mu}^{(t+\delta)/\mu - 1} \left\{ [M_j(\ell) - \overline{M}(\ell)]x_c \right\} \Gamma_{\{\alpha_j = \ell\}} \Gamma_{j+1}^i
\]

\( \rightarrow 0 \) as \( \mu \rightarrow 0 \) uniformly in \( t \in [0, T] \).
Observe that by virtue of Theorem 3.8, \( \{U_n : n \geq N_\mu \} \) is tight. It yields that

\[
U^{\mu,i}(t+s) - U^{\mu,i}(t) = \mu \sum_{j=t/\mu}^{(t+s)/\mu-1} [M_j(\alpha_j) \tilde{U}_{j+1}^i] I_{j+1}^i + \mu \sum_{j=t/\mu}^{(t+s)/\mu-1} [\tilde{W}_j(x, \alpha_j, \zeta)] I_{j+1}^i + \sqrt{\mu} \sum_{j=t/\mu}^{(t+s)/\mu-1} [\tilde{W}_j(x, \alpha_j, \zeta)] I_{j+1}^i + \sqrt{\mu} \sum_{j=t/\mu}^{(t+s)/\mu-1} [W_j(\alpha_j, \zeta)] I_{j+1}^i + o(1),
\]

where \( o(1) \to 0 \) in probability. The \( o(1) \) is obtained by use of the last line of (3.30), (A3), and Theorem 3.8. Using the methods presented for analyzing \( x^\mu(\cdot) \), we obtain the following lemma, whose details are omitted.

**Lemma 3.9.** \( \{U^{\mu,i}(\cdot), \alpha^\mu(\cdot)\} \) is tight in \( D([0,T] : \mathbb{R}^* \times \mathcal{M}) \).

Next, for \( n > 0 \), and each \( t \in \mathcal{M} \), define

\[
B^{\mu,i}_{n,1} = \sqrt{\mu} \sum_{j=N_\mu}^{N_\mu+n-1} [\tilde{W}_j(x, t, \zeta)] I_{j+1}^i, \quad B^{\mu,i}_{n,2} = \sqrt{\mu} \sum_{j=N_\mu}^{N_\mu+n-1} [W_j(t, \zeta)] I_{j+1}^i,
\]

\[
B^{\mu,i}_{1}(t) = B^{\mu,i}_{n,1}, \quad B^{\mu,i}_{2}(t) = B^{\mu,i}_{n,2}, \quad \text{for } t \in [\mu n, \mu n + \mu),
\]

\[
\tilde{B}^{\mu,i}_{n,1} = \sqrt{\mu} \sum_{j=N_\mu}^{N_\mu+n-1} [\tilde{W}_j(x, t, \zeta)] I_{j+1}^i, \quad \tilde{B}^{\mu,i}_{n,2} = \sqrt{\mu} \sum_{j=N_\mu}^{N_\mu+n-1} [W_j(t, \zeta)] I_{j+1}^i,
\]

\[
\tilde{B}^{\mu,i}_{1}(t) = \tilde{B}^{\mu,i}_{n,1}, \quad \tilde{B}^{\mu,i}_{2}(t) = \tilde{B}^{\mu,i}_{n,2}, \quad \text{for } t \in [\mu n, \mu n + \mu).
\]

Define also

\[
Z^{\mu,i}_n = \mu \sum_{j=N_\mu}^{N_\mu+n-1} Y_{j+1}^i, \quad Z^{\mu,i}(t) = Z^{\mu,i}_n, \quad \text{for } t \in [\mu n, \mu n + \mu),
\]

\[
b^{\mu,i}_{n,1} = \sqrt{\mu} \sum_{j=N_\mu}^{N_\mu+n-1} [\tilde{W}_j(x, \alpha_j, \zeta)] I_{j+1}^i, \quad b^{\mu,i}_{n,2} = \sqrt{\mu} \sum_{j=N_\mu}^{N_\mu+n-1} [W_j(\alpha_j, \zeta)] I_{j+1}^i,
\]

\[
b^{\mu,i}_{1}(t) = b^{\mu,i}_{n,1}, \quad b^{\mu,i}_{2}(t) = b^{\mu,i}_{n,2}, \quad \text{for } t \in [\mu n, \mu n + \mu)
\]

\[
\tilde{b}^{\mu,i}_{n,1} = \sqrt{\mu} \sum_{j=N_\mu}^{N_\mu+n-1} [\tilde{W}_j(x, \widetilde{\alpha}_j, \zeta)] I_{j+1}^i, \quad \tilde{b}^{\mu,i}_{n,2} = \sqrt{\mu} \sum_{j=N_\mu}^{N_\mu+n-1} [W_j(\widetilde{\alpha}_j, \zeta)] I_{j+1}^i,
\]

\[
\tilde{b}^{\mu,i}_{1}(t) = \tilde{b}^{\mu,i}_{n,1}, \quad \tilde{b}^{\mu,i}_{2}(t) = \tilde{b}^{\mu,i}_{n,2}, \quad \text{for } t \in [\mu n, \mu n + \mu).
\]

Using similar methods of the martingale averaging as in Theorem 3.4, we can show that \( B^{\mu,i}_{l}(\cdot) \) converges weakly to \( B^{\mu,i}_{l}(\cdot) = B^{\mu,i}_{l}((Z^l(\cdot))^{-1}) \) for \( l = 1, 2 \). It is also easy to see that \( B^{\mu,i}_{1}(\cdot) \) and \( B^{\mu,i}_{2}(\cdot) \) are independent.

**Theorem 3.10.** Under (A1)–(A3), there are independent standard Brownian
motions \( w_{i,1}(\cdot) \) and \( w_{i,2}(\cdot) \) for \( i \leq r \) such that the limits \( U^i(\cdot), i \leq r \), satisfy

\[
(3.34) \quad dU^i = \frac{M(\alpha(t)) U^i}{\pi^i(x_c, \alpha(t))} dt + \frac{\sigma_1^i(\alpha(t)) dw_{i,1}(t) + \sigma_2^i(\alpha(t)) dw_{i,2}(t)}{\sqrt{\pi^i(x_c, \alpha(t))}}, \quad i \leq r.
\]

**Proof.** The proof is similar in spirit to that of Theorem 3.4. So we will only point out the distinct features. Using the well-known results for mixing processes (see [2] and [5]), it is easily seen that \( \tilde{B}_{1}^{\mu,i,t}(\cdot) \) and \( \tilde{B}_{2}^{\mu,i,t}(\cdot) \) converge weakly to Brownian motions \( B_{1}^{\mu,i,t}(\cdot) \) and \( B_{2}^{\mu,i,t}(\cdot) \), with covariance \((\sigma_1^i(\cdot))^2 t\) and \((\sigma_2^i(\cdot))^2 t\), respectively. It is also easy to see that \( B_{1}^{\mu}(\cdot) \) and \( B_{2}^{\mu}(\cdot) \) are independent. Using the scaling argument as in the proof of Theorem 3.4, we can show that \( B_{1}^{\mu,i,t}(\cdot) \) converges weakly to \( B_{1}^{i,t}(\cdot) = \tilde{B}_{1}^{i,t}((Z^t(\cdot))^{-1}) \) for \( l = 1, 2 \). We need to prove the independence of the limit Brownian motions. Here we use an argument similar to [13, the last few lines of p. 239 and the first few lines of p. 240]. Let \( K_\mu \to \infty \) such that \( \sqrt{\mu}K_\mu \to 0 \) and let \( N_\mu^i = N_i(N_\mu) \). We work with

\[
\tilde{B}_{1}^{\mu,i,t}(t) = \sqrt{\mu} \sum_{j=\mathcal{N}_\mu^i+K_\mu}^{\mathcal{N}_\mu^i+K_\mu-1} [\tilde{W}_j(x_c, \tau, \zeta_j)^i], \quad \tilde{B}_{2}^{\mu,i,t}(t) = \sqrt{\mu} \sum_{j=\mathcal{N}_\mu^i+K_\mu}^{\mathcal{N}_\mu^i+K_\mu-1} [W_j(t)\tau]^i.
\]

For simplicity of notation, we take \( r = 2 \). We shall show that the limits of \( \tilde{B}_{1}^{\mu,i,t}(\cdot) \) are independent. Denote \( I_{mn}^\mu = I_{\{N_\mu^i = n\}}I_{\{N_\mu^j = m\}} \). For any bounded and continuous function \( H_1(\cdot) \) and \( H_2(\cdot) \), we have

\[
EH_1(\tilde{B}_{1}^{\mu,i,t}(t))H_2(\tilde{B}_{2}^{\mu,i,t}(t))
= \sum_{n,m} EH_1(\sqrt{\mu} \sum_{j=\mathcal{N}_\mu^i+K_\mu}^{\mathcal{N}_\mu^i+K_\mu-1} [\tilde{W}_j(x_c, \tau, \zeta_j)^i])H_2(\sqrt{\mu} \sum_{j=\mathcal{N}_\mu^i+K_\mu}^{\mathcal{N}_\mu^i+K_\mu-1} [W_j(t)\tau]^i)I_{mn}^\mu
= \sum_{n,m} EH_1(\sqrt{\mu} \sum_{j=\mathcal{N}_\mu^i+K_\mu}^{\mathcal{N}_\mu^i+K_\mu-1} [\tilde{W}_j(x_c, \tau, \zeta_j)^i])H_2(\sqrt{\mu} \sum_{j=\mathcal{N}_\mu^i+K_\mu}^{\mathcal{N}_\mu^i+K_\mu-1} [W_j(t)\tau]^i)E I_{mn}^\mu + o(1),
\]

where \( o(1) \to 0 \) as \( \mu \to 0 \). This, together with the arbitrariness of \( k \) and the weak convergence, implies the independence of the limit Brownian motions. Likewise, we can show the independence of the limit Brownian motions associated with \( \tilde{B}_{2}^{\mu,i,t}(\cdot) \).

We can then show

\[
\tilde{b}_{n,1}^{\mu,i} = \sqrt{\mu} \sum_{j=\mathcal{N}_\mu^i}^{\mathcal{N}_\mu^i+1-1} [\tilde{W}_j(x_c, \alpha_s, \zeta_j)^i]
= \sqrt{\mu} \sum_{i \in \mathcal{M}} \sum_{j=\mathcal{N}_\mu^i}^{\mathcal{N}_\mu^i+1-1} [\tilde{W}_j(x_c, \alpha_s, \zeta_j)^i] I_{\{\alpha_s = \nu\}}.
\]

Choose \( m_\mu, \delta_\mu \) etc. as in the convergence proof of the algorithm. Then

\[
(3.35) \quad \tilde{b}_{1}^{\mu,i}(t + s) - \tilde{b}_{1}^{\mu,i}(t) = \sum_{i \in \mathcal{M}} \frac{1}{\sqrt{m_\mu}} \sum_{l=t/\delta_\mu}^{(t+s)/\delta_\mu} \sum_{j=l m_\mu}^{l m_\mu + m_\mu - 1} [\tilde{W}_j(x_c, \alpha_{s}, \zeta_j)^i] I_{\{\alpha_{s} = \nu\}}.
\]
Let $\mu t_m \to u$. Then for all $lm_\mu \leq j \leq lm_\mu + m_\mu - 1$, $\mu k \to u$. Using the weak convergence of $\tilde{a}^t(\cdot)$ to $\alpha(\cdot)$ and the Skorohod representation, the limit in the last line of (3.35) is the same as that of

$$\sum_{i \in M} \sqrt{\tilde{b}_i} \frac{1}{\sqrt{\mu}} \sum_{t=t/\delta_\mu} \sum_{j=lm_\mu} \tilde{W}_j \cdot \{ \tilde{W}_j(x, t, \tilde{\zeta}_j) \}^i I_{\{ \alpha(u) = i \}}.$$ 

Thus, the limit of (3.35) is given by

$$\bar{b}_1(t + s) - \bar{b}_1(t) = \sum_{i \in M} \int_t^{t+s} \sigma_i^1(\cdot) I_{\{ \alpha(u) = i \}} \, dw_{i,1}(u)$$

$$= \int_t^{t+s} \sigma_i^1(\alpha(u)) \, dw_{i,1}(u),$$

where $w_{i,1}(\cdot)$ is a standard Brownian motion. Likewise, $\tilde{b}_2^i(\cdot)$ converges to a switched Brownian motion in the sense that $\bar{b}_2(t) = \int_0^s \sigma_i(\alpha(u)) \, dw_{i,2}(u)$. Thus $\bar{b}_2^i(\cdot) = \tilde{b}_2^i(\cdot) + \bar{b}_2^i(\cdot)$ converges weakly to $\bar{b}_2^i(\cdot)$ such that

$$\bar{b}_2(t) = \int_0^t \left[ \sigma_1^i(\alpha(u)) \, dw_{i,2}(u) + \sigma_2^i(\alpha(u)) \, dw_{i,2}(u) \right].$$

The last step is to combine the above estimates, together with the independence of the limit Brownian motions established, with a scaling argument as in the proof of Theorem 3.4. A few details are omitted. \[\Box\]

Remark 3.11. In view of the independence of the Brownian motions $w_{i,1}(\cdot)$ and $w_{i,2}(\cdot)$, there is a standard Brownian motion $w_i(\cdot)$ such that the switching diffusion (3.34) may also be written in an equivalent form as

$$dU_i = \frac{[M(\alpha(t))U_i]}{\pi_i(x, \alpha(t))} dt + \tilde{\sigma}_i(\alpha(t)) dw_i(t), \quad i \leq r,$$

where

$$[\tilde{\sigma}_i(t)]^2 = \frac{[\sigma_1^i(t)]^2 + [\sigma_2^i(t)]^2}{\pi_i(x, t)} \quad \ell \in M, \quad i \leq r.$$

4. Slowly varying ($\varepsilon \ll \mu$) and rapidly varying ($\mu \ll \varepsilon$) Markov chains.

This section is divided into two subsections. One is concerned with slowly varying Markov chains ($0 < \varepsilon \ll \mu$), whereas the other treats rapidly switching processes ($0 < \mu \ll \varepsilon$).

4.1. Slowly varying Markov chains. Suppose that $\varepsilon \ll \mu$, where $\varepsilon$ is the parameter appeared in the transition probability matrix of the Markov chain and $\mu$ is the step-size of the algorithm (2.4). Intuitively, because the Markov chain changes so slowly, the time-varying parameter process is essentially a constant. We reveal the asymptotic properties of the recursive algorithm. To facilitate the discussion and to simplify the notation, we take $\varepsilon = \mu^2$ in what follows.

Note that Lemma 3.2 still holds. We next analyze algorithm (2.4). As in the previous case, we can prove $\sup_{0 \leq t \leq O(1/\mu)} E|x_n^i| < \infty$. Define the piecewise constant interpolation $x^n(t) = x_n$ for $t \in [\mu n, \mu n + \mu)$. Then as in the previous section, we
have that \( \{x^{\mu,t}(\cdot)\} \) is tight in \( D([0,T],\mathbb{R}) \). We proceed to characterize its limit. The analysis is similar to that of Theorem 3.4, so we will omit most of the details.

The idea is that since the Markov chain is slowly varying. The parameter is almost a constant. Since \( \tilde{a}_0 = \sum_{i=1}^{m_0} p_i \tilde{a}_{i0} = 1 \), we obtain the desired result with \( \tilde{M}(t) = \tilde{a}_0 x(u, t) \) in (3.6) replaced by \( \sum_{i=1}^{m_0} p_i \tilde{M}(t) x(u, t) \). We summarize the discussions above into the following result.

**Theorem 4.1.** Assume the conditions of Theorem 3.4 with the modification that the step-size in (2.4) satisfies \( \varepsilon = \mu^2 \). Then \( x^{\mu}(\cdot) \) converges weakly to \( x(\cdot) \), which is a solution of the ODE

\[
\frac{dx^i(t)}{dt} = \sum_{i=1}^{m_0} p_i \frac{[M(t)x(t)]^i}{\pi^i(x(t), ij)},
\]

where \( \pi^i(x, t) \) is a standard Brownian motion and

\[
[\tilde{\sigma}_i(t)^2] = \sum_{i=1}^{m_0} p_i \frac{[\sigma^i_{1}(t)]^2 + [\sigma^i_{2}(t)]^2}{\pi^i(x, t)}, \quad i \leq r.
\]

Note that the interpolation of the centered and scaled sequence of errors has a diffusion limit in which the drift and diffusion coefficients are averaged out with respect to the initial probability distribution.

### 4.2. Fast changing Markov chains

This section takes up the issue that the Markov chain is fast varying compared to the adaptation. By that, we mean \( \mu \ll \varepsilon \). For concreteness of the discussion, we take a specific form of the step-size, namely, \( \varepsilon = \mu^{1/2} \). Intuitively, the Markov chain varies relatively fast and can be thought of as a noise process. Eventually it is averaged out.

For \( \alpha_{\ell m_\mu} = \ell \),

\[
P(\alpha_k = j | \alpha_{\ell m_\mu}) = \Xi_{ij}(\varepsilon lm_\mu, \varepsilon k) + O(\varepsilon + \exp(-\kappa)).
\]

In view of (3.5) and noting \( \varepsilon = \mu^{1/2} \) and irreducibility of \( Q \), we have \( \Xi_{ij}(\varepsilon lm_\mu, \varepsilon k) = \nu_j + O(\exp(-\kappa_0 \log(m_\mu)) \), where \( \nu_j \) is the \( j \)th component of the stationary distribution \( \nu = (\nu_1, \ldots, \nu_{m_0}) \) associated with the generator \( Q \) of the corresponding continuous-time Markov chain. This indicates that \( \Xi(s, t) \) can be approximated by a matrix \( \mathbf{I} \nu \) with identical rows. Thus we obtain the limit ODE.

**Theorem 4.3.** Assume the conditions of Theorem 3.4 with the modification that the step-size in (2.4) satisfies \( \varepsilon = \mu^{1/2} \). Then \( x^{\mu}(\cdot) \) converges weakly to \( x(\cdot) \), which is a solution of the ODE

\[
\frac{dx^i(t)}{dt} = \sum_{i=1}^{m_0} \nu_i \frac{[M(t)x(t)]^i}{\pi^i(x(t), ij)},
\]
In addition, for any \( t_\mu \to \infty \) as \( \mu \to 0 \) and for any \( \delta > 0 \), \( \lim_{\mu \to 0} \mathbb{P}(|x^\mu(\cdot + t_\mu) - x_c| \geq \delta) = 0 \).

Remark 4.4. Concerning the errors, for the fast changing Markov chain case, within a very short period of time, the system is replaced by an average with respect to the stationary distribution of the Markov chain. For the error analysis, furthermore, we may define \( \xi \_c \) \( U_n \) as in (3.29) and show that \( \{U_n : n \geq N_\mu \} \) is tight. If we let \( U^\mu(t) \) be the piecewise constant interpolation of \( U_n \) on \( t \in [(n - N_\mu)\mu, (n - N_\mu)\mu + \mu) \), similar to Remark 3.11, then \( U^\mu(\cdot) \) converges weakly to \( U(\cdot) \) such that \( U(\cdot) \) is the solution of the SDE

\[
dU^i = \sum_{i=1}^{m_0} \mu_i \left( \mathbb{M}(t)U(t) \right)^i dt + \mathbb{N}(t)dw_i(t),
\]

where \( w_i(\cdot) \) is a standard Brownian motion and

\[
|\mathbb{N}(t)|^2 = \sum_{i=1}^{m_0} \mu_i \left( \sigma_1^2 + \sigma_2^2 \right) \frac{\pi^i(x_c, t)}{\pi^i(x_c, t)}, \quad i \leq r.
\]

Remark 4.5. As was mentioned, for convenience of presentation, we chose \( \delta = \mu^2 \) and \( \varepsilon = \sqrt{\mu} \) for the slowly varying and fast varying cases. The specific forms of \( \mu \) and \( \varepsilon \) enable us to simplify the presentation. The convergence results remain essentially the same for the general cases \( \mu/\varepsilon \to 0 \) and \( \mu/\varepsilon \to \infty \).

5. Illustrative examples. This section presents several simulation examples. We call \((x_n - x_c)'(x_n - x_c)\) the consensus error variance at time \( n \).

Example 5.1. Suppose that the Markov chain \( \alpha_n \) has only two states, i.e., \( M = \{1, 2\} \). The transition probability matrix is

\[
P^\varepsilon = I + \varepsilon \begin{pmatrix} -0.4 & 0.4 \\ 0.3 & -0.3 \end{pmatrix}.
\]

For a given system of five subsystems, suppose the link gains are \( G_1 = \text{diag}(1, 0.3, 1.2, 4, 7, 10) \) and \( G_2 = \text{diag}(2, 0.5, 1, 6, 9, 14) \) with regime-switching at two different states. Suppose the initial states are \( x_0^1 = 12, x_0^2 = 34, x_0^3 = 56, x_0^4 = 8, x_0^5 = 76 \). The state average is \( \eta = 37.2 \) \( (x_c = \eta I) \). The initial consensus error is \( (x_0 - x_c)'(x_0 - x_c) = 3356.8 \). Take \( \varepsilon = 0.02 \) and step-size \( \mu = \varepsilon = 0.02 \). The updating algorithm runs for 3000 steps, and the stopped consensus error variance is \( (x_{3000} - x_c)'(x_{3000} - x_c) = 8.0166 \). In Figure 2, we plot the Markov chain state trajectories and the system state trajectories.

Example 5.2. Here we consider the case that the Markov chain changes very slowly compared with the adaptation step-size. That is, \( \varepsilon \ll \mu \). To be specific, suppose \( \varepsilon = \mu^2 \), where \( \mu = 0.02 \). The numerical results are shown in Figure 3. From the trajectory of the Markov chain, there is only one switching taking place in the first 1000 iterations. The convergence of the consensus is also demonstrated.

Example 5.3. Here we consider the fast changing Markov \( \mu \ll \varepsilon \). Specifically, we take \( \mu = \varepsilon^2 \) with \( \mu = 0.02 \). The corresponding trajectories are plotted in Figure 4. The frequent Markov switching is clearly seen.

6. Further remarks. For convenience and notational simplicity, we have used the current setup. Several extensions and generalizations can be carried out. So far,
the noise sequences are correlated random processes. For convenience, we used mixing-type noise processes. All of the development up to this point can be generalized to more complex $x$-dependent noise processes [14, sections 6.6 and 8.4].
To conclude, this paper provided a class of asynchronous SA algorithms for consensus-type problems with randomly switching topologies. This study extended the arenas for consensus-type control problems to randomly time-varying dynamics of networked systems.

Appendix. Consensus algorithm basics: Traditional setting. This appendix gives a brief account of the setup of consensus under simple conditions. Consider a networked system of \( r \) nodes, given by

\[
(A.1)\quad x_{n+1}^i = x_n^i + u_n^i, \quad i = 1, \ldots, r,
\]

where \( u_n^i \) is the node control for the \( i \)th node, or in a vector form, \( x_{n+1} = x_n + u_n \) with \( x_n = [x_n^1, \ldots, x_n^r]' \), \( u_n = [u_n^1, \ldots, u_n^r]' \). The nodes are linked by a sensing network, represented by a directed graph \( G \) whose element \((i, j)\) indicates estimation of the state \( x_n^i \) by node \( i \) via a communication link, and a permitted control \( v_n^{ij} \) on the link. For node \( i \), \((i, j) \in G\) is a departing edge and \((l, i) \in G\) is an entering edge. The total number of communication links in \( G \) is \( l_r \). From its physical meaning, node \( i \) can always observe its own state, which will not be considered as a link in \( G \).

We consider link controls among nodes permitted by \( G \). The node control \( u_n^i \) is determined by the link control \( v_n^{ij} \). Since a positive transportation of quantity \( v_n^{ij} \) on \((i, j)\) means a loss of \( v_n^{ij} \) at node \( i \) and a gain of \( v_n^{ji} \) at node \( j \), the node control at node \( i \) is \( u_n^i = -\sum_{(i, j) \in G} v_n^{ij} + \sum_{(j, i) \in G} v_n^{ji} \). The most relevant implication in this control scheme is that for all \( n \), \( \sum_{i=1}^r x_n^i = \sum_{i=1}^r x_0^i := \eta r \) for some \( \eta \in \mathbb{R} \) that is the average of \( x_0 \). That is, \( \eta = \sum_{i=1}^r x_0^i / r \). Consensus control seeks control algorithms that achieve \( x_n \to \eta \mathbb{I} \), where \( \mathbb{I} \) is the column vector of all 1s. A link \((i, j) \in G\) entails an estimate, denoted by \( \hat{x}_{n}^{ij} \), of \( x_n^i \) by node \( i \) with estimation error \( d_{n}^{ij} \), i.e.,

\[
(A.2)\quad \hat{x}_{n}^{ij} = x_n^i + d_{n}^{ij}.
\]

The estimation error \( d_{n}^{ij} \) is usually a function of the signal \( x_n^i \) itself and depends on communication channel noises \( \xi_n^{ij} \) in a nonadditive and nonlinear relation

\[
(A.3)\quad d_{n}^{ij} = g(x_n^i, \xi_n^{ij})
\]

and can be spatially and temporally dependent. Most existing literature considers much simplified noise classes \( d_{n}^{ij} = \xi_n^{ij} \) with i.i.d. assumptions.

Such extensions are necessary when dealing with networked systems. A sampled and quantized signal \( x \) in a networked system enters a communication transmitter as a source. To enhance channel efficiency and reduce noise effects, source symbols are encoded [6, 18]. Typical block or convolutional coding schemes such as Hamming, Reed–Solomon, or, more recently, the low-density parity-check (LDPC) code and Turbo code, often introduce a nonlinear mapping \( v = f_1(x) \). The code word \( v \) is then modulated into a waveform \( s = f_2(v) = f_2(f_1(x)) \) which is then transmitted. Even when the channel noise is additive, namely, the received waveform is \( w = s + d \) where \( d \) is the channel noise, after the reverse process of demodulation and decoding, we have \( y = g(w) = g(s + d) = g(f_2(f_1(x)) + d) \). As a result, the error term \( g(f_2(f_1(x)) + d) - x \) in general is nonadditive and signal dependent. In addition, block and convolution coding schemes introduce temporally dependent noises. In our formulation, this aspect is reflected in dependent \( \phi \)-mixing noises on \( \xi_n^{ij} \). These will be detailed later.

For simplification on system derivations, we use first \( d_{n}^{ij} = \xi_n^{ij} \) in this section. Let \( \eta_n \) and \( \xi_n \) be the \( l_r \)-dimensional vectors that contain all \( \hat{x}_{n}^{ij} \) and \( \xi_n^{ij} \) in a selected order,
respectively. Then, (A.2) can be written as \( \tilde{\eta}_n = H_1 x_n + \xi_n \), where \( H_1 \) is an \( l_x \times r \) matrix whose rows are elementary vectors such that if the \( \ell \)th element of \( \tilde{\zeta}_n \) is \( \tilde{\omega}^{ij} \), then the \( \ell \)th row in \( H_1 \) is the row vector of all zeros except for a “1” at the \( j \)th position. Each sensing link provides information \( \delta_n^{ij} = x_n^{ij} - \tilde{x}_n^{ij} \), an estimated difference between \( x_n^{ij} \) and \( \tilde{x}_n^{ij} \). This information may be represented, in the same arrangement as \( \tilde{\eta}_n \), by a vector \( \delta_n \) of size \( l_x \) containing all \( \delta_n^{ij} \) in the same order as \( \tilde{\eta}_n \). \( \delta_n \) can be written as \( \delta_n = H_2 x_n - \tilde{\eta}_n = H_2 x_n - H_1 x_n - \xi_n = H x_n - \xi_n \), where \( H_2 \) is an \( l_x \times r \) matrix whose rows are elementary vectors such that if the \( \ell \)th element of \( \tilde{\zeta}_n(k) \) is \( \tilde{\omega}^{ij} \), then the \( \ell \)th row in \( H_2 \) is the row vector of all zeros except for a “1” at the \( i \)th position, and \( H = H_2 - H_1 \). The reader is referred to [3] for basic matrix properties in graphs and to [27] for matrix iterative schemes. Due to network constraints, the information \( \delta_n^{ij} \) can only be used by nodes \( i \) and \( j \). When the control is linear, time invariant, and memoryless, we have \( v_n^{ij} = \mu g_{ij} \delta_n^{ij} \), where \( g_{ij} \) is the link control gain on \((i, j)\) and \( \mu \) is a global scaling factor that will be used in state updating algorithms as the recursive step-size. Let \( G \) be the \( l_x \times l_x \) diagonal matrix that has \( g_{ij} \) as its diagonal element. In this case, the node control becomes \( u_n = -\mu H' G \delta_n \). For convergence analysis, we note that \( \mu \) is a global control variable, and we may represent \( u_n \) equivalently as \( u_n = -\mu (H' G x_n - H' G \xi_n) = \mu (M x_n + W \xi_n) \), with \( M = -H' G H \) and \( W = H' G \).

Under the link-based state control \( v_n^{ij} \), the state updating scheme (A.1) becomes

\[
\begin{align*}
x_{n+1} & = x_n - \mu H' G \delta_n.
\end{align*}
\]

Since \( \mathbf{1}' M = 0, \mathbf{1}' W = 0, \mathbf{1}' x_{n+1} = \mathbf{1}' x_n = \mathbf{r} \eta \) hold for all \( n \), which is a natural constraint to the SA algorithm. Starting at \( x_0 \), \( x_n \) is updated iteratively by using (A.4), which for the analysis is

\[
\begin{align*}
x_{n+1} & = x_n + \mu (M x_n + W \xi_n).
\end{align*}
\]

Throughout the paper, the noise \( \{ \xi_n \} \) is allowed to be correlated, both spatially and temporally. We will assume the following conditions.

(A0) (i) All link gains are positive, \( g_{ij} > 0 \). (ii) \( \mathcal{G} \) contains a spanning tree.

Recall that a square matrix \( \tilde{Q} = (\tilde{q}_{ij}) \) is a generator of a continuous-time Markov chain if \( \tilde{q}_{ij} \geq 0 \) for all \( i \neq j \) and \( \sum_j \tilde{q}_{ij} = 0 \) for each \( i \). Also, a generator or the associated continuous-time Markov chain is irreducible if the system of equations

\[
\begin{align*}
\nu \tilde{Q} & = \nu, \\
\nu \mathbf{1} & = 1
\end{align*}
\]

has a unique solution, where \( \nu = [\nu_1, \ldots, \nu_r] \in \mathbb{R}^{1 \times r} \) with \( \nu_i > 0 \) for each \( i = 1, \ldots, r \) is the associated stationary distribution. Assume that the noise is unbounded but has bounded \((2 + \Delta)\)th moments. In addition, it is a sequence of correlated noise, much beyond the usual i.i.d. noise classes. A \( \phi \)-mixing sequence has the property that the remote past and the distant future are asymptotically independent. The asymptotic independence is reflected by the condition on the underlying mixing measure. The proof of the following theorem is in [33].

**Theorem A.1.** Under assumption (A0), (1) \( M \) has rank \( r - 1 \) and is negative semidefinite, and (2) \( M \) is a generator of a continuous-time Markov chain and is irreducible.
REFERENCES


