Asymptotic properties of nonlinear autoregressive Markov processes with state-dependent switching

Fubao Xi\textsuperscript{a}, G. Yin\textsuperscript{b,}\textsuperscript{*}

\textsuperscript{a} Department of Mathematics, Beijing Institute of Technology, Beijing 100081, People's Republic of China
\textsuperscript{b} Department of Mathematics, Wayne State University, Detroit, MI 48202, USA

\begin{abstract}
In this paper, we consider a class of nonlinear autoregressive (AR) processes with state-dependent switching, which are two-component Markov processes. The state-dependent switching model is a nontrivial generalization of Markovian switching formulation and it includes the Markovian switching as a special case. We prove the Feller and strong Feller continuity by means of introducing auxiliary processes and making use of the Radon–Nikodym derivatives. Then, we investigate the geometric ergodicity by the Foster–Lyapunov inequality. Moreover, we establish the $V$-uniform ergodicity by means of introducing additional auxiliary processes and by virtue of constructing certain order-preserving couplings of the original as well as the auxiliary processes. In addition, illustrative examples are provided for demonstration.

© 2009 Elsevier Inc. All rights reserved.
\end{abstract}

1. Introduction

In this work, we concern ourselves with a discrete-time, nonlinear autoregressive model modulated by a random switching process. The resulting stochastic process has components, namely, the primary sequence and the modulating sequence. The two components are jointly Markov. As a convention, we call the primary sequence the “state” and view the modulating sequence as the “mode”. In the previous work, a commonly used formulation of randomly switching AR models assumes the modulating component being a Markov chain independent of the state, which is termed a Markovian switching autoregressive process. In this paper, we consider a much more difficult case, namely, the modulating component depends on the state. As a result, this switching process alone is not Markov, but only the two-component process bundled together is a Markov process. When the state-dependent switching process disappears, the model reduces to Markovian switching case. Thus, our model includes the Markovian switching models as a special case.

The purpose of the modulating switching process is to describe uncertainty due to random environment and other stochastic discrete events that are not representable by the primary component. The formulation of state-dependent mode enables us to describe more complex systems and their inherent uncertainty and randomness. However, it adds much difficulty in analysis. Our study is largely motivated by applications arising in nonlinear time series, discrete optimization, communication networks, signal processing, and financial engineering among others. We focus on asymptotic properties such as regularity, Feller properties, and ergodicity. Because our formulation includes the Markovian switching, all results obtained in this paper also hold for Markovian switching AR models.

\textsuperscript{*} Corresponding author.
E-mail addresses: xfb@bit.edu.cn (F. Xi), gyn@math.wayne.edu (G. Yin).

0047-259X/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
Consider a finite set $S := \{1, 2, \ldots, n_0\}$ and suppose that $(X_n, Z_n)$ is a strong Markov process whose state space is $\mathbb{R}^d \times S$. The evolution of the process is given by

$$X_n = f_n(X_{n-1}) + \varepsilon_n, \quad X_n \in \mathbb{R}^d,$$

and

$$P(Z_n = l | Z_{n-1} = k, X_{n-1} = x) = p_{kl}(x), \quad k, l \in S, \ x \in \mathbb{R}^d.$$

Note that (1.1) delineates the dynamics of the primary component—the state, whereas (1.2) describes the evolution of the jumps. In case $Z_n$ is a discrete-time Markov chain taking values in $S$, the transition probabilities in (1.2) become the usual Markovian one without $x$ dependence. The difficulty we encounter here is the $x$-dependent switching process since the processes $X_n$ and $Z_n$ are interdependent.

In this work, the error process $\{\varepsilon_n\}$ is assumed to be a sequence of independent and identically distributed (i.i.d.) $\mathbb{R}^d$-valued random variables, $\{f_k : k \in S\}$ is a collection of nonlinear autoregressive functions, and $P(x) = (p_{kl}(x))$ is a transition probability matrix for each $x \in \mathbb{R}^d$. We call the strong Markov process $(X_n, Z_n)$ a nonlinear autoregressive (AR) process with state-dependent switching. In particular, when the functions $p_{kl}(x)$ in (1.2) are independent of $x$ for all $k, l \in S$ and the second component $Z_n$, independent of the error process $\varepsilon_n$, is a Markov chain itself, the corresponding strong Markov process $(X_n, Z_n)$ then can be called a nonlinear autoregressive (AR) process with Markovian switching. Clearly, the state-dependent switching is a nontrivial generalization of the Markovian counterpart. On the other hand, for each $k \in S$, we can determine a nonlinear autoregressive process $X_n^{(k)}$ by

$$X_n^{(k)} = f_k(X_{n-1}^{(k)}) + \varepsilon_n, \quad X_n^{(k)} \in \mathbb{R}^d.$$

Here each $f_k(\cdot)$ together $\varepsilon_n$ can be called a regime. Equivalently, for each $k \in S$, $X_n^{(k)}$ can also be viewed as a regime. Therefore, we may also call $(X_n, Z_n)$ a regime-switching autoregressive process.

Regime-switching autoregressive models have received enormous attention in the past two decades. In the late 1980s, a series of papers were published regarding regime-switching time series models [1–3]. The regime-switching formulation soon attracted needed attention and resulted in an extensive study on econometric series modeling. Nonlinear AR processes with Markovian switching was studied in [4]; it was noted that the model can be regarded as an extension of hidden Markov models. Among other things, $V$-uniform ergodicity for nonlinear AR processes with Markovian switching was obtained in the aforementioned paper. One should also be mentioned that in econometrics, in 1982, a related model, namely, autoregressive conditional heteroscedasticity (ARCH) model came into being [5], which stimulated much of subsequent study in such models and generalizations. One of the main ingredients is to consider the variance of the current error term as a function of the variances of the that of the previous terms. It is commonly used in financial engineering to capture the time-varying volatility clustering. Nowadays, generalized autoregressive conditional heteroscedasticity (GARCH) models have been widely used by researchers and practitioners in financial market analysis. Nonlinear autoregressive models have also been used in many fields of science and engineering. For example, hidden Markov models and autoregressive models in connection with stochastic approximation algorithms and applications to multiuser detection in wireless networks were considered in [6]. We note that a nonlinear AR process with random switching can also be regarded as a discrete-time hybrid system. Recently, much work has been devoted to continuous-time hybrid systems [7–17] and the references therein, whereas the investigation on the discrete-time hybrid systems is still relatively scarce.

**Example 1.1.** As a motivation of our study, consider the following stochastic approximation problem. Suppose that $g(\cdot, \cdot) : \mathbb{R}^d \times \mathbb{R}^d \mapsto \mathbb{R}^d$. Consider a stochastic recursive algorithm of the form

$$\theta_{n+1} = \theta_n + \epsilon g(\theta_n, X_n),$$

where $\epsilon > 0$ is a sufficiently small number representing the stepsize of the algorithm, $\{X_n\}$ serves as a noise process that is given by (1.1), and $\{\theta_n\}$ is a sequence of estimates recursively constructed. The purpose of (1.4) is either for finding the roots of a function whose precise form is unknown or too complicated to compute, or for locating the minimizer of an appropriate “cost” function. In each case, only noise corrupted observations or measurements are available. For further motivation and an up-to-date treatment of stochastic approximation methods, we refer the reader to [18]. In examining the stochastic approximation problem, the long-time behavior of $X_n$ (i.e., ergodicity) is of crucial importance. Many stochastic approximation problems arising in applications require us to treat random noises of various forms. Because randomly varying switching nonlinear AR processes naturally arise in a wide variety of cases, it is necessary to study the convergence of stochastic approximation under such random influence. The results to be presented in this paper pave a way for handling the asymptotic behavior of $X_n$ thereby contribute to the study of convergence and rate of convergence of the associated stochastic approximation problems.

In this paper, we provide a comprehensive and systematic study on asymptotic properties of AR processes with state-dependent switching. In addition to Feller properties, we consider the geometric ergodicity and $V$-uniform ergodicity for the nonlinear AR process with state-dependent switching $(X_n, Z_n)$ defined by (1.1) and (1.2).
The rest of the paper is arranged as follows. Section 2 proves Feller and strong Feller continuity for \((X_n, Z_n)\) by means of introducing an auxiliary process \((V_n, \Psi_n)\) and by making use of the Radon–Nikodym derivative of the transition probability of \((X_n, Z_n)\) with respect to the transition probability of \((V_n, \Psi_n)\). Section 3 proves the aperiodicity for \((X_n, Z_n)\) and investigates its geometric ergodicity by using the Foster–Lyapunov inequality. Based on coupling techniques, Section 4 introduces another auxiliary process \((U_n, \Phi_n)\) and constructs an order-preserving coupling of \((X_n, Z_n)\) and \((U_n, \Phi_n)\). In Section 5, we show the so-called contraction inequality (see (9) in [4]) for the transition probability of \((X_n, Z_n)\) by virtue of the above order-preserving coupling and then prove the \(V\)-uniform ergodicity for \((X_n, Z_n)\). A number of illustrative examples are provided in Sections 3–5 for demonstration purposes. Finally, the paper is concluded with further remarks in Section 6.

2. Feller properties

In this section we prove the Feller continuity and strong Feller continuity for the nonlinear AR process with state-dependent switching \((X_n, Z_n)\). To do so, we make the following assumption.

Assumption 2.1. Assume that all the functions \(f_k(x), p_{kl}(x), k, l \in \mathbb{S}\), are continuous and that the random variable \(\varepsilon_1\) has a density with respect to the Lebesgue measure on \(\mathbb{R}^d\).

Now we introduce an auxiliary process \((V_n, \Psi_n)\). Let the first component \(V_n\) satisfy

\[
V_n = f_{\Psi_n}(V_{n-1}) + \varepsilon_n, \quad V_n \in \mathbb{R}^d,
\]

and the second component \(\Psi_n\) be a Markov chain with transition probabilities

\[
P(\Psi_n = l|\Psi_{n-1} = k) = \frac{1}{n_0}, \quad k, l \in \mathbb{S},
\]

where \(n_0\) is the number of the elements in \(\mathbb{S}\). Note that the auxiliary process \((V_n, \Psi_n)\) is a nonlinear AR process with Markovian switching. Note that the transition probabilities of the Markov chain are uniformly distributed among the Markov states in \(\mathbb{S}\). Moreover, the Markov chain is irreducible. For definiteness, we denote the process \((V_n, \Psi_n)\) determined by (2.1) and (2.2) with initial condition \((V_0, \Psi_0) = (x, k)\) by \((V^{x,k}_n, \Psi^{x,k}_n)\). Likewise, we denote the process \((X_n, Z_n)\) determined by (1.1) and (1.2) with initial condition \((X_0, Z_0) = (x, k)\) by \((X^{x,k}_n, Z^{x,k}_n)\).

Now we define a metric \(\lambda(\cdot, \cdot)\) on \(\mathbb{R}^d \times \mathbb{S}\) as follows:

\[
\lambda((x, m), (y, n)) = |x - y| + d(m, n),
\]

where

\[
d(m, n) = \begin{cases} 0, & m = n, \\ 1, & m \neq n. \end{cases}
\]

Let \(\mathcal{B}(\mathbb{R}^d \times \mathbb{S})\) be the Borel \(\sigma\)-algebra on \(\mathbb{R}^d \times \mathbb{S}\). Therefore, \((\mathbb{R}^d \times \mathbb{S}, \lambda(\cdot, \cdot), \mathcal{B}(\mathbb{R}^d \times \mathbb{S}))\) is a locally compact and separable metric space. We denote the one-step transition probabilities of the strong Markov processes \((X_n, Z_n)\) and \((V_n, \Psi_n)\) by \(P((x, k), A) : (x, k) \in \mathbb{R}^d \times \mathbb{S}, A \in \mathcal{B}(\mathbb{R}^d \times \mathbb{S}))\) and \(Q((x, k), A) : (x, k) \in \mathbb{R}^d \times \mathbb{S}, A \in \mathcal{B}(\mathbb{R}^d \times \mathbb{S}))\), respectively. It can be seen that for each \((x, k) \in \mathbb{R}^d \times \mathbb{S}, P((x, k), \cdot)\) is absolutely continuous with respect to \(Q((x, k), \cdot)\) and the corresponding Radon–Nikodym derivative has the following form:

\[
\frac{dP((x, k), \cdot)}{dQ((x, k), \cdot)}|_{(y, 0)} = n_0 p_{\Psi_1}(x).
\]

Lemma 2.2. Suppose that Assumption 2.1 holds. For any given \(k \in \mathbb{S}\), we have that

\[
E\lambda\left((V_1^{x,k}, \Psi_1^{x,k}), (V_1^{y,k}, \Psi_1^{y,k})\right) \to 0
\]

as \(|x - y| \to 0\).

Proof. From (2.1) and (2.2), \(V_1^{x,k} = f_{\Psi_1}(x) + \varepsilon_1\) and \(V_1^{y,k} = f_{\Psi_1}(y) + \varepsilon_1\). Thus, it follows from the continuity of \(f_l(x)\) for all \(l \in \mathbb{S}\) (see Assumption 2.1) that

\[
E|V_1^{x,k} - V_1^{y,k}| = E|f_{\Psi_1}(x) - f_{\Psi_1}(y)| = \frac{1}{n_0} \sum_{l \in \mathbb{S}} |f_l(x) - f_l(y)| \to 0
\]

as \(|x - y| \to 0\). This clearly implies (2.4). \(\square\)

Proposition 2.3. Suppose that Assumption 2.1 holds. The one-step transition probability \(P((x, k), A) : (x, k) \in \mathbb{R}^d \times \mathbb{S}, A \in \mathcal{B}(\mathbb{R}^d \times \mathbb{S}))\) of the strong Markov process \((X_n, Z_n)\) is Feller continuous.

Proof. For any given bounded continuous function \(f(x, k)\) on \(\mathbb{R}^d \times \mathbb{S}\), from (2.3) we have that for all \((x, k) \in \mathbb{R}^d \times \mathbb{S},

\[
Ef(X_1^{x,k}, Z_1^{x,k}) = n_0 Ef(V_1^{x,k}, \Psi_1^{x,k})p_{\Psi_1}(x).
\]

(2.5)
Next, combining (2.4) with the continuity of \( f(x, k) \), we also have that

\[
\begin{align*}
&f(V_1^{x,k}, \Psi_1^k) \to f(V_1^{y,k}, \Psi_1^k) \quad \text{in probability} \\
\text{as } |x - y| \to 0.
\end{align*}
\]

Then, similarly to the proof of Proposition 1.2 in [19], for any given \( \delta > 0 \), using (2.5), we have

\[
\begin{align*}
\left| Ef(X_1^{x,k}, Z_1^{x,k}) - Ef(X_1^{y,k}, Z_1^{y,k}) \right| &\leq n_0 E|f(V_1^{x,k}, \Psi_1^k)p_{x\Psi_1^k}(x) - f(V_1^{y,k}, \Psi_1^k)p_{y\Psi_1^k}(y)| + \delta n_0 Ep_{x\Psi_1^k}(y) \notag \\
&\leq n_0 \|f\| \|p_{x\Psi_1^k}(x) - p_{y\Psi_1^k}(y)\| + \delta n_0 Ep_{y\Psi_1^k}(y) \notag \\
&+ 2n_0 \|f\| \int_{(f(V_1^{x,k}, \Psi_1^k) - f(V_1^{y,k}, \Psi_1^k)) \geq \delta} p_{y\Psi_1^k}(y) dP \\
&= (I) + (II) + (III).
\end{align*}
\]

(2.7)

Here and hereafter, we put \( \|f\| := \sup \{|f(x, k)| : (x, k) \in \mathbb{R}^d \times \mathbb{S}\} \). From the continuity of \( p_{kl}(x) \) for all \( k, l \in \mathbb{S} \) (see Assumption 2.1), we see that term (I) in (2.7) tends to zero as \( |x - y| \to 0 \). From (2.6), we derive that term (III) in (2.7) also tends to zero as \( |x - y| \to 0 \). At the same time, we also know that term (II) in (2.7) can be arbitrarily small since the multiplier \( \delta \) is arbitrary. Finally, combining these three facts with (2.7) and that \( \mathbb{S} \) has discrete metric together, we obtain the desired result. \( \square \)

To establish the strong Feller continuity for \((X_n, Z_n)\), let us fix a probability measure \( \mu(\cdot) \) which is equivalent to the product measure on \( \mathbb{R}^d \times \mathbb{S} \) of the Lebesgue measure on \( \mathbb{R}^d \) and the counting measure on \( \mathbb{S} \). Moreover, we also need to prove two lemmas. We first give the following elementary lemma.

**Lemma 2.4.** For any given bounded measurable function \( f(x, k) \) on \( \mathbb{R}^d \times \mathbb{S} \) and any given positive number \( \delta > 0 \), there exists a compact subset \( D \subset \mathbb{R}^d \) such that \( \mu(D^c \times \mathbb{S}) < \delta \) and \( f\mid_{D \times \mathbb{S}} \), the function \( f(x, k) \) confined on \( D \times \mathbb{S} \), is uniformly continuous.

**Proof.** Although this lemma can also be derived from the Luzin Theorem (cf. Theorem 6.3 in Chapter 5 of [20]), now we would like to prove it by the monotone class theorem. Denote by \( \mathcal{L} \) the family of the bounded measurable functions on \( \mathbb{R}^d \times \mathbb{S} \). Set

\[
\mathcal{L} := \{f(x, k) : \exists \text{ a compact subset } D \subset \mathbb{R}^d \text{ such that } \mu(D^c \times \mathbb{S}) < \delta \text{ and } f\mid_{D \times \mathbb{S}} \text{ is uniformly continuous}\}.
\]

According to the definition of \( \mathcal{L} \)-system (refer to Definition 1.34 in [21]), one can verify that \( \mathcal{L} \) is an \( \mathcal{L} \)-system. Moreover, let \( \mathcal{C} \) denote the set of all the open sets in \( \mathbb{R}^d \times \mathbb{S} \). Obviously, \( \mathcal{C} \) is a \( \pi \)-system. Note that \( \mathcal{L} \) contains the set of all bounded Lipschitz continuous functions on \( \mathbb{R}^d \times \mathbb{S} \). By virtue of the monotone class theorem (refer to Theorem 1.35 in [21] again), we then get that \( \mathcal{L} \) contains the set of all bounded measurable functions on \( \mathbb{R}^d \times \mathbb{S} \). This completes the proof. \( \square \)

**Lemma 2.5.** Suppose that Assumption 2.1 holds. For any given bounded measurable function \( f(x, k) \) on \( \mathbb{R}^d \times \mathbb{S} \), we have that

\[
\begin{align*}
f(V_1^{x,k}, \Psi_1^k) &\to f(V_1^{y,k}, \Psi_1^k) \quad \text{in probability} \\
as |x - y| \to 0.
\end{align*}
\]

(2.8)

**Proof.** Since the random variable \( \varepsilon_1 \) has a density with respect to the Lebesgue measure on \( \mathbb{R}^d \), we easily know that the one-step transition probability \( Q((x, k), \cdot) \) of the auxiliary process \((V_n, \Psi_n)\) is absolutely continuous with respect to the probability measure \( \mu(\cdot) \). Thus, by Radon–Nikodym theorem, we obtain that \((V_n, \Psi_n)\) has transition probability density such that for any \((x, k) \in \mathbb{R}^d \times \mathbb{S}\), \(A \in \mathcal{B}(\mathbb{R}^d)\) and \(l \in \mathbb{S}\),

\[
P((V_1, \Psi_1) \in A \times \{l\} | V_0 = x, \Psi_0 = k) = \int_A p((x, k), (y, l)) \mu(dy \times \{l\}).
\]

(2.9)

Since the auxiliary process \((V_n, \Psi_n)\) is a nonlinear AR process with Markovian switching, it has the strong Feller property (see Lemma 1 of [4]). Hence, for any sequence \(\{x_n\}\) tending to \(x\) and for any given \(g(y, l) \in L^\infty(\mu)\), we have

\[
\sum_{k \in \mathbb{S}} \int g(y, l) p((x_n, k), (y, l)) \mu(dy \times \{l\}) \to \sum_{k \in \mathbb{S}} \int g(y, l) p((x, k), (y, l)) \mu(dy \times \{l\})
\]

as \(n \to \infty\). Namely, when \(n \to \infty\), \(p((x_n, k), \cdot)\) converges weakly to \(p((x, k), \cdot)\) in \(L_1(\mu)\) (cf. Definition 5.8 in Chapter 7 of [20]). Thus, by Dunford–Petit theorem (cf. Theorem 5.10 in Chapter 7 of [20]), we obtain that the family \(\{p((x_n, k), \cdot) : n \geq 1\}\) is uniformly integrable in \(L_1(\mu)\). Hence for any given \(\varepsilon > 0\), there exists \(\delta > 0\) such that for all \(A \in \mathcal{B}(\mathbb{R}^d)\), if \(\mu(A \times \mathbb{S}) < \delta\), then for all \(n \geq 1\),

\[
P((V_1, \Psi_1) \in A \times \mathbb{S} | V_0 = x_n, \Psi_0 = k) \leq \sum_{l \in \mathbb{S}} \int_A p((x_n, k), (y, l)) \mu(dy \times \{l\}) < \varepsilon
\]

(2.10)
and
\[
P( (V_1, \Psi_1) \in A \times S | V_0 = x, \Psi_0 = k ) = \sum_{l \in S} \int_A p(x, k, (y, l)) \mu(dy \times \{l\}) < \varepsilon. \tag{2.11}
\]

On the other hand, by Lemma 2.4, we can find a compact subset \( D \subset \mathbb{R}^d \) such that \( \mu(D^c \times \mathbb{R}) < \delta \) and \( f|_{D \times \mathbb{R}} \) is uniformly continuous. Namely, for any given \( \eta > 0 \), there exists \( \delta_1 > 0 \) such that for all \( (x, k), (x', k) \in D \times \mathbb{S} \), if \( |x - x'| < \delta_1 \), then \( |f(x, k) - f(x', k)| < \eta \) for all \( k \in \mathbb{S} \). Therefore, from (2.10) and (2.11) we arrive at
\[
P( |f(V_1^{x,k}, \Psi_1^k) - f(V_1^{x',k}, \Psi_1^k)| > \eta ) \leq P( |V_1^{x,k} - V_1^{x',k}| > \delta_1 ) + P( (V_1^{x,k}, \Psi_1^k) \notin D \times \mathbb{S} ) + P( (V_1^{x',k}, \Psi_1^k) \notin D \times \mathbb{S} ) \]
\[
\leq P( |V_1^{x,k} - V_1^{x',k}| > \delta_1 ) + 2\varepsilon. \tag{2.12}
\]

Meanwhile, recalling that \( V_1^{x,k} = f_{\varphi_1^k}(x_0) + \varepsilon_1 \) and \( V_1^{x',k} = f_{\varphi_1^k}(x) + \varepsilon_1 \), by the proof of Lemma 2.2, we have that
\[
P( |V_1^{x,k} - V_1^{x',k}| > \delta_1 ) \leq \frac{1}{n_0\delta_1} \sum_{l \in \mathbb{S}} |f_l(x_0) - f_l(x)| \to 0 \quad \text{as} \quad n \to \infty. \tag{2.13}
\]

Inserting (2.13) into (2.12) and noting that \( \varepsilon \) and \( \eta \) are arbitrary, we conclude that (2.8) holds. This completes the proof. \( \square \)

**Theorem 2.6.** Suppose that Assumption 2.1 holds. The one-step transition probability \( P((x, k), A) : (x, k) \in \mathbb{R}^d \times \mathbb{S}, A \in \mathcal{B}(\mathbb{R}^d \times \mathbb{S}) \) of the strong Markov processes \((X_n, Z_n)\) is strongly Feller continuous.

**Proof.** According to the definition of strong Feller continuity, it is enough to prove that for any bounded measurable function \( f(x, k) \) on \( \mathbb{R}^d \times \mathbb{S} \), \( Ef(1_{X_1^k, Z_1^k}) \) is bounded continuous in both \( x \) and \( k \). Since \( \mathbb{S} \) is bounded and has discrete metric, it is sufficient to prove that
\[
|Ef(1_{X_1^k, Z_1^k}) - Ef(1_{X_1^k, Z_1^k})| \to 0 \tag{2.14}
\]
as \( |x - y| \to 0 \). Indeed, using (2.8) instead of (2.6), we can obtain (2.14) by the proof of Proposition 2.3. The proof is complete. \( \square \)

**Remark 2.7.** We should point out that for the continuous-time models, in Section 1.3 of [19], the strong Feller continuity was proved for a degenerate diffusion process using an auxiliary diffusion process with transition probability density. Moreover, in [12], the Feller continuity was proved for a diffusion process with state-dependent switching via a diffusion process with Markovian switching. Here we consider the similar problems for a nonlinear autoregressive process with state-dependent switching which is a discrete-time model.

3. Geometric ergodicity

In this section we prove the aperiodicity for \((X_n, Z_n)\) and investigate its geometric ergodicity by using the Foster-Lyapunov inequality. To do so, we first make the following assumption.

**Assumption 3.1.** Assume that \( p_{i,l}(x) > 0 \) for all \( k \neq l \in \mathbb{S} \) and \( x \in \mathbb{R}^d \), and that the random variable \( \varepsilon_1 \) has an everywhere positive and continuous density \( g(x) \) with respect to the Lebesgue measure \( m(\cdot) \) on \( \mathbb{R}^d \).

**Lemma 3.2.** Suppose that Assumptions 2.1 and 3.1 hold. Then all compact subsets of \( \mathbb{R}^d \times \mathbb{S} \) are pettie and \((X_n, Z_n)\) is aperiodic (see [22] for the definitions of petite sets and aperiodicity).

**Proof.** By Assumption 3.1, we know that \((X_n, Z_n)\) is irreducible with respect to the reference measure \( \mu(\cdot) \) defined in Section 2. From Proposition 2.3, we know that \((X_n, Z_n)\) \has the Feller property. Obviously, \( \text{supp } \mu(\cdot) = \mathbb{R}^d \times \mathbb{S} \) which has non-empty interior. Combining these three facts with Theorem 3.4 in [23], we obtain that all compact subsets of \( \mathbb{R}^d \times \mathbb{S} \) are petite.

Next we prove the aperiodicity. Let \( p^{(x,k)} \) denote the distribution of \((X_n, Z_n)\) starting from \((x, k)\). Let \( C \subset \mathbb{R}^d \) be a non-empty compact subset with \( m(C) > 0 \). Then, by the continuity of the functions \( p_{i,l}(x), f_i(x), k, l \in \mathbb{S} \), and \( g(x) \), we have that for all \((x, k) \in C \times \mathbb{S}, A \in \mathcal{B}(\mathbb{R}^d) \) and \( l \in \mathbb{S} \),
\[
p^{(x,k)}( (X_1, Z_1), A \times \{l\} ) = P((X_1, Z_1) \in A \times \{l\} | X_0 = x, Z_0 = k )
= P(Z_1 = l | X_0 = x, Z_0 = k ) P(X_1 \in A | X_0 = x, Z_0 = k, Z_1 = l )
= p_{i,l}(x) \int_A g(y - f_i(x)) dy
\geq \int_A \min\{p_{i,l}(x) g(y - f_i(x)) : x \in C, k, l \in \mathbb{S}\} dy, \tag{3.1}
\]
where we have used that \( f_i(x) + \varepsilon_1 \) has the density \( g(y - f_i(x)) \) since \( \varepsilon_1 \) has the density \( g(x) \). In particular, we have that for all \((x,k) \in C \times S \) and \( l \in S \),

\[
P^{(x,k)}((X, l) C \times \{l\}) \geq \int_C \min\{p_{il}(x)g(y - f_i(x)) : x \in C, k, l \in S\}dy
\]

\[
\geq \min\{p_{il}(x)g(y - f_i(x)) : x, y \in C, k, l \in S\}m(C) > 0.
\]

(3.2)

Therefore, in view of the definition of aperiodicity given in Section 5.4.3 of [22], from (3.1) and (3.2) we can derive that \((X_n, Z_n)\) is aperiodic. \( \square \)

Now we proceed to investigate the geometric ergodicity for \((X_n, Z_n)\). As in [23], for any positive function \( V(x, k) \geq 1 \) defined on \( \mathbb{R}^d \times S \) and any signed measure \( \omega(\cdot) \) defined on \( B(\mathbb{R}^d \times S) \) we write

\[
\|\omega\|_V = \text{sup}\{|\omega(g)| : \text{all measurable } g(x, k) \text{ satisfying } |g| \leq V\},
\]

(3.3)

where \( \omega(g) \) denotes the integral of function \( g \) with respect to measure \( \omega \). Moreover, for a function \( \infty > V(x, k) \geq 1 \) on \( \mathbb{R}^d \times S \), \((X_n, Z_n)\) is said to be geometrically ergodic if there exists a probability measure \( \pi(\cdot) \), a positive constant \( \theta < 1 \) and a finite-valued function \( \Theta(x, k) \) such that

\[
\|P^n(\cdot) - \pi(\cdot)\|_V \leq \Theta(x, k)\theta^n
\]

(3.4)

for all \( n \geq 0 \) and all \((x,k) \in \mathbb{R}^d \times S \). Moreover, a nonnegative function \( \bar{V}(x, k) \) defined on \( \mathbb{R}^d \times S \) is called a norm-like function if \( \bar{V}(x, k) \to \infty \) as \( |x| \to \infty \) for all \( k \in S \). Now we proceed to investigate the geometric ergodicity for \((X(n), Z(n))\).

For this we make the following assumption on the Foster–Lyapunov inequality.

**Assumption 3.3.** Assume that there exist a norm-like function \( \bar{V}(x, k) \) and constants \( 0 \leq \alpha < 1 \) and \( 0 \leq \beta < \infty \) such that

\[
P\bar{V}(x, k) \leq \alpha \bar{V}(x, k) + \beta, \quad (x, k) \in \mathbb{R}^d \times S.
\]

(3.5)

**Theorem 3.4.** Suppose that Assumptions 2.1, 3.1 and 3.3 hold. Then \((X_n, Z_n)\) is geometrically ergodic with \( V(x, k) = \bar{V}(x, k) + 1 \) and \( \Theta(x, k) = B(\bar{V}(x, k) + 1) \), where \( B \) is a finite constant.

**Proof.** Clearly, (3.4) implies the discrete drift condition (DD4) in [23]. Thus, by virtue of Theorem 6.3 in [23] and Lemma 3.2, we obtain the desired result. \( \square \)

**Remark 3.5.** Theorem 3.4 is a rather general result and its key condition is the existence of Foster–Lyapunov function \( \bar{V}(\cdot) \) in (3.5). For a given concrete model, how to identify a Foster–Lyapunov function is a very important issue. It is known that finding the right Foster–Lyapunov functions is never a simple or systematic task. Hence, one naturally hope to find some explicit conditions on the functions \( f_k(\cdot) \) and \( p_{il}(\cdot) \) (1.1) and (1.2) which guarantee the existence of Foster–Lyapunov function \( \bar{V}(\cdot) \) in (3.5). We will accomplish this task in Section 5.

**Example 3.6.** Let \( d = 1 \) and \( S = \{1, 2\} \). Consider the following simple autoregressive process with Markovian switching:

\[
X_n = \lambda Z_n X_{n-1} + \varepsilon_n, \quad X_n \in \mathbb{R}^1
\]

(3.6)

where \((Z_n)\) is a Markov chain with transition probability matrix \( P = (p_{kl}) \) given by

\[
p_{11} = \frac{4}{5}, \quad p_{12} = \frac{1}{5}, \quad p_{21} = \frac{9}{10}, \quad p_{22} = \frac{1}{10};
\]

(3.7)

\( \varepsilon_n \) has the one-dimensional standard normal distribution; and constants \( \lambda_1 = 1 / \sqrt{5} \) and \( \lambda_2 = \sqrt{2} \). Intuitively, The first component of \((X_n, Z_n)\) can be regarded as the result of the following two autoregressive processes:

\[
X^{(1)}_n = \frac{1}{\sqrt{5}} X^{(1)}_{n-1} + \varepsilon_n \quad \text{and} \quad X^{(2)}_n = \sqrt{2} X^{(2)}_{n-1} + \varepsilon_n
\]

switching back and forth from one to the other according to the movement of \((Z_n)\). It is easy to see that \( X^{(1)}_n \) is stable while \( X^{(2)}_n \) is unstable since the corresponding coefficients \( 1 / \sqrt{5} < 1 \) whereas \( \sqrt{2} > 1 \). But, we can prove that the switching autoregressive process \((X_n, Z_n)\) is geometrically ergodic. To this end, by virtue of Theorem 3.4, we need only verify that Assumption 3.3 holds since Assumptions 2.1 and 3.1 clearly hold.

To do so, we define a norm-like function \( \bar{V}(x, k) \) on \( \mathbb{R}^1 \times \{1, 2\} \) as \( \bar{V}(x, k) = a_k x^2 \) with \( a_1 = 1 \) and \( a_2 = 2 \). Note that

\[
P\bar{V}(x, k) = \sum_{l=1}^{2} \int_{\mathbb{R}^1} \bar{V}(y, l)P((x, k), dy \times \{l\}).
\]

(3.8)

Denote by \( \psi(x) \) the density of standard normal distribution. Then \( \lambda_k x + \varepsilon_1 \) has the density \( \psi(y - \lambda_k x) \). Therefore, it follows
from (3.8) that
\[
P\widehat{V}(x, k) = \sum_{l=1}^{2} \int_{\mathbb{R}^1} a_l y^2 p_{kl}(y - \lambda_l x) dy
\]
\[
= \sum_{l=1}^{2} p_{kl} a_l \int_{\mathbb{R}^1} y^2 \psi(y - \lambda_l x) dy
\]
\[
= \sum_{l=1}^{2} p_{kl} a_l (1 + \lambda_l^2 x^2),
\]
which can be rewritten as
\[
P\widehat{V}(x, k) = \frac{p_{k1} a_1 \lambda_1^2 + p_{k2} a_2 \lambda_2^2}{a_k} \widehat{V}(x, k) + (p_{k1} a_1 + p_{k2} a_2).
\]
(3.9)

Recalling the concrete the values of \(p_{kl}, a_k\) and \(\lambda_k\) defined above, we know that the coefficients
\[
p_{11} a_1 \lambda_1^2 + p_{12} a_2 \lambda_2^2 = \frac{24}{25} < 1 \quad \text{and} \quad p_{21} a_1 \lambda_1^2 + p_{22} a_2 \lambda_2^2 = \frac{29}{100} < 1.
\]
Clearly, this and (3.9) together imply (3.5), and hence Assumption 3.3 holds.

**Example 3.7.** Let \(Z_n^+\) be the usual integer lattice. Consider the following regime-switching Bernoulli random walk \((X_n, Z_n)\) on \(Z_n^+ \times \{1, 2\}\):
\[
X_n = X_{n-1} + \xi(Z_n)\left(\xi(x, n-1) = 1 \right) + \chi(Z_n = 0) \chi(\xi(Z_n) = 1),
\]
(3.10)
\[
P(\xi(Z_n) = 1 | Z_n = 1) = 1 - P(\xi(Z_n) = -1 | Z_n = 1) = p_1 = \frac{1}{4},
\]
(3.11)
\[
P(\xi(Z_n) = 1 | Z_n = 2) = 1 - P(\xi(Z_n) = -1 | Z_n = 2) = p_2 = \frac{3}{4},
\]
(3.12)
where \((Z_n)\) is a Markov chain with transition probability matrix \(P = (p_{kl})\) given by
\[
p_{11} = \frac{24}{25}, \quad p_{12} = \frac{1}{25}, \quad p_{21} = \frac{9}{10}, \quad p_{22} = \frac{1}{10}.
\]
(3.13)

Obviously, \((X_n, Z_n)\) has the strong Feller property since its state space is discrete. Clearly, for \(\xi(Z_n) = 1\) and \(\xi(Z_n) = 2\), we have two corresponding Bernoulli walks \(X_n^{(1)}\) and \(X_n^{(2)}\) respectively (cf. Section 15.5.1 of [22] or Section 7 of [24]). Moreover, it is easy to see that \(X_n^{(1)}\) is recurrent whereas \(X_n^{(2)}\) is transient. But, we can prove that the regime-switching Bernoulli random walk \((X_n, Z_n)\) is geometrically ergodic.

To do so, we define a norm-like function \(\widehat{V}(x, k)\) on \(Z_n^+ \times \{1, 2\}\) as \(\widehat{V}(x, k) = \gamma^x = (\sqrt{3})^x\). Note that for all \(x \geq 1\) and \(k \in \{1, 2\}\),
\[
P\widehat{V}(x, k) = \sum_{i=1}^{2} (p_i p_{ki} \gamma^{x+i} + (1 - p_i) p_{ki} \gamma^{x-1})
\]
\[
= \sum_{i=1}^{2} (p_i p_{ki} \gamma + (1 - p_i) p_{ki} \gamma^{-1}) \widehat{V}(x, k).
\]
(3.14)
Recalling the concrete the values of \(p_i, p_{kl}\) and \(\gamma\) defined above, we know that the coefficients
\[
\sum_{i=1}^{2} (p_i p_{1i} \gamma + (1 - p_i) p_{1i} \gamma^{-1}) = \frac{77}{150} \sqrt{3} \approx 0.8891 < 1
\]
and
\[
\sum_{i=1}^{2} (p_i p_{2i} \gamma + (1 - p_i) p_{2i} \gamma^{-1}) = \frac{8}{15} \sqrt{3} \approx 0.9238 < 1.
\]
Analogously to Example 3.6, from this and (3.14), we can prove that \((X_n, Z_n)\) is geometrically ergodic.

**Example 3.8.** This is a continuation of the stochastic approximation problem in Example 1.1. Assume the conditions of Theorem 3.4 hold for the process \(\{(X_n, Z_n)\}\). Suppose that for each \(x, g(\cdot, x)\) is a continuous function and that \(\{X_n\}\) is a
sequence of bounded noise. By Theorem 3.4, for each \( \theta \) and some \( \ell > 0 \), \( \frac{1}{n} \sum_{k=\ell}^{\ell+n-1} E_{\mathcal{F}_k} g(\theta, X_k) \rightarrow \overline{g}(\theta) \) in probability as \( n \rightarrow \infty \), where \( E_{\mathcal{F}_k} \) denotes the conditional expectation up to “time” \( \ell \), and \( \overline{g}(\theta) = \int g(\theta, x) \pi(dx) \), with \( \pi(x) \) being the invariant measure given in Theorem 3.4. Define a piecewise constant interpolation \( \theta^\epsilon(t) = \theta_t \) for \( t \in [n\epsilon, n\epsilon + \epsilon) \). Using the techniques of martingale averaging [18, Chapter 8], we can show that \( \theta^\epsilon(\cdot) \) converges weakly to \( \theta(\cdot) \), which is a solution of the martingale problem with a degenerate generator, or equivalently, \( \theta(\cdot) \) is the solution of the ordinary differential equation \( \frac{d}{dt} \theta(t) = \overline{g}(\theta(t)) \), provided the differential equation above has a unique solution. Note that the stationary point of the above differential equation, namely, the solution of \( \overline{g}(\theta) = 0 \) corresponds to either the root of \( \overline{g} \) or the minimizer of an objective function \( J(\cdot) \) with \( \nabla J(\theta) = \overline{g}(\theta) \). Suppose that \( \theta_\ast \) is the unique solution of \( \overline{g}(\theta) = 0 \) and it is stable in the sense of Lyapunov [18, p. 104]. Choose a sequence \( t_\epsilon \rightarrow \infty \) as \( \epsilon \rightarrow 0 \). Then it can be shown \( \theta^\epsilon(\cdot + t_\epsilon) \) converges in probability to \( \theta_\ast \).

Moreover, we can study the rates of convergence of the algorithm. Theorem 3.4 implies that \( \{(X_n, Z_n)\} \) is mixing with exponential mixing rate [25]. The exponential mixing is a consequence of the geometric ergodicity. Define \( u_n = (\theta - \theta_\ast) / \sqrt{\epsilon} \).

Taking truncated Taylor expansions of \( g(\theta, x) \) about \( \theta_\ast \) and under suitable conditions such \( g_{\theta,\theta}(\theta, x) \) the Hessian of \( g \) with respect to \( \theta \) is bounded together additional conditions, we can write \( u_{n+1} = u_n + \epsilon g_{\theta,\theta}(\theta_\ast, X_n)u_n + \sqrt{\epsilon} E g(\theta_\ast, X_n) + o(\epsilon) \), where \( o(\epsilon) \rightarrow 0 \) in probability. Assume \( \overline{g}_{\theta,\theta}(\theta_\ast) \) is Hurwitz (with all of its eigenvalue having negative real parts). Define \( u^\epsilon(t) = u_n \) for \( t \in [\epsilon(n - N_\epsilon), \epsilon(n - N_\epsilon + 1)] \). Then we can show \( \frac{1}{n} \sum_{k=\ell}^{\ell+n-1} E_{\mathcal{F}_k} g(\theta_\ast, X_k) \rightarrow \overline{g}(\theta_\ast) \) in probability as \( n \rightarrow \infty \) and \( \sqrt{\epsilon} \sum_{k=0}^{[t]} g(\theta_\ast, X_k) \) converges weakly to a Brownian motion \( \Sigma^{1/2} w(t) \), where \([z]\) denotes the integer part of \( z \), \( w(\cdot) \) is a standard Brownian motion, \( z^\epsilon \) denotes the transpose of \( z \), and

\[
\Sigma = Eg(\theta_\ast, X_0)g(\theta_\ast, X_0) + \sum_{k=1}^{\infty} Eg(\theta_\ast, X_k)g(\theta_\ast, X_k) + \sum_{k=1}^{\infty} Eg(\theta_\ast, X_k)g(\theta_\ast, X_0).
\]

Thus, we can show that \( u^\epsilon(\cdot) \) converges weakly to the solution of the diffusion

\[
du = \overline{g}_\theta(\theta_\ast) u dt + \Sigma^{1/2} dw.
\]

As in [18, Chapter 10], the scaling factor \( \sqrt{\epsilon} \) together with the stationary covariance \( \Sigma \), gives us the rate of convergence. Note that \( \Sigma \) can be found from the Lyapunov equation \( \overline{g}_\theta(\theta_\ast) \Sigma + \Sigma \overline{g}_\theta(\theta_\ast) = -\Sigma \). Note also that if \( g(\theta, x) = \overline{g}(\theta) + x \), then the noise appears in an additive form. In lieu of the boundedness condition for the noise, we require only \( \sup_n E|X_n|^{2 + \Delta} < \infty \) for some \( \Delta > 0 \). Then the results still hold. Interested reader is referred to [18] for further details.

4. Order-preserving coupling

As is well known, the coupling methods have a very wide range of applications (refer to [21]). One application is that the study of complex processes is boiled down to the study of some simple ones. For this the order-preserving couplings usually play an important role. For such an application in what follows, in this section we construct an order-preserving coupling. To do so, we need to introduce the following definition of the stochastic comparability.

Definition 4.1. Two transition probability matrices \( p^{(1)} = (p_{kl}^{(1)}) \) and \( p^{(2)} = (p_{kl}^{(2)}) \) on \( S \) are said to be stochastically comparable if

\[
\sum_{j \geq m} p_{kj}^{(1)} \leq \sum_{j \geq m} p_{kj}^{(2)}, \quad m \in S, \quad k \leq l \in S.
\]

We write it as \( p^{(1)} \preceq p^{(2)} \) briefly.

Remark 4.2. A real function \( g(k) \) on \( S \) is called monotone if for all \( k \leq l \in S, g(k) \leq g(l) \). One can check the definition \( p^{(1)} \preceq p^{(2)} \) given in (4.1) is equivalent to that

\[
\sum_{j \in S} p_{kj}^{(1)} g(j) \leq \sum_{j \in S} p_{kj}^{(2)} g(j)
\]

for any monotonic function \( g(j) \) on \( S \) and any \( k \leq l \in S \).

Stochastic comparability for jump processes was studied in [26,27]. Moreover, it was proved in [28] that two transition probabilities on a Polish space, are stochastically comparable if and only if there exists an order-preserving Markovian coupling of them (see Section 2 of [28] for the details).

Assumption 4.3. Assume that \( (\Phi_n) \) is a positive recurrent Markov chain on \( S \) with transition probability matrix \( P = (p_{kl}) \) such that \( P(x) \preceq P \) for all \( x \in \mathbb{R}^d \).

By the Markov chain \( (\Phi_n) \) with transition probability matrix \( P = (p_{kl}) \), we introduce another auxiliary process \( (U_n, \Phi_n) \) such that the first component \( U_n \) satisfy

\[
U_n = f_{\Phi_n}(U_{n-1}) + \varepsilon_n, \quad U_n \in \mathbb{R}^d.
\]
Proposition 4.4. Suppose that Assumption 4.3 holds. There exists an order-preserving Markovian coupling \((X_n, Z_n, U_n, \Phi_n)\) of \((X_n, Z_n)\) and \((U_n, \Phi_n)\) such that for any positive integer \(n \geq 1, x, y \in \mathbb{R}^d\) and \(k \leq l \in \mathbb{S},\)

\[P(Z_1 \leq \Phi_1, Z_2 \leq \Phi_2, \ldots, Z_n \leq \Phi_n | X_0 = x, Z_0 = k, U_0 = y, \Phi_0 = l) = 1.\]  \hfill (4.4)

**Proof.** In view of the proof methods of Theorems 2.3 and 2.4 in [28], we readily have an order-preserving Markovian coupling \((X_n, Z_n, U_n, \Phi_n)\) such that for \(x, y \in \mathbb{R}^d\) and \(k \leq l \in \mathbb{S},\)

\[P(Z_1 \leq \Phi_1 | X_0 = x, Z_0 = k, U_0 = y, \Phi_0 = l) = 1.\]

Then, from the Markov property it follows that

\[P(Z_1 \leq \Phi_1, Z_2 \leq \Phi_2 | X_0 = x, Z_0 = k, U_0 = y, \Phi_0 = l) = 1.\]

Furthermore, we can derive the desired result (4.4) from the Markov property. \(\square\)

**Remark 4.5.** For any positive integer \(n \geq 1,\) a multivariate function \(f(k_1, k_2, \ldots, k_n)\) on \(\mathbb{S}^n\) is said to be monotone if

\[f(k_1, k_2, \ldots, k_n) \leq f(l_1, l_2, \ldots, l_n)\]

holds for all \((k_1, k_2, \ldots, k_n), (l_1, l_2, \ldots, l_n) \in \mathbb{S}^n\) satisfying \(k_1 \leq l_1, k_2 \leq l_2, \ldots, k_n \leq l_n.\) Clearly, by virtue of the marginality of the order-preserving coupling constructed in Proposition 4.4, for any monotonic multivariate function \(f(k_1, k_2, \ldots, k_n)\) on \(\mathbb{S}^n\) we have that

\[E(f(Z_1, Z_2, \ldots, Z_n) | X_0 = x, Z_0 = k) \leq E(f(\Phi_1, \Phi_2, \ldots, \Phi_n) | U_0 = y, \Phi_0 = l)\]  \hfill (4.5)

holds for any \(x, y \in \mathbb{R}^d\) and \(k \leq l \in \mathbb{S}.

To conclude this section, we provide two examples about the stochastic comparability of transition probability matrices of \(P(x) = (p_{ui}(x))\) and \(P = (p_{ui})\).

**Example 4.6.** Take \(d = 1, \mathbb{S} = \{1, 2\}\) and let

\[P(x) = (p_{ui}(x)) = \begin{pmatrix} 3 & 1 \\ 4 & 4 + 4 \cos x \end{pmatrix} \quad \text{and} \quad P = (p_{ui}) = \begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix}.
\]

According to Definition 4.1, it can be verified that \(P(x) \preceq P\) for all \(x \in \mathbb{R}^1.\)

**Example 4.7.** Take \(d = 1, \mathbb{S} = \{1, 2, 3\}\) and let

\[P(x) = (p_{ui}(x)) = \begin{pmatrix} 5 & -1 & 1 \\ 14 & \sin x \cos x + 1 & \sin x^2 \\ 2 & 3 & 1 - \frac{1}{6} \cos x \end{pmatrix} \quad \text{and} \quad P = (p_{ui}) = \begin{pmatrix} 1 & 2 & 2 \\ 7 & 7 & 7 \\ 1 & 1 & 1 \end{pmatrix}.
\]

According to Definition 4.1, it can also be verified that \(P(x) \preceq P\) for all \(x \in \mathbb{R}^1.\)

5. \(V\)-uniform ergodicity

In this section, we investigate the \(V\)-uniform ergodicity for the nonlinear AR process with state-dependent switching \((X_n, Z_n).\) We first recall the following definition from [22]. For a function \(\infty > V(x, k) \geq 1\) on \(\mathbb{R}^d \times \mathbb{S},\) Markov process
(Xn, Zn) is said to be V-uniformly ergodic if there exists a probability measure π(·) such that
\[
\sup \left\{ \frac{\| P((x, k), \cdot) - \pi(\cdot) \|_V}{V(x, k)} : (x, k) \in \mathbb{R}^d \times S \right\} \to 0 \quad \text{as} \ n \to \infty. \tag{5.1}
\]

Next, we proceed to investigate the V-uniform ergodicity for the strong Markov process (Xn, Zn). For this we make some assumptions.

**Assumption 5.1.** Assume that there exist positive constants (a0, b0, c0) such that for all (x, k) ∈ \mathbb{R}^d × S,
\[
|f_k(x)| \leq a_k |x| + b_k \tag{5.2}
\]
with order a1 ≤ a2 ≤ ··· ≤ an0.

**Assumption 5.2.** Assume that \( p_{kl}(x) > 0 \) for all k ≠ l ∈ S and x ∈ \mathbb{R}^d, that the random variable \( \varepsilon_1 \) has an everywhere positive density with respect to the Lebesgue measure on \( \mathbb{R}^d \) and that there exists a positive number \( \gamma > 0 \) such that \( E|\varepsilon_1|^{\gamma} < \infty \).

**Assumption 5.3.** Assume that \((\Phi_n)\) is a positive recurrent Markov chain on S with transition probability matrix \( P = (p_{kl}) \) and invariant probability measure \( \nu = (\nu_1, \nu_2, \ldots, \nu_m) \) such that \( P(x) \leq \beta \) for all \( x \in \mathbb{R}^d \) and \( \beta := \sum_{i \in S} \nu_i \log a_i < 0 \).

Note that **Assumption 5.3** implies **Assumption 4.3**. Now we are in a position to formulate the main result of this section.

**Theorem 5.4.** Suppose that **Assumptions 2.1 and 5.1–5.3** hold. The nonlinear AR process with state-dependent switching \((X_n, Z_n)\) is V-uniformly ergodic.

**Proof.** Following the proof of Theorem 1 in [4], for any positive integer \( n \geq 1 \), by **Assumption 5.1** we also have
\[
|X_1|^{1/n} = |f_{n-1}(X_{n-1}) + \varepsilon_n|^{1/n} \leq (a_{n-1}|X_{n-1}| + b_{n-1} + |\varepsilon_n|)^{1/n} \leq (a_{n-1})^{1/n}|X_{n-1}|^{1/n} + (b_{n-1} + |\varepsilon_n|)^{1/n}.
\]
Furthermore, recursively as in the proof of (8) in [4], we also have
\[
|X_1|^{1/n} \leq (a_1 \cdots a_{n-1})^{1/n}|X_0|^{1/n} + (b_1 \cdots b_{n-1} + |\varepsilon_1|)^{1/n} + \sum_{i=1}^{n-1} (a_1 \cdots a_{n-1})^{1/n}(b_1 \cdots b_{n-1} + |\varepsilon_i|)^{1/n}. \tag{5.3}
\]
On the other hand, since \((\Phi_n)\) is a positive recurrent Markov chain on finite set S, it follows from Birkhoff ergodic theorem and **Assumption 5.3** that for any initial condition \( \Phi_0 = k \in S \),
\[
\lim_{n \to \infty} \frac{1}{n} \left( \log(a_{\Phi_1}) + \cdots + \log(a_{\Phi_n}) \right) \to \sum_{i \in S} \nu_i \log a_i < 0, \quad \text{a.s.}
\]
Similarly to the proof of Theorem 1 in [4] again, from this we obtain that for any \((x, k) \in \mathbb{R}^d \times S \),
\[
\lim_{n \to \infty} E\left( (a_{\Phi_1} \cdots a_{\Phi_n})^{1/n} \mid U_0 = x, \Phi_0 = k \right) = a_{1}^{\nu_1} \cdots a_{n_0}^{\nu_n} < 1. \tag{5.4}
\]
Note that \( f(k_1, k_2, \ldots, k_n) = a_{k_1} a_{k_2} \cdots a_{k_n} \) is a monotonic multivariate function on \( \mathbb{R}^n \) due to that \( a_1 \leq a_2 \leq \cdots \leq a_{n_0} \). Combining (4.5) and (5.4), we derive that for any \((x, k) \in \mathbb{R}^d \times S \),
\[
\lim_{n \to \infty} E\left( (a_{Z_1} \cdots a_{Z_n})^{1/n} \mid X_0 = x, Z_0 = k \right) \leq a_{1}^{\nu_1} \cdots a_{n_0}^{\nu_n} < 1. \tag{5.5}
\]
Thus, by **Assumption 5.2**, there exists a positive integer \( m \geq 1/\gamma \) such that
\[
\alpha_m := \sup \left\{ E\left( (a_{Z_1} \cdots a_{Z_m})^{1/m} \mid X_0 = x, Z_0 = k \right) : (x, k) \in \mathbb{R}^d \times S \right\} < 1.
\]
Therefore, taking conditional expectation of the inequality (5.3), with arbitrary initial condition \((X_0, Z_0) = (x, k) \in \mathbb{R}^d \times S \), yields
\[
E\left( |X_1|^{1/m} \mid X_0 = x, Z_0 = k \right) = E_m|x|^{1/m} + \beta_m, \tag{5.6}
\]
where
\[
\beta_m := \sup \left\{ E\left( (b_{Z_m} + |\varepsilon_m|)^{1/m} + \sum_{i=1}^{m-1} (a_{Z_m} \cdots a_{Z_{i+1}})^{1/m}(b_{Z_i} + |\varepsilon_i|)^{1/m} \mid X_0 = x, Z_0 = k \right) : (x, k) \in \mathbb{R}^d \times S \right\}.
\]
By $1/m \leq \gamma$ and Assumption 5.2 we know that $\beta_m < +\infty$. Set the Lyapunov function $V(x, k) = |x|^{1/m} + 1$ on $\mathbb{R}^d \times S$ and denote by

$$P^m = \{p^m((x, k), A) : (x, k) \in \mathbb{R}^d \times S, A \in \mathcal{B}(\mathbb{R}^d \times S)\}$$

the $m$-step transition probability of $(X_n, Z_n)$. Clearly, (5.6) implies that for $(x, k) \in \mathbb{R}^d \times S$,

$$P^m V(x, k) \leq \alpha_n V(x, k) + \beta_m + 1 - \alpha_n,$$  \hspace{1cm} (5.7)

where $P^m V(x, k) := \sum_{y \in S} \int_{S \times \mathbb{R}^d} V(y, l) p^m((x, k), dy \times \{l\})$. As explained in the proof of Theorem 1 in [4], the geometric drift condition (i.e., the contraction inequality) (5.7) for the $m$-step transition probability $P^m((x, k), \cdot)$ implies that the same also holds for the one-step transition probability $P((x, k), \cdot)$ with some larger Lyapunov function $V_0(x, k) \geq V(x, k)$ (see [22], pp. 386–387 for the details).

As in the proof of Lemma 3.2, by Assumption 5.2 and Proposition 2.3, we can prove that all compact subsets of $\mathbb{R}^d \times S$ are petite and $(X_n, Z_n)$ is aperiodic. Hence, by virtue of Theorem 16.1.2 in [22], $(X_n, Z_n)$ is $V_0$-uniformly ergodic. □

Example 5.5. Let $d = 1$ and $S = \{1, 2\}$, and take $P(x) = (p_{1h}(x))$ and $P = (p_{1h})$ as in Example 4.6. Moreover, set $a_1 = \frac{1}{2}$ and $a_2 = \frac{4}{7}$, and let $\varepsilon_1$ have the standard normal distribution $N(0, 1)$. By an elementary calculation, we get that $v = (v_1, v_2) = (\frac{1}{2}, \frac{7}{8})$. Furthermore, it is also verifiable that Assumptions 2.1 and 5.1–5.3 all hold. In view of Theorem 5.4, we derive that the nonlinear AR process with state-dependent switching $(X_n, Z_n)$ is $V$-uniformly ergodic.

Example 5.6. Let $d = 1$ and $S = \{1, 2, 3\}$, and take $P(x) = (p_{ih}(x))$ and $P = (p_{ih})$ as in Example 4.7. Moreover, set $a_1 = \frac{1}{4}$, $a_2 = \frac{1}{2}$ and $a_3 = 2$, and let $\varepsilon_1$ have the standard normal distribution $N(0, 1)$. By an elementary calculation, we get that $v = (v_1, v_2, v_3) = (\frac{35}{113}, \frac{42}{113}, \frac{36}{113})$. Furthermore, it can also be verified that Assumptions 2.1 and 5.1–5.3 all hold. In view of Theorem 5.4, we derive that the nonlinear AR process with state-dependent switching $(X_n, Z_n)$ is also $V$-uniformly ergodic.

6. Further remarks

This paper developed Feller properties as well as ergodicity for a class of nonlinear autoregressive processes with state-dependent switching. The desired results were obtained by introducing auxiliary processes, the use of Radon–Nikodym derivatives, and coupling techniques. For further investigation, several problems are worthwhile to look into. In the current setup, (1.1), the random process $\{\varepsilon_n\}$ is a sequence of independent and identically distributed random variables. It is of practical interest to consider the problem that such a noise process has dependent structure. For example, what can we say if the process is certain mixing process with mixing rate decaying sufficiently fast. Second, in lieu of (1.1), can we handle $X_n = f_m(X_{n-1}, X_{n-2}, \ldots, X_{n-q}) + \varepsilon_n$, for some $q > 1$? This will be appealing for people who need to deal with autoregressive processes in applications with order of regressions beyond order 1. Currently, the current setup requires the process $(X_n, Z_n)$ to be Markov. What happens if such Markovian assumption is no longer available or violated? Can we obtain similar results for process being non-Markovian but close to or approximately Markovian? All of these questions require much thoughts and investigation.

Acknowledgments

We thank Professors Mu-Fa Chen, Yong-Hua Mao, and Yu-Hui Zhang for the discussions, comments, and suggestions. We also thank the referees and the editors for their comments and suggestions.

The research of first author was supported in part by the National Natural Science Foundation of China under Grant No. 10671037.

The research of second author was supported in part by the National Science Foundation under DMS-0907753 and in part by the National Security Agency under grant MSPF-068-029.

References