J. Math. Anal. Appl. 553 (2026) 129838

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

journal homepage: www.elsevier.com/locate/jmaa

Regular Articles

Stabilization of regime-switching processes based on discrete time observations: Existence of invariant probability measure

Jianrui Li, Jinghai Shao *, Xiao-Yu Zhao

Center for Applied Mathematics, Tianjin University, Tianjin 300072, China

ARTICLE INFO

Article history:
Received 23 March 2025
Available online 27 June 2025
Submitted by B. Jourdain

Keywords:
Regime-switching
Discrete time observations
Invariant measure
Stability in distribution

ABSTRACT

Given an unstable SDE with regime-switching, we provide explicit conditions to stabilize this system in distribution. Since the instability of original system allows that the drift coefficient does not satisfy any dissipative condition, we hence essentially use the delay term to ensure the existence of invariant probability measure. This result remains meaningful even for stochastic processes without regime-switching. Two methods are used: the first method uses the Krylov-Bogoliubov theorem by viewing the segment process as a Markov process in the infinite dimension path space. The second method takes advantage of the special structure of the controlled system to construct a family of embedded finite dimension Markov processes.

1. Introduction

Consider a Markovian regime-switching process

$$\mathrm{d}\widetilde{X}(t) = b(\widetilde{X}(t), \Lambda(t))\mathrm{d}t + \sigma(\widetilde{X}(t), \Lambda(t))\mathrm{d}W(t), \tag{1.1}$$

where $(\Lambda(t))$ is a Markov chain on a countable state space $\mathcal{S} = \{1, 2, \dots, N\}$, $N \le \infty$, $(W(t))$ is a d-dimension Wiener process, $b : \mathbb{R}^d \times \mathcal{S} \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \times \mathcal{S} \to \mathbb{R}^{d\times d}$. If $\widetilde{X}(\cdot) \equiv 0$ is an equilibrium state of the original system (1.1), but it is not stable, one wants to add additional term to stabilize it. For the sake of saving cost and being more realistic, Mao [16] proposed to stabilize the previous system using a feedback control based on discrete time observations of $(\widetilde{X}(t))$ as follows:

$$\mathrm{d}\widetilde{X}(t) = \big(b(\widetilde{X}(t), \Lambda(t)) - u(\widetilde{X}(\delta_t), \Lambda(t))\big)\mathrm{d}t + \sigma(\widetilde{X}(t), \Lambda(t))\mathrm{d}W(t), \tag{1.2}$$

where $\delta_t = [t/\delta]\delta$ for some $\delta > 0$ and $[t/\delta]$ denotes the integer part of t/δ.

* Corresponding author.
E-mail addresses: jianruili@tju.edu.cn (J. Li), shaojh@tju.edu.cn (J. Shao), zhxy_0628@tju.edu.cn (X.-Y. Zhao).

https://doi.org/10.1016/j.jmaa.2025.129838

Inspired by this work, many researches have been devoted to this topic. For instance, [17,26] aimed to find a better lower bound of time duration δ between two consecutive observations; [12,23] and [20,22] used different methods to stabilize the unstable system based on both discrete time observations of $(\widetilde{X}(t))$ and $(\Lambda(t))$; [11] designed a periodically intermittent controller based on discrete time observations of $(\widetilde{X}(t))$; [13] and [6] designed a controller based on the response lags and discrete time observations of $(\widetilde{X}(t))$ and $(\Lambda(t))$.

All the previously mentioned works focus on the stability of equilibrium state $X(\cdot) \equiv 0$ in various senses including moments stability and almost sure stability. However, the stability of equilibrium state is sometimes too strong, so it is of meaning to investigate the solutions to be stable in distribution. The stability in distribution has been studied in many works (with some subtle differences). For instance, [3] for diffusion processes with degenerate diffusion coefficients; [27] for SDEs with Markovian regime-switching; [28] for delay-dependent SDEs with Markovian regime-switching. Our purpose of this work is to stabilize the unstable system (1.1) in distribution via feedback controls based on discrete time observations of $(\widetilde{X}(t))$ and $(\Lambda(t))$.

A little precisely, we aim to design a controller based on the discrete time observations such that the obtained stochastic system

$$\mathrm{d}X(t) = \big(b(X(t), \Lambda(t)) - u(X(\delta_t), \Lambda(\delta_t))\big)\mathrm{d}t + \sigma(X(t), \Lambda(t))\mathrm{d}W(t) \tag{1.3}$$

admits a limit in distribution as $t \to \infty$, where $\delta_t = [t/\delta]\delta$. Since the controlled system $(X(t), \Lambda(t))$ in (1.3) becomes a path-dependent SDE, it is natural to investigate the existence of limit distribution for the segment process (X_t, Λ_t), where $X_t \in \mathscr{C}([-\delta, 0]; \mathbb{R}^d)$ is defined as $X_t(s) = X((t+s) \vee 0)$, $-\delta \le s \le 0$, and $\Lambda_t \in \mathscr{D}([-\delta, 0]; \mathcal{S})$ defined as $\Lambda_t(s) = \Lambda((t+s) \vee 0)$ for $s \in [-\delta, 0]$.

SDE (1.3) is a kind of stochastic functional differential equation (SFDE). Many literatures have been devoted to investigating the stationary distributions of SFDEs, see, for instance, [1,5,7,9] and [2] for SFDEs with regime-switching. As shown in [5], a typically crucial condition to ensure the existence of invariant probability measure for the SFDE

$$\mathrm{d}X(t) = f(X_t)\mathrm{d}t + g(X_t)\mathrm{d}W(t)$$

is in the form

$$\langle f(\phi), \phi(0)\rangle \le -c|\phi(0)|^p, \quad \phi \in \mathscr{C}([-\delta, 0]; \mathbb{R}^d) \tag{1.4}$$

for some $p \in (0, 1]$, $c > 0$. Namely, the drift coefficient f acting on $\phi \in \mathscr{C}([-\delta, 0]; \mathbb{R}^d)$ must be dissipative w.r.t. the current state $\phi(0)$.

To see the difficulty to stabilize SDE (1.3) in distribution, let us consider a simple example. The coefficient b can be extended as a function on $\mathscr{C}([-\delta, 0]; \mathbb{R}^d)$ by setting $b(\phi, i) = b(\phi(0), i)$, $\phi \in \mathscr{C}([-\delta, 0]; \mathbb{R}^d)$, $i \in \mathcal{S}$. Since the original system is unstable, it is possible that $b(x, i)$ is in the form

$$b(\phi, i) = \phi(0), \quad \phi \in \mathscr{C}([-\delta, 0]; \mathbb{R}^d),\ i \in \mathcal{S}, \quad \text{and hence } \langle b(\phi, i), \phi(0)\rangle = |\phi(0)|^2,$$

which obviously violates (1.4) and cannot satisfy any dissipative condition. Therefore, the main challenge to stabilize (1.3) lies in how to use the delay term $u(\phi(\delta_t), \Lambda(\delta_t))$ to ensure the existence of invariant probability measure of (X_t, Λ_t) in the infinite dimensional space $\mathscr{C}([-\delta, 0]; \mathbb{R}^d) \times \mathscr{D}([-\delta, 0]; \mathcal{S})$.

We use two methods to realize the stabilization of (1.3) in distribution in this work. Our first method is based on the idea of [9] and [16]. We view (X_t, Λ_t) as a Markov process in the path space $\mathscr{C}([-\delta, 0]; \mathbb{R}^d) \times \mathscr{D}([-\delta, 0]; \mathcal{S})$, and notice that the difference $\mathbb{E}|X(t) - X(\delta_t)|^6$ can be controlled by $\mathbb{E}|X(t)|^6$ when δ is small (see Lemma 2.3), which guarantees the existence of an invariant probability measure for (X_t, Λ_t)

under appropriate explicit conditions (see Theorem 2.8). Moreover, in this case we actually show that the distribution of (X_t, Λ_t) converges weakly to its invariant probability measure in $\mathscr{C}([-\delta,0];\mathbb{R}^d)\times\mathscr{D}([-\delta,0];\mathcal{S})$.

Our second method takes advantage of the special structure of SFDE (1.3). To show the idea, let us consider only a simple case here. Note that for $r \in (0,\delta)$, the process $\big(X(n\delta+r), X(n\delta), \Lambda(n\delta)\big)_{n\geq 0}$ is a Markov process on $\mathbb{R}^{2d}\times\mathcal{S}$. This permits us to use the criterion on the existence of invariant probability measure for the finite dimension Markov process (cf. e.g. [14,15]). Then, via the Kolmogorov extension theorem, we obtain a probability measure μ on $\mathscr{C}([-\delta,0];\mathbb{R}^d)\times\mathscr{D}([-\delta,0];\mathcal{S})$ such that the distribution of $(X_{n\delta}, \Lambda_{n\delta})$ converges in finite dimension projection to μ as $n\to\infty$. The convergence in finite dimension projection is weaker than the weak convergence in $\mathscr{C}([-\delta,0];\mathbb{R}^d)\times\mathscr{D}([-\delta,0];\mathcal{S})$. But, the conditions on the term $u(x,i)$ are also much weaker than those imposed in the first method. See Theorem 3.3 and Remark 3.5 below.

This work is organized as follows. In Section 2, we investigate the existence and uniqueness of invariant probability measure for SDE (1.3) in the infinite dimensional path space $\mathscr{C}([-\delta,0];\mathbb{R}^d)\times\mathscr{D}([-\delta,0];\mathcal{S})$. In Section 3, we first verify the structure of embedded Markov processes of SDE (1.3). Then, using the criterion on the existence of finite dimensional Markov processes, we obtain the existence of a unique invariant probability measure for each embedded Markov process. At last, we obtain the desired probability measure on $\mathscr{C}([-\delta,0];\mathbb{R}^d)\times\mathscr{D}([-\delta,0];\mathcal{S})$ by virtue of the Kolmogorov extension theorem, which is the limit distribution of $(X_{n\delta}, \Lambda_{n\delta})$ as $n\to\infty$ in the sense of convergence in finite dimension projection.

2. An approach in an infinite dimensional path space

In this section we shall view (X_t,Λ_t) as a Markov process in the path space, and use the Krylov-Bogoliubov theorem to show the existence of the invariant probability measure. Then, by showing the t_0-regularity of the corresponding semigroup, the uniqueness of the invariant probability measure is also shown.

Denote $\mathscr{C}([-\delta,0];\mathbb{R}^d)$ the continuous path space from $[-\delta,0]$ to $\mathbb{R}^d$, endowed with the uniform norm $\|\phi\| = \sup_{s\in[-\delta,0]}|\phi(s)|$. Define $\mathscr{D}([-\delta,0];\mathcal{S})$ the collection of right continuous functions from $[-\delta,0]$ to $\mathcal{S}$ with left limits, endowed with Skorokhod's topology (cf. [4]). In this work $\mathscr{L}_\xi$ denotes the distribution of random variable ξ. $\|\nu_1-\nu_2\|_{\mathrm{var}} := 2\sup_{A\in\mathscr{B}(E)}|\nu_1(A)-\nu_2(A)|$ denotes the total variation distance between two probability measures ν_1 and ν_2 on a measurable space $(E,\mathscr{B}(E))$.

Define the segment process (X_t,Λ_t) on $\mathscr{C}([-\delta,0];\mathbb{R}^d)\times\mathscr{D}([-\delta,0];\mathcal{S})$ for the solution $(X(t),\Lambda(t))$ to SDE (1.3) by

$$X_t(s) = X((t+s)\vee 0),\quad \Lambda_t(s) = \Lambda((t+s)\vee 0),\quad s\in[-\delta,0],\ t\geq 0.$$

We collect the conditions on the coefficients used in this work here.

(H1) There exist constants $\alpha,\beta\geq 0$ such that

$$\max\{|b(x,i)|^2, \|\sigma(x,i)\|_{\mathrm{HS}}^2\} \leq \alpha|x|^2+\beta,\quad x\in\mathbb{R}^d,\ i\in\mathcal{S},$$

where $\|\sigma(x,i)\|_{\mathrm{HS}}^2 := \mathrm{tr}(\sigma\sigma^*)(x,i)$ and $\sigma^*(x,i)$ denotes the transpose of matrix $\sigma(x,i)$.

(H2) There exist constants η, $\hat\eta$, c_0, $c_1, K_u\geq 0$ such that

$$|u(x,i)-u(y,i)|\leq K_u|x-y|,\quad |u(x,i)|^2\leq c_0+c_1|x|^2,$$
$$\eta|x|^2\leq\langle x, u(x,i)\rangle\leq\hat\eta|x|^2,\quad x,y\in\mathbb{R}^d,\ i\in\mathcal{S}.$$

(H3) There is a constant $\tilde K$ such that

$$|b(x,i)-b(y,i)|+\|\sigma(x,i)-\sigma(y,i)\|_{\mathrm{HS}}\leq\tilde K|x-y|,\quad x,\ y\in\mathbb{R}^d,\ i\in\mathcal{S}.$$

(H4) There exists a constant $c_2 > 0$ such that

$$\xi^*\sigma(x,i)\sigma^*(x,i)\xi \geq c_2|\xi|^2, \quad \forall \xi \in \mathbb{R}^d, x \in \mathbb{R}^d, i \in \mathcal{S}.$$

(H5) $Q = (q_{ij})_{i,j\in\mathcal{S}}$ is a conservative, irreducible transition rate matrix. Assume that $M := \sup\{-q_{ii}; i \in \mathcal{S}\} < \infty$ and there exist a function $H : \mathcal{S} \to [1,\infty)$, constants $\kappa_1, \kappa_2 > 0$ such that

$$\lim_{j\to\infty} H(j) = \infty, \quad QH(i) := \sum_{j\in\mathcal{S}} q_{ij}H(j) \leq -\kappa_1 H(i) + \kappa_2, \quad i \in \mathcal{S}. \tag{2.1}$$

The process $(\Lambda(t))$ is a continuous-time Markov chain with transition rate matrix $(q_{ij})_{i,j\in\mathcal{S}}$. $(\Lambda(t))$ and $(W(t))$ are assumed to be mutually independent throughout this work. It is known that the drift condition (2.1) means that the process $(\Lambda(t))_{t\geq 0}$ is exponentially ergodic, and there is a stationary distribution γ of $(\Lambda(t))$ on $\mathcal{S}$ such that

$$\lim_{t\to\infty} \|\mathbb{P}\big(\Lambda(t) \in \cdot\,|\Lambda(0) = i\big) - \gamma\|_{\mathrm{var}} = 0, \qquad \forall\, i \in \mathcal{S}.$$

Remark 2.1. Under the conditions (H1)-(H3), SDE (1.3) admits a unique strong solution (cf. e.g. [20]). Condition (H4) is used to ensure the validation of φ-irreducibility of skeleton process of $(X(t), \Lambda(t))$ in order to apply the Lyapunov criterion studied in Meyn and Tweedie [15], which can be replaced by weaker conditions such as hypoelliptic condition. We refer to [24] for the study of strong Feller property of regime-switching processes under the hypoelliptic condition.

Definition 2.2. Let $(X(t), \Lambda(t))_{t\geq 0}$ be the solution to (1.3).

(i) We say that $(X(t), \Lambda(t))_{t\geq 0}$ is *strongly stable in distribution* if the distribution of (X_t, Λ_t) converges weakly to some probability measure μ on $\mathscr{C}([-\delta, 0]; \mathbb{R}^d) \times \mathscr{D}([-\delta, 0]; \mathcal{S})$ for every initial value $(X(0), \Lambda(0)) = (x, i) \in \mathbb{R}^d \times \mathcal{S}$ as $t \to \infty$.
(ii) It is said $(X(t), \Lambda(t))_{t\geq 0}$ is *weakly stable in distribution* if there is a probability measure μ on $\mathscr{C}([-\delta, 0]; \mathbb{R}^d) \times \mathscr{D}([-\delta, 0]; \mathcal{S})$ such that for any finite dimension projection map $\pi_F \times \pi_\Lambda$, $\mathscr{L}_{(X_{(n+1)\delta}, \Lambda_{(n+1)\delta})} \circ (\pi_F \times \pi_\Lambda)^{-1}$ converges weakly to $\mu \circ (\pi_F \times \pi_\Lambda)^{-1}$ for every initial value $(X(0), \Lambda(0)) = (x, i) \in \mathbb{R}^d \times \mathcal{S}$ as $n \to \infty$, where

$$\pi_F : \mathscr{C}([-\delta, 0]; \mathbb{R}^d) \to \mathbb{R}^{(k+1)d},\ \phi \mapsto \big(\phi(-(\delta - r_k)), \ldots, \phi(-(\delta - r_1)), \phi(-\delta)\big)$$
$$\pi_\Lambda : \mathscr{D}([-\delta, 0]; \mathcal{S}) \longrightarrow \mathcal{S}, \quad \psi \mapsto \psi(-\delta),$$

associated with $F = \{0 = r_0 < r_1 < \ldots < r_k < \delta\}$ for some $k \in \mathbb{N}$.

In view of the fact that the Markov chain $(\Lambda(t))$ is assumed to be exponentially ergodic with stationary distribution γ, the limit of segment process $\mathscr{L}_{\Lambda_t}$ in $\mathscr{D}([-\delta, 0]; \mathbb{R}^d)$ is clearly equal to $\gamma^{[-\delta,0]}$, denoting the infinite product measure of $(\mathcal{S}, \gamma)$ with index set $[-\delta, 0]$. Therefore, in the definition of weak stability in distribution, we only consider the projection π_Λ to map $\psi \in \mathscr{D}([-\delta, 0]; \mathcal{S})$ to $\psi(-\delta) \in \mathcal{S}$, which is sufficient to guarantee the finite projection process $(\pi_F \times \pi_\Lambda)(X_{(n+1)\delta}, \Lambda_{(n+1)\delta})$, $n \geq 0$, to be a Markov process (see Lemma 3.1).

In practical applications, the stability of a given system is often checked or used on a sequence of discrete times $\tau, 2\tau, \ldots, n\tau, \ldots$ with a time step size $\tau > 0$. This induces many works to study how to stabilize a system based on discrete time observations (cf. [12,13,16] and references therein). If a system is weak stable in distribution, then this system is stable enough with respect to the discrete observations times $\tau, 2\tau, \ldots$ with suitable $\tau > 0$.

Lemma 2.3. *Assume* (H1), (H2) *and* (H5) *hold. If* $\delta > 0$ *is sufficiently small so that*

$$K(\delta) := 2^5(12\alpha + c_1)\delta \mathrm{e}^{(6+60\alpha+2c_1+12\beta+2c_0)\delta} < 1, \tag{2.2}$$

then for any $t > 0$,

$$\mathbb{E}|X(t) - X(\delta_t)|^6 \le \frac{K(\delta)}{1-K(\delta)}\mathbb{E}|X(t)|^6 + \frac{(6\beta + c_0)\delta \mathrm{e}^{(6+60\alpha+2c_1+12\beta+2c_0)\delta}}{1-K(\delta)}. \tag{2.3}$$

Proof. Note that for each $t \in [n\delta, (n+1)\delta)$ with $n \in \mathbb{Z}_+$, $\delta_t = n\delta$ is a fixed point. Then, applying Itô's formula and (H1), (H2), we obtain that

$$\begin{aligned}
&\mathrm{d}|X(t) - X(\delta_t)|^6 \\
&= 3|X(t) - X(\delta_t)|^4\big(2\langle X(t) - X(\delta_t), b(X(t), \Lambda(t)) - u(X(\delta_t), \Lambda(\delta_t))\rangle \\
&\quad + \|\sigma(X(t), \Lambda(t))\|_{\mathrm{HS}}^2\big)\mathrm{d}t + 6|X(t) - X(\delta_t)|^4\langle X(t) - X(\delta_t), \sigma(X(t), \Lambda(t))\mathrm{d}W(t)\rangle \\
&\quad + 12|X(t) - X(\delta_t)|^2\Big[\sum_{k=1}^{d}\Big(\sum_{l=1}^{d}\sigma_{lk}(X(t), \Lambda(t))(X_l(t) - X_l(\delta_t))\Big)^2\Big]\mathrm{d}t \\
&\le 3|X(t) - X(\delta_t)|^4\big\{2|X(t) - X(\delta_t)|^2 + |b(X(t), \Lambda(t))|^2 + |u(X(\delta_t), \Lambda(\delta_t))|^2 \\
&\quad + \|\sigma(X(t), \Lambda(t))\|_{\mathrm{HS}}^2\big\}\mathrm{d}t + 12|X(t) - X(\delta_t)|^4\|\sigma(X(t), \Lambda(t))\|_{\mathrm{HS}}^2\mathrm{d}t \\
&\quad + 6|X(t) - X(\delta_t)|^4\langle X(t) - X(\delta_t), \sigma(X(t), \Lambda(t))\mathrm{d}W(t)\rangle \\
&\le 3|X(t) - X(\delta_t)|^4\big\{2|X(t) - X(\delta_t)|^2 + 6\alpha|X(t)|^2 + 6\beta + c_0 + c_1|X(\delta_t)|^2\big\}\mathrm{d}t \\
&\quad + 6|X(t) - X(\delta_t)|^4\langle X(t) - X(\delta_t), \sigma(X(t), \Lambda(t))\mathrm{d}W(t)\rangle.
\end{aligned}$$

By Young's inequality, for any $x, y \ge 0$, $x^4y^2 \le \frac{2}{3}x^6 + \frac{1}{3}y^6$. Hence,

$$\begin{aligned}
&\mathrm{d}|X(t) - X(\delta_t)|^6 \\
&\le \big\{(6 + 36\alpha)|X(t) - X(\delta_t)|^6 + (36\alpha + 3c_1)|X(t) - X(\delta_t)|^4|X(\delta_t)|^2 \\
&\quad + 3(6\beta + c_0)|X(t) - X(\delta_t)|^4\big\}\mathrm{d}t + 6|X(t) - X(\delta_t)|^4\langle X(t) - X(\delta_t), \sigma(X(t), \Lambda(t))\mathrm{d}W(t)\rangle \\
&\le \big\{(6 + 60\alpha + 2c_1 + 12\beta + 2c_0)|X(t) - X(\delta_t)|^6 + (12\alpha + c_1)|X(\delta_t)|^6 + 6\beta + c_0\big\}\mathrm{d}t \\
&\quad + 6|X(t) - X(\delta_t)|^4\langle X(t) - X(\delta_t), \sigma(X(t), \Lambda(t))\mathrm{d}W(t)\rangle.
\end{aligned}$$

By taking expectation in both sides of the previous inequality, we get

$$\begin{aligned}
\mathbb{E}|X(t) - X(\delta_t)|^6 &\le \big((12\alpha + c_1)\mathbb{E}|X(\delta_t)|^6 + 6\beta + c_0\big)\delta \\
&\quad + \int_{\delta_t}^{t}\big\{(6 + 60\alpha + 2c_1 + 12\beta + 2c_0)\mathbb{E}|X(s) - X(\delta_s)|^6\big\}\mathrm{d}s.
\end{aligned}$$

Gronwall's inequality implies that

$$\begin{aligned}
\mathbb{E}|X(t) - X(\delta_t)|^6 &\le \Big((12\alpha + c_1)\delta\mathbb{E}|X(\delta_t)|^6 + (6\beta + c_0)\delta\Big)\mathrm{e}^{(6+60\alpha+2c_1+12\beta+2c_0)\delta} \\
&\le \Big(2^5(12\alpha + c_1)\delta\big(\mathbb{E}|X(t) - X(\delta_t)|^6 + \mathbb{E}|X(t)|^6\big) + (6\beta + c_0)\delta\Big)\mathrm{e}^{(6+60\alpha+2c_1+12\beta+2c_0)\delta}.
\end{aligned}$$

Therefore, if $K(\delta) < 1$, this inequality deduces the estimate (2.3). □

Theorem 2.4. *Assume* (H1), (H2) *and* (H5) *hold. Suppose that* $\eta > \alpha + 1$ *sufficiently large and* $\delta > 0$ *sufficiently small such that* $K(\delta) < 1$,

$$\theta(\eta,\delta) := 72\sqrt{2}\delta e^{3(2\eta-2\alpha-2)\delta}\kappa(2\eta-2\alpha-2)\big(\alpha^{\frac{3}{2}} + \frac{1}{2}\big) + 36\delta^3 \frac{K_u^6 K(\delta)}{1-K(\delta)} < 1, \tag{2.4}$$

and

$$\frac{9e^{-6(\eta-\alpha-1)\delta}}{1-\theta(\eta,\delta)} < 1, \tag{2.5}$$

where $\kappa(2\eta-2\alpha-2) = C_3 3^{-\frac{1}{4}}\Gamma\left(\frac{1}{6}\right)^{\frac{3}{2}}\Gamma\left(\frac{1}{8}\right)^{2}(2\eta-2\alpha-2)^{-\frac{1}{2}}$, $\Gamma(\cdot)$ *denotes the Gamma function, and* C_3 *is the universal constant in Burkholder-Davis-Gundy's inequality. Then*

$$\sup_{t\ge 0} \mathbb{E}\left[\|X_t\|^6\right] < \infty,$$

where we put $X(s) = X(0)$ *for* $s \le 0$.

Proof. Let $Z(t) = |X(t)|^2$, then by (H1), (H2) and Itô's formula,

$$\begin{aligned} dZ(t) &\le \Big(Z(t) + |b(X(t),\Lambda(t))|^2 + \|\sigma(X(t),\Lambda(t))\|_{HS}^2 - 2\langle X(t), u(X(\delta_t),\Lambda(\delta_t))\rangle\Big)dt \\ &\quad + 2\langle X(t), \sigma(X(t),\Lambda(t))dW(t)\rangle \\ &\le \Big(Z(t) + 2(\alpha Z(t)+\beta) - 2\langle X(t), u(X(t),\Lambda(\delta_t))\rangle \\ &\quad + 2\langle X(t), u(X(t),\Lambda(\delta_t)) - u(X(\delta_t),\Lambda(\delta_t))\rangle\Big)dt \\ &\quad + 2\langle X(t), \sigma(X(t),\Lambda(t))dW(t)\rangle \\ &\le \Big(2\beta + (2\alpha+1)Z(t) - 2\eta Z(t) + Z(t) + K_u^2|X(t)-X(\delta_t)|^2\Big)dt \\ &\quad + 2\langle X(t), \sigma(X(t),\Lambda(t))dW(t)\rangle \\ &= \Big(2\beta + (2\alpha+2-2\eta)Z(t) + K_u^2|X(t)-X(\delta_t)|^2\Big)dt \\ &\quad + 2\langle X(t), \sigma(X(t),\Lambda(t))dW(t)\rangle. \end{aligned}$$

For simplicity of notation, let

$$\lambda = 2\eta - 2\alpha - 2,$$

which is positive by assumption. So,

$$d\big(e^{\lambda t}Z(t)\big) \le e^{\lambda t}\big(2\beta + K_u^2|X(t)-X(\delta_t)|^2\big)dt + 2e^{\lambda t}\langle X(t), \sigma(X(t),\Lambda(t))dW(t)\rangle.$$

This implies that for $0 \le s < t$,

$$\begin{aligned} e^{\lambda t}Z(t) &\le e^{\lambda s}Z(s) + \int_s^t e^{\lambda r}\big(2\beta + K_u^2|X(r)-X(\delta_r)|^2\big)dr \\ &\quad + \int_s^t 2e^{\lambda r}\langle X(r), \sigma(X(r),\Lambda(r))dW(r)\rangle. \end{aligned} \tag{2.6}$$

There is a 1-dimension Brownian motion $(B(t))_{t\geq 0}$ w.r.t. the same filtration such that

$$\langle X(s), \sigma(X(s), \Lambda(s))\mathrm{d}W(s)\rangle = \chi(s,\omega)\mathrm{d}B(s), \tag{2.7}$$

where

$$\chi(s,\omega) = \Big(\sum_{j=1}^{d}\Big(\sum_{i=1}^{d} X_i(s)\sigma_{ij}(X(s),\Lambda(s))\Big)^2\Big)^{\frac{1}{2}}.$$

According to [9, Lemma 2.2],

$$\mathbb{E}\sup_{0\leq t\leq T}\Big|\int_0^t \mathrm{e}^{-\lambda(t-s)}\chi(s,\omega)\mathrm{d}B(s)\Big|^3 \leq \kappa(\lambda)\mathbb{E}\Big[\int_0^T |\chi(s,\omega)|^3\mathrm{d}s\Big], \tag{2.8}$$

where

$$\kappa(\lambda) = C_3 3^{-\frac{1}{4}}\Gamma\Big(\frac{1}{6}\Big)^{\frac{3}{2}}\Gamma\Big(\frac{1}{8}\Big)^2\lambda^{-\frac{1}{2}}, \tag{2.9}$$

and C_3 is the universal constant in Burkholder-Davis-Gundy's inequality, i.e. for a martingale $(M_t)_{t\geq 0}$,

$$\mathbb{E}\big[\sup_{s\leq t}|M_s|^3\big] \leq C_3\mathbb{E}\big[\langle M\rangle_t^{3/2}\big].$$

Indeed, $\kappa(\lambda)$ is derived from the argument of [9, Lemma 2.2] by taking there $p = 3$, $\alpha = \frac{5}{12}, \mu = \lambda$.

Due to (H1),

$$\chi(s,\omega)^2 \leq |X(s)|^2\|\sigma(X(s),\Lambda(s))\|_{\mathrm{HS}}^2 \leq |X(s)|^2\big(\alpha|X(s)|^2 + \beta\big). \tag{2.10}$$

As $(a+b+c)^3 \leq 9(a^3+b^3+c^3)$ for $a,b,c\geq 0$, (2.3), (2.6), (2.8) and (2.10) yield that for $t\neq k\delta$ for some $k\in\mathbb{Z}_+$,

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{\delta_t\leq s\leq t}\Big(\mathrm{e}^{\lambda s}Z(s)\Big)^3\Big]\\
&\leq 9\mathrm{e}^{3\lambda\delta_t}\mathbb{E}Z(\delta_t)^3 + 9\delta^2\int_{\delta_t}^t \mathrm{e}^{3\lambda r}\mathbb{E}\Big[\big(2\beta + K_u^2|X(r)-X(\delta_t)|^2\big)^3\Big]\mathrm{d}r\\
&\quad + 72\mathrm{e}^{3\lambda(\delta_t+\delta)}\kappa(\lambda)\mathbb{E}\Big[\int_{\delta_t}^t |\vartheta(r,\omega)|^3\mathrm{d}r\Big]\\
&\leq 9\mathrm{e}^{3\lambda\delta_t}\mathbb{E}Z(\delta_t)^3 + 9\delta^2\int_{\delta_t}^t \mathrm{e}^{3\lambda r}\Big(32\beta^3 + 4K_u^6\mathbb{E}\big[|X(r)-X(\delta_t)|^6\big]\Big)\mathrm{d}r\\
&\quad + 72\mathrm{e}^{3\lambda(\delta_t+\delta)}\kappa(\lambda)\sqrt{2}\int_{\delta_t}^t \mathbb{E}\Big[\alpha^{\frac{3}{2}}|X(r)|^6 + \frac{1}{2}\beta^3 + \frac{1}{2}|X(r)|^6\Big]\mathrm{d}r\\
&\leq 9\mathrm{e}^{3\lambda\delta_t}\mathbb{E}Z(\delta_t)^3 + 96\delta^2\beta^3\frac{1}{\lambda}\big(\mathrm{e}^{3\lambda t}-\mathrm{e}^{3\lambda\delta_t}\big) + 36\sqrt{2}\mathrm{e}^{3\lambda(\delta_t+\delta)}\kappa(\lambda)\beta^3\delta
\end{aligned}$$

$$+ 36\delta^3 \frac{K_u^6 K(\delta)}{1-K(\delta)} \mathbb{E}\Big[\sup_{\delta_t \le s \le t} \mathrm{e}^{3\lambda s} Z(s)^3\Big] + \frac{72\delta^3 K_u^6 \beta}{1-K(\delta)} \frac{\mathrm{e}^{3\lambda t} - \mathrm{e}^{3\lambda \delta_t}}{\lambda} \mathrm{e}^{(6+60\alpha+\hat{\eta}+12\beta)\delta}$$
$$+ 72\sqrt{2}\delta \mathrm{e}^{3\lambda\delta} \kappa(\lambda)\big(\alpha^{\frac{3}{2}} + \frac{1}{2}\big)\mathbb{E}\Big[\sup_{\delta_t \le s \le t} \mathrm{e}^{3\lambda s} Z(s)^3\Big].$$

Let

$$\theta_1(\lambda, \delta) = 72\sqrt{2}\delta \mathrm{e}^{3\lambda\delta}\kappa(\lambda)\big(\alpha^{\frac{3}{2}} + \frac{1}{2}\big) + 36\delta^3 K_u^6 \frac{K(\delta)}{1-K(\delta)}, \tag{2.11}$$

$$\theta_2(\lambda, \delta) = 96\delta^2\beta^3 \frac{\mathrm{e}^{3\lambda\delta} - 1}{\lambda} + 36\sqrt{2}\mathrm{e}^{3\lambda\delta}\kappa(\lambda)\beta^3\delta + \frac{72\delta^3 K_u^6 \beta}{1-K(\delta)} \frac{\mathrm{e}^{3\lambda\delta} - 1}{\lambda} \mathrm{e}^{(6+60\alpha+\hat{\eta}+12\beta)\delta}. \tag{2.12}$$

If $\theta_1(\lambda, \delta) < 1$, then for $k \ge 0$,

$$\mathbb{E}\Big[\sup_{k\delta \le s < (k+1)\delta} \mathrm{e}^{3\lambda s} Z(s)^3\Big] \le \frac{9}{1-\theta_1(\lambda,\delta)} \mathbb{E}\Big[\mathrm{e}^{3\lambda k\delta} Z(k\delta)^3\Big] + \frac{\theta_2(\lambda,\delta)\mathrm{e}^{3\lambda k\delta}}{1-\theta_1(\lambda,\delta)}. \tag{2.13}$$

Define a function $V : \mathscr{C}([-\delta, 0]; \mathbb{R}^d) \to \mathbb{R}$ by

$$V(\zeta) = \sup_{-\delta \le s \le 0} \mathrm{e}^{3\lambda s}|\zeta(s)|^6, \quad \zeta \in \mathscr{C}([-\delta, 0]; \mathbb{R}^d).$$

It is easy to see

$$V(\zeta) = \mathrm{e}^{-3\lambda\delta} \sup_{0 \le s \le \delta} \mathrm{e}^{3\lambda s}|\zeta(s-\delta)|^6 \ge |\zeta(0)|^6.$$

Applying (2.13) to $k = 0$, we obtain that

$$\mathbb{E}\big[V(X_\delta)\big] \le \frac{9\mathrm{e}^{-3\lambda\delta}}{1-\theta_1(\lambda,\delta)} \mathbb{E}\Big[|X(0)|^6\Big] + \frac{\theta_2(\lambda,\delta)\mathrm{e}^{-3\lambda\delta}}{1-\theta_1(\lambda,\delta)}.$$

Applying (2.13) to $k \ge 1$,

$$\begin{aligned}\mathbb{E}\big[V(X_{(k+1)\delta})\big] &= \mathrm{e}^{-3\lambda(k+1)\delta}\mathbb{E}\Big[\sup_{k\delta \le s \le (k+1)\delta} \mathrm{e}^{3\lambda s}|X(s)|^6\Big] \\ &\le \frac{9\mathrm{e}^{-3\lambda\delta}}{1-\theta_1(\lambda,\delta)} \mathbb{E}\big[|X(k\delta)|^6\big] + \frac{\theta_2(\lambda,\delta)\mathrm{e}^{-3\lambda\delta}}{1-\theta_1(\lambda,\delta)} \\ &\le \frac{9\mathrm{e}^{-3\lambda\delta}}{1-\theta_1(\lambda,\delta)} \mathbb{E}\big[V(X_{k\delta})\big] + \frac{\theta_2(\lambda,\delta)\mathrm{e}^{-3\lambda\delta}}{1-\theta_1(\lambda,\delta)}.\end{aligned}$$

Therefore, if

$$\theta_3(\lambda, \delta) := \frac{9\mathrm{e}^{-3\lambda\delta}}{1-\theta_1(\lambda,\delta)} < 1, \tag{2.14}$$

then iterating above estimate, we have

$$\mathbb{E}\big[V(X_{(k+1)\delta})\big] \le \theta_3(\lambda,\delta)\mathbb{E}\big[|X(0)|^6\big] + \frac{1}{1-\theta_3(\lambda,\delta)} \cdot \frac{\theta_2(\lambda,\delta)\mathrm{e}^{-3\lambda\delta}}{1-\theta_1(\lambda,\delta)}, \quad \forall k \ge 1.$$

Combining these estimates with the fact $\|X_t\|^6 \le \|X_{k\delta}\|^6 + \|X_{(k+1)\delta}\|^6$ for $t \in (k\delta, (k+1)\delta)$, we finally get

$$\sup_{t\ge 0} \mathbb{E}\left[\|X_t\|^6\right] \le 2\theta_3(\lambda,\delta)\mathrm{e}^{3\lambda\delta}\mathbb{E}\left[|X(0)|^6\right] + \frac{2}{1-\theta_3(\lambda,\delta)} \cdot \frac{\theta_2(\lambda,\delta)}{1-\theta_1(\lambda,\delta)} < \infty,$$

and complete the proof. □

Remark 2.5. Note that

$$\lim_{\delta\downarrow 0} K(\delta) = 0, \quad \lim_{\delta\downarrow 0}\lim_{\eta\to\infty} \theta(\eta,\delta) = 0,$$

then by choosing η, δ such that $\lim\limits_{\delta\to 0,\eta\to\infty} \mathrm{e}^{-3\eta\delta} = 0$, the conditions in Theorem 2.4 could be verified for suitable δ and η.

Proposition 2.6. *Under the conditions of Theorem 2.4, the class of distributions $\{\mathscr{L}_{(X_t,\Lambda_t)}; t \ge 0\}$ is tight.*

Proof. Let us first show that $\{\mathscr{L}_{\Lambda_t}; t \ge 0\}$ is tight. According to Kurtz's tightness criterion ([10, Theorem 8.6, p. 137]), we only need to show there exists a sequence of nonnegative random variables $\gamma_t(\varepsilon)$ such that $\lim_{\varepsilon\to 0}\sup_{t\ge 0}\mathbb{E}[\gamma_t(\varepsilon)] = 0$, and

$$\mathbb{E}\left[\mathbf{1}_{\Lambda_t(s+u)\ne\Lambda_t(s)}\middle|\mathscr{F}_s^{\Lambda_t}\right] \le \mathbb{E}\left[\gamma_t(\varepsilon)|\mathscr{F}_s^{\Lambda_t}\right] \tag{2.15}$$

for $0 \le u \le \varepsilon$, $-\delta \le s \le 0$, where $\mathscr{F}_s^{\Lambda_t} = \sigma\{\Lambda_t(r); -\delta \le r \le s\}$. Due to (H5),

$$\mathbb{P}\big(\Lambda_t(s+r) = \Lambda_t(s), \forall\, r \in [0,u]\big) \ge \mathbb{E}\big[\exp\big(-\sup_{j\in\mathcal{S}} q_j u\big)\big] \ge \exp(-Mu).$$

Then, we take $\gamma_t(\varepsilon) = 1 - \mathrm{e}^{-M\varepsilon}$ to arrive at

$$\begin{aligned}\mathbb{E}\left[\mathbf{1}_{\Lambda_t(s+u)\ne\Lambda_t(s)}\middle|\mathscr{F}_s^{\Lambda_t}\right] &\le 1 - \mathbb{P}\big(\Lambda_t(s+r) = \Lambda_t(s), \forall\, r \in [0,u]\big)\\ &\le 1 - \mathrm{e}^{-M\varepsilon} = \gamma_t(\varepsilon), \quad \forall\, 0 \le u \le \varepsilon.\end{aligned}$$

It is clear that (2.15) is verified, and we conclude that $\{\mathscr{L}_{\Lambda_t}; t \ge 0\}$ is tight.

By Itô's formula and (H1), (H2), for any $0 \le s_1 < s_2 \le \delta$,

$$\begin{aligned}&\mathbb{E}|X_t(-s_1) - X_t(-s_2)|^4 = \mathbb{E}|X((t-s_1)\vee 0) - X((t-s_2)\vee 0)|^4\\ &\le \mathbb{E}\Big[\Big|\int_{(t-s_2)\vee 0}^{(t-s_1)\vee 0} b(X(s),\Lambda(s)) - u(X(\delta_s),\Lambda(\delta_s))\mathrm{d}s + \int_{(t-s_2)\vee 0}^{(t-s_1)\vee 0} \sigma(X(s),\Lambda(s))\mathrm{d}W(s)\Big|^4\Big]\\ &\le 8(s_2-s_1)^3\mathbb{E}\Big[\int_{(t-s_2)\vee 0}^{(t-s_1)\vee 0} \big(2\alpha|X(s)|^2 + 2\beta + 2c_0 + 2c_1|X(\delta_s)|^2\big)^2\mathrm{d}s\Big]\\ &\quad + 288(s_2-s_1)\mathbb{E}\Big[\int_{(t-s_2)\vee 0}^{(t-s_1)\vee 0} \big(\alpha|X(s)|^2+\beta\big)^2\mathrm{d}s\Big].\end{aligned}$$

Hence, according to Theorem 2.4, $\sup_{t\ge 0}\mathbb{E}|X(t)|^6 < \infty$, then

$$\sup_{t\ge 0}\mathbb{E}|X_t(-s_1) - X_t(-s_2)|^4 \le C(s_2-s_1)^2$$

for some constant $C > 0$. By virtue of [4, Theorem 12.3], $\{\mathscr{L}_{X_t}; t \geq 0\}$ is tight. In all, it is easy to see that $\{\mathscr{L}_{(X_t,\Lambda_t)}; t \geq 0\}$ is tight. □

Next, we study the continuous dependence on initial values of $(X(t), \Lambda(t))$. As $\mathcal{S}$ is endowed with discrete topology, we only need to focus on the component $(X(t))$. Let $(X^{(m)}(t), \Lambda(t))_{t\geq 0}$ and $(X(t), \Lambda(t))_{t\geq 0}$ be the solution to the following SDEs respectively

$$\begin{aligned} X^{(m)}(t) = x_m + \int_0^t b(X^{(m)}(s), \Lambda(s)) - u(X^{(m)}(\delta_s), \Lambda(\delta_s))\mathrm{d}s \\ + \int_0^t \sigma(X^{(m)}(s), \Lambda(s))\mathrm{d}W(s), \end{aligned} \tag{2.16}$$

$$\begin{aligned} X(t) = x_0 + \int_0^t b(X(s), \Lambda(s)) - u(X(\delta_s), \Lambda(\delta_s))\mathrm{d}s \\ + \int_0^t \sigma(X(s), \Lambda(s))\mathrm{d}W(s), \end{aligned} \tag{2.17}$$

where $i \in \mathcal{S}$, $x_m, x_0 \in \mathbb{R}^d$. Let $X^{(m)}(s) = x_m$, $X(s) = x_0$ for $s \leq 0$.

Lemma 2.7. *Assume* (H1)-(H3) *hold. Then for any bounded continuous function* $F : \mathscr{C}([-\delta, 0]; \mathbb{R}^d) \to \mathbb{R}$ *it holds*

$$\lim_{m\to\infty} \mathbb{E}F(X_t^{(m)}) = \mathbb{E}F(X_t), \quad t > 0,$$

if $\lim_{m\to\infty} x_m = x_0$.

Proof. We derive from Itô's formula that

$$\begin{aligned} \mathrm{d}|X^{(m)}(t) - X(t)|^2 = &\big\{2\langle b(X^{(m)}(t), \Lambda(t)) - b(X(t), \Lambda(t)), X^{(m)}(t) - X(t)\rangle \\ &+ 2\langle u(X(\delta_t), \Lambda(\delta_t)) - u(X^{(m)}(\delta_t), \Lambda(\delta_t)), X^{(m)}(t) - X(t)\rangle\big\}\mathrm{d}t \\ &+ \|\sigma(X^{(m)}(t), \Lambda(t)) - \sigma(X(t), \Lambda(t))\|_{\mathrm{HS}}^2 \mathrm{d}t + \mathrm{d}M_t, \end{aligned}$$

where

$$M_t = \int_0^t 2\langle X^{(m)}(s) - X(s), \big(\sigma(X^{(m)}(s), \Lambda(s)) - \sigma(X(s), \Lambda(s))\big)\mathrm{d}W(s)\rangle.$$

Hence, by (H2) and (H3),

$$\|X_t^{(m)} - X_t\|^2 \leq 2|x_m - x_0|^2 + \int_0^t (2\tilde{K}^2 + K_u^2 + 2)\|X_s^{(m)} - X_s\|^2 \mathrm{d}s + \sup_{0\leq s\leq t} M_s.$$

Furthermore,

$$\mathbb{E}\|X_t^{(m)} - X_t\|^4 \le 3\Big(4|x_m - x_0|^4 + (2\tilde{K}^2 + K_u^2 + 2)^2 t \int_0^t \mathbb{E}\big[\|X_s^{(m)} - X_s\|^4\big]\mathrm{d}s$$

$$+ 16\tilde{K}^2 \int_0^t \mathbb{E}\big[|X^{(m)}(s) - X(s)|^4\big]\mathrm{d}s\Big).$$

It follows from Gronwall's inequality that

$$\mathbb{E}\|X_t^{(m)} - X_t\|^4 \le 12|x_m - x_0|^4 \exp\big(3(2\tilde{K}^2 + K_u^2 + 2)^2 t + 48\tilde{K}^2\big).$$

So, $X_t^{(m)}$ converges to X_t in probability as $m \to \infty$, and the dominated convergence theorem yields that

$$\lim_{m\to\infty} \mathbb{E}F(X_t^{(m)}) = \mathbb{E}F(X_t), \quad t > 0,$$

which yields the Feller property of (X_t) in $\mathscr{C}([-\delta,0];\mathbb{R}^d)$. □

Theorem 2.8. *Suppose the conditions of Theorem 2.4 hold. In addition, assume* (H3) *and* (H4) *hold. Then the process $(X_t, \Lambda_t)_{t\ge 0}$ admits a unique invariant probability measure μ, and the distribution of (X_t, Λ_t) in $\mathscr{C}([-\delta,0];\mathbb{R}^d) \times \mathscr{D}([-\delta,0];\mathcal{S})$ converges weakly to μ. Hence, $(X(t), \Lambda(t))$ is strongly stable in distribution.*

Proof. According to the Krylov-Bogoliubov Theorem (cf. [8, Section 3.1]), the existence of invariant probability measure μ follows from Proposition 2.6 and Lemma 2.7.

To see the uniqueness of μ, we need to check the semigroup P_t corresponding to (X_t, Λ_t) is t_0-regular for some $t_0 > 0$. By definition (cf. [8, Section 4.1]), if all the transition probabilities $P_{t_0}((\phi,i),\cdot)$, $(\phi,i) \in \mathscr{C}([-\delta,0];\mathbb{R}^d) \times \mathcal{S}$, are mutually equivalent, then P_t is t_0-regular. For SDE (1.3), the initial value for $(X(t))$ satisfies that $X_0(s) \equiv X(0) = x$ for all $s \in [-\delta,0]$. So we can write $P_{t_0}((\phi,i),\cdot) = P_{t_0}((\phi(0),i),\cdot)$. Since (q_{ij}) is irreducible, the transition probability $P_{ij}(t) := \mathbb{P}(\Lambda(t) = j|\Lambda(0) = i)$ of the Markov chain $(\Lambda(t))$ satisfies that $P_{ij}(t) > 0$ for all $i,j \in \mathcal{S}$ and all $t > 0$. Therefore, the initial state i of $(\Lambda(t))$ has no impact on the equivalence of $P_{t_0}((\phi(0),i),\cdot)$.

To emphasize the initial state of solution to SDE (1.3), we denote $X^{x,i}(t)$ the solution to SDE (1.3) with $X(0) = x, \Lambda(0) = i$. Due to (H1)-(H3), particularly the nondegenerate condition (H4), for all $x \in \mathbb{R}^d$, $i \in \mathcal{S}$, $t > 0$, the distributions of $X^{x,i}(t)$ are all equivalent to the Lebesgue measure. Therefore, given any $t_0 > 0$, for all $x \in \mathbb{R}^d$, $i \in \mathcal{S}$, the finite dimensional projection of the distributions of $(X_{t_0}^{x,i}, \Lambda_{t_0})$ on $\mathscr{C}([-\delta,0];\mathbb{R}^d) \times \mathscr{D}([-\delta,0];\mathcal{S})$ are all equivalent, which yields that $P_{t_0}((x,i),\cdot)$, $(x,i) \in \mathbb{R}^d \times \mathcal{S}$, are mutually equivalent. Then, applying [8, Theorem 4.2.1], μ is the unique invariant probability measure of $P_t((x,i),\cdot)$ on $\mathscr{C}([-\delta,0];\mathbb{R}^d) \times \mathscr{D}([-\delta,0];\mathcal{S})$ and $P_t((x,i),\cdot)$ converges weakly to μ. By Definition 2.2, $(X(t), \Lambda(t))$ is strongly stable in distribution. □

Remark 2.9. Theorem 2.4 cannot be proved by a direct application of [9, Theorem 3.2], since we cannot verify simultaneously the conditions $(\mathbf{H_0})$ and $(\mathbf{H_2})$ there for the SDE (1.3). The control of $\mathbb{E}|X(t) - X(\delta_t)|^6$ via $\mathbb{E}|X(t)|^6$ plays essential role in our argument.

3. An approach via embedding Markov processes in finite dimensional spaces

In this section instead of viewing $(X(t))_{t\ge 0}$ via the segment process $(X_t)_{t\ge 0}$ in the path space $\mathscr{C}([-\delta,0];\mathbb{R}^d)$, we shall apply the special history dependent structure of the process $(X(t))_{t\ge 0}$ to construct a family of finite dimensional embedded Markov process and to prove its existence of invariant probability measure. Then, via Kolmogorov's extension theorem, we can find a probability measure μ on the path space

$\mathscr{C}([-\delta,0];\mathbb{R}^d)$ such that any finite dimensional projection of the distribution of the segment process X_t will converge to the corresponding finite dimensional projection of μ. Notice that this does not mean that the distribution of X_t in $\mathscr{C}([-\delta,0];\mathbb{R}^d)$ converges weakly to μ; see [4, Chapter 3] for more discussions. Compared with Theorem 2.4, in this situation the condition on the term $u(x,i)$ will be weakened. Our approach in this section is based on the following observation:

- For each $k\geq 1$ and any $0<r_1<\ldots<r_k<\delta$, the process $\big(X(n\delta+r_k),\ldots,X(n\delta+r_1),X(n\delta),\Lambda(n\delta)\big)_{n\geq 0}$ is a Markov process on $\mathbb{R}^{kd}\times\mathcal{S}$.

In order to present a strict argument, we recall Skorokhod's representation for continuous time Markov chains, which has been extensively studied in the study of regime-switching processes; see, e.g. [25,20] and the recent work [21] on its application to study the ergodicity of state-dependent regime-switching processes.

Precisely, define a sequence of intervals Δ_{ij} on $[0,\infty)$ associated with the transition rate matrix $(q_{ij})_{i,j\in\mathcal{S}}$ of $(\Lambda_t)_{t\geq 0}$ in the following way:

$$\Delta_{12}=[0,q_{12}),\quad \Delta_{13}=[q_{12},q_{12}+q_{13}),\ldots,\Delta_{1N}=\Big[\sum_{1<j<N}q_{1j},q_1\Big),$$
$$\Delta_{21}=[q_1,q_1+q_{21}),\quad \Delta_{23}=[q_1+q_{21},q_1+q_{21}+q_{23}),\ldots,$$

and so on. For convenience of notation, put $\Delta_{ii}=\emptyset$. Then, introduce an auxiliary function $\vartheta:\mathcal{S}\times[0,\infty)\to\mathbb{R}$ defined by

$$\vartheta(i,z)=\sum_{j\in\mathcal{S},j\neq i}(j-i)\mathbf{1}_{\Delta_{ij}}(z).$$

Let $\mathcal{N}(\mathrm{d}t,\mathrm{d}z)$ be a Poisson random measure with intensity $\mathrm{d}t\times\mathrm{d}z$, which is independent of the Wiener process $(W(t))_{t\geq 0}$. Consequently, $(\Lambda(t))_{t\geq 0}$ can be expressed as a solution to the following SDE:

$$\Lambda(t)=\Lambda(0)+\int_0^t\int_{[0,\infty)}\vartheta(\Lambda(s),z)\mathcal{N}(\mathrm{d}s,\mathrm{d}z). \tag{3.1}$$

Lemma 3.1. *Assume* (H1)-(H3) *and* (H5) *hold. Let* $(X(t),\Lambda(t))$ *be the solution to* (1.3)*. Then*

(i) $(X(n\delta),\Lambda(n\delta))_{n\geq 0}$ *is a discrete-time Markov process on* $\mathbb{R}^d\times\mathcal{S}$.
(ii) *For each* $k\geq 1$ *and any* $0<r_1<\ldots<r_k<\delta$*, the process* $\big(X(n\delta+r_k),\ldots,X(n\delta+r_1),X(n\delta),\Lambda(n\delta)\big)_{n\geq 0}$ *is a Markov process on* $\mathbb{R}^{(k+1)d}\times\mathcal{S}$.

Proof. (i) Based on Skorokhod's representation (3.1) for $(\Lambda(t))_{t\geq 0}$, we can rewrite $(X(t),\Lambda(t))_{t\geq 0}$ as a solution to the SDE below:

$$\begin{aligned}\mathrm{d}\begin{pmatrix}X(t)\\ \Lambda(t)\end{pmatrix}&=\begin{pmatrix}b(X(t),\Lambda(t))-u(X(\delta_t),\Lambda(\delta_t))\\ 0\end{pmatrix}\mathrm{d}t+\begin{pmatrix}\sigma(X(t),\Lambda(t))\\ 0\end{pmatrix}\mathrm{d}W(t)\\ &\quad+\int_{[0,\infty)}\begin{pmatrix}0\\ \vartheta(\Lambda(s),z)\end{pmatrix}\mathcal{N}(\mathrm{d}s,\mathrm{d}z).\end{aligned} \tag{3.2}$$

Under conditions (H1)-(H3), SDE (3.2) admits a unique strong solution (cf. [20, Theorem 2.4]). This means

$$\big(X((n+1)\delta),\Lambda((n+1)\delta)\big)=F\big(X(n\delta),\Lambda(n\delta),(W(t),\mathcal{N}(t))_{t\in[n\delta,(n+1)\delta)}\big)$$

for some measurable functional F. Refer to the argument of [19, Theorem 7.1.2] on the Markov property of solution to SDE for more details. So, $\big(X((n+1)\delta), \Lambda((n+1)\delta)\big)$ is independent of its history $(X(s), \Lambda(s))_{s\in[0,n\delta)}$. Consequently, $(X(n\delta), \Lambda(n\delta))_{n\geq 0}$ becomes a Markov process.

We proceed to give out the transition probability measure of the process $(X(n\delta), \Lambda(n\delta))_{n\geq 0}$. For $x \in \mathbb{R}^d$, $i \in \mathcal{S}$, consider the SDE

$$\tilde{X}(t) = x - u(x,i)t + \int_0^t b(\tilde{X}(s), \tilde{\Lambda}(s))\mathrm{d}s + \int_0^t \sigma(\tilde{X}(s), \tilde{\Lambda}(s))\mathrm{d}W(s),$$

where $(\tilde{\Lambda}(t))$ is a Markov chain on $\mathcal{S}$ with transition rate matrix (q_{ij}) and being independent of the Wiener process $(W(t))$. Under conditions (H1)-(H3), the previous SDE admits a unique strong solution, and define

$$P_t((x,i); A\times\{j\}) = \mathbb{P}\big(\tilde{X}(t) \in A,\ \tilde{\Lambda}(t) = j \big| \tilde{X}(0) = x, \tilde{\Lambda}(0) = i\big)$$

for $A \in \mathscr{B}(\mathbb{R}^d)$, $j \in \mathcal{S}$, $t > 0$. Then the transition probability measure of the Markov process $(X(n\delta), \Lambda(n\delta))_{n\geq 0}$ is given by

$$\mathbb{P}(X((n+1)\delta) \in A, \Lambda((n+1)\delta) = j | X(n\delta) = x, \Lambda(n\delta) = i) = P_\delta((x,i); A\times\{j\}). \tag{3.3}$$

Moreover, under the uniform ellipticity condition (H4), the density of $P_t((x,i), \cdot\times\{j\})$ w.r.t. the Lebesgue measure exists, denoted by $p_t\big((x,i);(z,j)\big)$, and so

$$P_t((x,i); A\times\{j\}) = \int_A p_t\big((x,i);(z,j)\big)\mathrm{d}z. \tag{3.4}$$

Combining (H4) with the irreducibility of (q_{ij}), it also holds

$$p_t((x,i);(z,j)) > 0, \quad x, z \in \mathbb{R}^d,\ i,j \in \mathcal{S}.$$

(ii) For simplicity of notation, we only consider the case $k = 1$ with $r_1 = r \in (0, \delta)$. Combining (3.2) with the following observation

$$\begin{aligned} &X((n+1)\delta + r) - X((n+1)\delta) \\ &= \int_{(n+1)\delta}^{(n+1)\delta+r} \Big(b(X(s), \Lambda(s)) - u(X((n+1)\delta), \Lambda((n+1)\delta))\Big)\mathrm{d}s \\ &\quad + \int_{(n+1)\delta}^{(n+1)\delta+r} \sigma(X(s), \Lambda(s))\mathrm{d}W(s), \end{aligned}$$

we obtain that $(X((n+1)\delta + r, X((n+1)\delta), \Lambda((n+1)\delta)$ depends only on $(X(n\delta + r), X(n\delta), \Lambda(n\delta))$ and $(W(t))_{t\in[n\delta,(n+1)\delta+r)}$, $(\mathcal{N}(t))_{t\in[n\delta,(n+1)\delta+r)}$. Hence, $(X(n\delta+r), X(n\delta), \Lambda(n\delta))_{n\geq 0}$ becomes a Markov process.

To characterize the transition probability of $(X(n\delta+r), X(n\delta), \Lambda(n\delta))_{n\geq 0}$, we introduce another auxiliary SDE: for given $y \in \mathbb{R}^d$, $i \in \mathcal{S}$,

$$\bar{X}(t) = x - u(y,i)(t-r) + \int_r^t b(\bar{X}(s), \bar{\Lambda}(s))\mathrm{d}s + \int_r^t \sigma(\bar{X}(s), \bar{\Lambda}(s))\mathrm{d}W(s),\ t \geq r,$$

where $(\bar{\Lambda}(t))_{t\geq 0}$ satisfying $\bar{\Lambda}(r)=k\in\mathcal{S}$ is a Markov chain with transition rate matrix (q_{ij}) and independent of $(W(t))$. Also, the previous SDE admits a unique strong solution under conditions (H1)-(H3). Define

$$\bar{P}^{(y,i)}_{t-r}((x,k);A\times\{j\})=\mathbb{P}\big(\bar{X}(t)\in A,\bar{\Lambda}(t)=j\big|\bar{X}(r)=x,\bar{\Lambda}(r)=k\big) \tag{3.5}$$

for $A\in\mathscr{B}(\mathbb{R}^d)$, $j\in\mathcal{S}, t>r$. After these preparations the transition probability of $(X(n\delta+r),X(n\delta),\Lambda(n\delta))_{n\geq 0}$ is given by

$$\begin{aligned}
&\mathbb{P}\big(X((n+1)\delta+r)\in A_1,X((n+1)\delta)\in A_2,\Lambda((n+1)\delta)=j\big|X(n\delta+r)=x,X(n\delta)=y,\Lambda(n\delta)=i\big)\\
&=\mathbb{P}\big(X((n+1)\delta+r)\in A_1\big|X((n+1)\delta)\in A_2,\Lambda((n+1)\delta)=j\big)\\
&\quad\cdot\sum_{k\in\mathcal{S}}\mathbb{P}\big(X((n+1)\delta)\in A_2,\Lambda((n+1)\delta)=j\big|X(n\delta+r)=x,\Lambda(n\delta+r)=k,X(n\delta)=y,\Lambda(n\delta)=i\big)\\
&\quad\cdot\mathbb{P}(X(n\delta+r)\in \mathrm{d}x,\Lambda(n\delta+r)=k\big|X(n\delta)=y,\Lambda(n\delta)=i)\\
&\quad\cdot\frac{1}{\mathbb{P}\big(X(n\delta+r)\in\mathrm{d}x\big|X(n\delta)=y,\Lambda(n\delta)=i\big)}\\
&=\frac{1}{\sum_{k\in\mathcal{S}}p_r\big((y,i);(x,k)\big)}\sum_{k\in\mathcal{S}}\int_{A_2}P_r((z,j);A_1\times\mathcal{S})\bar{P}^{(y,i)}_{\delta-r}\big((x,k);\mathrm{d}z\times\{j\}\big)p_r\big((y,i);(x,k)\big)
\end{aligned} \tag{3.6}$$

for $A_1,\ A_2\in\mathscr{B}(\mathbb{R}^d)$, $i,j\in\mathcal{S}$. □

Notice that the previous observations (i) and (ii) on the Markov property of embedded processes of $(X(t),\Lambda(t))_{t\geq 0}$ do not hold in general for other type of stochastic functional differential equations. For instance, it does not hold for the following stochastic delay differential equation:

$$\mathrm{d}Y(t)=b(Y(t),Y(t-1))\mathrm{d}t+\mathrm{d}W(t).$$

To this equation, one can only view the segment process $\{Y_t;t\geq 0\}$, defined by $Y_t(u)=Y(t+u)$ for $u\in[-1,0]$, as a Markov process in the infinite dimensional space $\mathscr{C}([-1,0];\mathbb{R}^d)$ (cf. [18, Theorem 1.1 of Chapter III]).

Lemma 3.2. *Under the conditions* (H1)-(H3) *and* (H5), *if*

$$\tilde{K}(\delta):=2(4\alpha+c_1)\delta\mathrm{e}^{(2+4\alpha)\delta}<1, \tag{3.7}$$

then

$$\mathbb{E}\big[|X(t)-X(\delta_t)|^2\big]\leq\frac{\tilde{K}(\delta)}{1-\tilde{K}(\delta)}\mathbb{E}|X(t)|^2+\frac{(2\beta+c_0)\delta\mathrm{e}^{(2+4\alpha)\delta}}{1-\tilde{K}(\delta)},\quad t\geq 0. \tag{3.8}$$

Proof. By Itô's formula, due to (H1), (H2),

$$\begin{aligned}
&\mathrm{d}|X(t)-X(\delta_t)|^2\\
&=\big[2\langle X(t)-X(\delta_t),b(X(t),\Lambda(t))-u(X(\delta_t),\Lambda(\delta_t))\rangle+\|\sigma(X(t),\Lambda(t))\|^2_{\mathrm{HS}}\big]\mathrm{d}t\\
&\quad+2\langle X(t)-X(\delta_t),\sigma(X(t),\Lambda(t))\mathrm{d}W(t)\rangle\\
&\leq\big[2|X(t)-X(\delta_t)|^2+|b(X(t),\Lambda(t))|^2+|u(X(\delta_t),\Lambda(\delta_t))|^2+\|\sigma(X(t),\Lambda(t))\|^2_{\mathrm{HS}}\big]\mathrm{d}t\\
&\quad+2\langle X(t)-X(\delta_t),\sigma(X(t),\Lambda(t))\mathrm{d}W(t)\rangle
\end{aligned}$$

$$
\begin{aligned}
&\le \big[(2+4\alpha)|X(t)-X(\delta_t)|^2+(4\alpha+c_1)|X(\delta_t)|^2+2\beta+c_0\big]\mathrm{d}t\\
&\quad + 2\langle X(t)-X(\delta_t),\sigma(X(t),\Lambda(t))\mathrm{d}W(t)\rangle.
\end{aligned}
$$

Then we obtain from Gronwall's inequality that

$$
\begin{aligned}
&\mathbb{E}\big[|X(t)-X(\delta_t)|^2\big]\\
&\le (4\alpha+c_1)\delta \mathrm{e}^{(2+4\alpha)\delta}\mathbb{E}\big[|X(\delta_t)|^2\big]+(2\beta+c_0)\delta\mathrm{e}^{(2+4\alpha)\delta}\\
&\le 2(4\alpha+c_1)\delta \mathrm{e}^{(2+4\alpha)\delta}\mathbb{E}\big[|X(t)-X(\delta_t)|^2\big]+2(4\alpha+c_1)\delta \mathrm{e}^{(2+4\alpha)\delta}\mathbb{E}\big[|X(t)|^2\big]\\
&\qquad +(2\beta+c_0)\delta\mathrm{e}^{(2+4\alpha)\delta},
\end{aligned}
$$

which yields the estimate (3.8) immediately if $\tilde{K}(\delta)<1$. □

After these preparations, we shall apply the Lyapunov criterion on geometric ergodicity of discrete time Markov processes to study the long time behavior of the process $(X(t))$. Such kind of criterion was extensively studied in Meyn and Tweedie [15]. In addition, Mattingly et al. [14] provided a self-contained proof in terms of reachability structure which arises in many applications to SDEs.

Theorem 3.3. *Assume that* (H1)-(H4) *and* (H5) *hold. If* $\tilde{K}(\delta)<1$ *and*

$$
2\eta-2-2\alpha-\frac{K_u^2\tilde{K}(\delta)}{1-\tilde{K}(\delta)}>0,
$$

then $(X(t),\Lambda(t))$ *is weakly stable in distribution. Precisely, there is a probability measure* μ *on* $\mathscr{C}([-\delta,0];\mathbb{R}^d)\times\mathscr{D}([-\delta,0];\mathcal{S})$ *such that for every finite set* $F=\{0=r_0<r_1<\ldots<r_m<\delta\}$ $\mathscr{L}_{(X_{(n+1)\delta},\Lambda_{(n+1)\delta})}\circ(\pi_F\times\pi_\Lambda)^{-1}$ *will converge weakly to* $\mu\circ(\pi_F\times\pi_\Lambda)^{-1}$ *as* $n\to\infty$.

Proof. Based on the observations in Lemma 3.1, we shall apply the Foster-Lyapunov drift condition and Kolmogorov extension theorem to prove this theorem. See Meyn and Tweedie [15, Theorem 15.0.1] or Mattingly et al. [14, Theorem 2.5] for the Foster-Lyapunov drift condition in the study of ergodicity of Markov processes.

Step 1. We shall prove the exponential ergodicity of any finite dimensional embedded Markov processes associated with $(X(t),\Lambda(t))_{t\ge0}$ based on Lemma 3.1. For each finite set $F=\{0=r_0<r_1<\ldots<r_k<\delta\}$ for some $k\in\mathbb{N}$, let

$$
Y_n=(X(n\delta+r_k),\ldots,X(n\delta+r_1),X(n\delta)),\quad Z_n=(Y_n,\Lambda(n\delta)),\quad n\ge0.
$$

Then $(Z_n)_{n\ge0}$ is a Markov process on the state space $\mathbb{R}^{(k+1)d}\times\mathcal{S}$ due to Lemma 3.1. Put $\mathscr{F}_n^Z=\sigma\big\{Z_m;\ 0\le m\le n\big\}$ and $\mathscr{F}_t=\sigma\{(W(s),\Lambda(s));0\le s\le t\}$. By virtue of non-degenerate condition (H4) and irreducibility of (q_{ij}), in view of the transition probability measure of (Z_n) given in (3.6) for the case $k=1$, it holds that $(Z_n)_{n\ge0}$ is φ-irreducible with the choice that

$$
\varphi(\mathrm{d}x,\mathrm{d}i)=\mathrm{Leb}(\mathrm{d}x)\times\delta_{\mathcal{S}},
$$

where $\mathrm{Leb}(\mathrm{d}x)$ denotes the Lebesgue measure on $\mathbb{R}^{kd}$ and $\delta_{\mathcal{S}}$ the Dirac measure over $\mathcal{S}$. The aperiodicity of (Z_n) is immediate from the positivity of the transition probability due to (3.6). What we need to do is to construct the desired Lyapunov function satisfying the drift condition. Namely, construct a function $V:\mathbb{R}^{(k+1)d}\times\mathcal{S}\to[1,\infty)$ such that

$$\mathbb{E}\big[V(Z_{n+1})|\mathscr{F}_n^Z\big] \le \theta V(Z_n) + \tilde{C} \tag{3.9}$$

for some constants $\theta \in (0,1)$, $\tilde{C} < \infty$.

We shall construct the desired Lyapunov function based on the following estimation. Due to (H1) and (H2),

$$\begin{aligned} \mathrm{d}|X(t)|^2 &= \big(2\langle X(t), b(X(t),\Lambda(t)) - u(X(\delta_t),\Lambda(\delta_t))\rangle + \|\sigma(X(t),\Lambda(t))\|_{\mathrm{HS}}^2\big)\mathrm{d}t \\ &\quad + 2\langle X(t), \sigma(X(t),\Lambda(t))\mathrm{d}W(t)\rangle \\ &\le \big[|X(t)|^2 + |b(X(t),\Lambda(t))|^2 + \|\sigma(X(t),\Lambda(t))\|_{\mathrm{HS}}^2 - 2\langle X(t), u(X(\delta_t),\Lambda(\delta_t))\rangle \\ &\quad + 2\langle X(t), u(X(t),\Lambda(\delta_t)) - u(X(\delta_t),\Lambda(\delta_t))\rangle\big]\mathrm{d}t \\ &\quad + 2\langle X(t), \sigma(X(t),\Lambda(t))\mathrm{d}W(t)\rangle \\ &\le \big[(2+2\alpha-2\eta)|X(t)|^2 + 2\beta + K_u^2|X(t)-X(\delta_t)|^2\big]\mathrm{d}t \\ &\quad + 2\langle X(t), \sigma(X(t),\Lambda(t))\mathrm{d}W(t)\rangle. \end{aligned}$$

According to Lemma 3.2,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\big[|X(t)|^2\big|\mathscr{F}_s\big] &\le \Big(2+2\alpha-2\eta+\frac{K_u^2\tilde{K}(\delta)}{1-\tilde{K}(\delta)}\Big)\mathbb{E}\big[|X(t)|^2\big|\mathscr{F}_s\big] \\ &\quad + \frac{K_u^2(2\beta+c_0)\delta}{1-\tilde{K}(\delta)}\mathrm{e}^{(2+4\alpha)\delta} + 2\beta. \end{aligned}$$

Setting

$$\theta_4(\eta,\delta) = 2\eta - 2 - 2\alpha - \frac{K_u^2\tilde{K}(\delta)}{1-\tilde{K}(\delta)}, \tag{3.10}$$

$$\theta_5(\delta) = \frac{K_u^2(2\beta+c_0)\delta}{1-\tilde{K}(\delta)}\mathrm{e}^{(2+4\alpha)\delta} + 2\beta, \tag{3.11}$$

and if $\theta_4(\eta,\delta) > 0$, we obtain that for $t > s \ge 0$,

$$\mathbb{E}\big[|X(t)|^2\big|\mathscr{F}_s\big] \le |X(s)|^2\mathrm{e}^{-\theta_4(\eta,\delta)(t-s)} + \frac{\theta_5(\delta)}{\theta_4(\eta,\delta)}\Big(1-\mathrm{e}^{-\theta_4(\eta,\delta)(t-s)}\Big). \tag{3.12}$$

In addition, by virtue of (2.1), it follows from Itô's formula that

$$\mathbb{E}\big[H(\Lambda(t))\big|\mathscr{F}_s\big] \le \mathrm{e}^{-\kappa_1(t-s)}H(\Lambda(s)) + \frac{\kappa_2}{\kappa_1}\big(1-\mathrm{e}^{-\kappa_1(t-s)}\big), \quad t > s \ge 0. \tag{3.13}$$

According to the estimates (3.12) and (3.13), we construct the desired Lyapunov function V in the following way: for arbitrary $\varepsilon_0,\ldots,\varepsilon_{k-1} \in (0,1)$, let

$$V(x_k, x_{k-1},\ldots,x_0,i) = |x_k|^2 + \sum_{\ell=0}^{k-1}\varepsilon_\ell|x_\ell|^2 + H(i), \quad x_\ell \in \mathbb{R}^d,\ \ell\in\{0,\ldots,k\},\ i\in\mathcal{S}.$$

Obviously, we have $V \ge 1$ as $H \ge 1$. Notice that $\mathscr{F}_n^Z \subset \mathscr{F}_{n\delta+r_k}$, and $X(n\delta+r_k), \Lambda(n\delta) \in \mathscr{F}_n^Z$ by the definition of (Z_n). Invoking the estimates (3.12) and (3.13), we obtain that

$$
\begin{aligned}
&\mathbb{E}\Big[V(Z_{n+1})\big|\mathscr{F}_n^Z\Big] = \mathbb{E}\Big[\mathbb{E}\Big[V(Z_{n+1})\big|\mathscr{F}_{n\delta+r_k}\Big]\Big|\mathscr{F}_n^Z\Big]\\
&= \mathbb{E}\Big[\mathbb{E}\Big[|X((n+1)\delta+r_k)|^2 + \sum_{\ell=0}^{k-1}\varepsilon_\ell|X((n+1)\delta+r_\ell)|^2 + H(\Lambda((n+1)\delta))\big|\mathscr{F}_{n\delta+r_k}\Big]\Big|\mathscr{F}_n^Z\Big]\\
&\le \mathbb{E}\Big[|X(n\delta+r_k)|^2\Big(\mathrm{e}^{-\theta_4(\eta,\delta)\delta} + \sum_{\ell=0}^{k-1}\varepsilon_\ell\mathrm{e}^{-\theta_4(\eta,\delta)(\delta+r_\ell-r_k)}\Big) + H(\Lambda(n\delta+r_k))\mathrm{e}^{-\kappa_1(\delta-r_k)}\Big|\mathscr{F}_n^Z\Big]\\
&\quad + \frac{\theta_5(\delta)}{\theta_4(\eta,\delta)}\Big(1-\mathrm{e}^{-\theta_4(\eta,\delta)\delta} + \sum_{\ell=0}^{k-1}\varepsilon_\ell\Big(1-\mathrm{e}^{-\theta_4(\eta,\delta)(\delta+r_\ell-r_k)}\Big)\Big) + \frac{\kappa_2}{\kappa_1}\big(1-\mathrm{e}^{-\kappa_1(\delta-r_k)}\big)\\
&\le |X(n\delta+r_k)|^2\Big(\mathrm{e}^{-\theta_4(\eta,\delta)\delta} + \sum_{\ell=0}^{k-1}\varepsilon_\ell\mathrm{e}^{-\theta_4(\eta,\delta)(\delta+r_\ell-r_k)}\Big) + H(\Lambda(n\delta))\mathrm{e}^{-\kappa_1\delta}\\
&\quad + \frac{\theta_5(\delta)}{\theta_4(\eta,\delta)}\Big(1-\mathrm{e}^{-\theta_4(\eta,\delta)\delta} + \sum_{\ell=0}^{k-1}\varepsilon_\ell\Big(1-\mathrm{e}^{-\theta_4(\eta,\delta)(\delta+r_\ell-r_k)}\Big)\Big) + \frac{\kappa_2}{\kappa_1}\big(2-\mathrm{e}^{-\kappa_1 r_k}-\mathrm{e}^{-\kappa_1(\delta-r_k)}\big)\\
&\le V(Z_n)\max\Big\{\mathrm{e}^{-\theta_4(\eta,\delta)\delta} + \sum_{\ell=0}^{k-1}\varepsilon_\ell\Big(1-\mathrm{e}^{-\theta_4(\eta,\delta)(\delta+r_\ell-r_k)}\Big), \mathrm{e}^{-\kappa_1\delta}\Big\}\\
&\quad + \frac{\theta_5(\delta)}{\theta_4(\eta,\delta)}\Big(1-\mathrm{e}^{-\theta_4(\eta,\delta)\delta} + \sum_{\ell=0}^{k-1}\varepsilon_\ell\Big(1-\mathrm{e}^{-\theta_4(\eta,\delta)(\delta+r_\ell-r_k)}\Big)\Big) + \frac{\kappa_2}{\kappa_1}\big(2-\mathrm{e}^{-\kappa_1 r_k}-\mathrm{e}^{-\kappa_1(\delta-r_k)}\big).
\end{aligned}
$$

As a consequence, since $\theta_4(\eta,\delta) > 0$ and $\varepsilon_0,\dots,\varepsilon_{k-1} \in (0,1)$ can be taken arbitrarily small, we can find $\varepsilon_0,\dots,\varepsilon_{k-1} \in (0,1)$ such that

$$
\mathrm{e}^{-\theta_4(\eta,\delta)\delta} + \sum_{\ell=0}^{k-1}\varepsilon_\ell\Big(1-\mathrm{e}^{-\theta_4(\eta,\delta)(\delta+r_\ell-r_k)}\Big) < 1,
$$

and further we have

$$
\theta := \max\Big\{\mathrm{e}^{-\theta_4(\eta,\delta)\delta} + \sum_{\ell=0}^{k-1}\varepsilon_\ell\Big(1-\mathrm{e}^{-\theta_4(\eta,\delta)(\delta+r_\ell-r_k)}\Big), \mathrm{e}^{-\kappa_1\delta}\Big\} < 1.
$$

Therefore, the function V satisfied the drift condition (3.9). Due to [14, Theorem 2.5], the Markov process $(Z_n)_{n\ge 0}$ is exponentially ergodic. Namely, there are constants c_3, $\lambda_3 > 0$ and a probability measure $\Gamma_{r_k,\dots,r_1,r_0,i}$ on $\mathbb{R}^{(k+1)d} \times \mathcal{S}$ such that

$$
\big|\mathbb{E}\big[f(Z_n)\big] - \Gamma_{r_k,\dots,r_1,r_0,i}(f)\big| \le c_3\mathrm{e}^{-\lambda_3 n} \tag{3.14}
$$

for all $f : \mathbb{R}^{(k+1)d} \times \mathcal{S} \to \mathbb{R}$ such that $|f| \le V$. As $V \ge 1$, (3.14) implies that

$$
\|\mathscr{L}_{(X(n\delta+r_k),\dots,X(n\delta+r_1),X(n\delta),\Lambda(n\delta))} - \Gamma_{r_k,\dots,r_1,r_0,i}\|_{\mathrm{var}} \le c_3\mathrm{e}^{-\lambda_3 n}. \tag{3.15}
$$

Besides, by the ergodicity of $(\Lambda(t))$ due to (H5), (3.14) also yields the projection measure of $\Gamma_{r_k,\dots,r_1,r_0,i}$ into $\mathcal{S}$ must equal to γ, i.e. $\Gamma_{r_k,\dots,r_1,r_0,i} \circ \pi_\Lambda^{-1} = \gamma$.

Step 2. Let $\mathcal{D}$ be the collection of all finite subsets of $[0,\delta)$, ordered by set inclusion. By the arbitrariness of the finite set F in the argument of **Step 1**, we obtain that for each $F = \{0 = r_0 < r_1 < \dots < r_k < \delta\}$, there is a probability measure $\Gamma_F := \Gamma_{r_k,\dots,r_1,r_0}$ on the product space $\mathbb{R}^F \times \mathcal{S}$, where $\mathbb{R}^F := \mathbb{R}^{(k+1)d}$. Therefore, we obtain a family of probability measures $\{\Gamma_F;\ F \in \mathcal{D}\}$. We shall check its consistent property below.

For $F, \widetilde{F} \in \mathcal{D}$ satisfying $\widetilde{F} \subset F$, denote by $\pi_{F,\widetilde{F}}$ the projection map from $\mathbb{R}^F$ to $\mathbb{R}^{\widetilde{F}}$. The Kolmogorov consistent condition obviously holds for the finite dimension marginal distribution of $(X_{(n+1)\delta})$ for every $n \in \mathbb{N}$, that is,

$$\mathscr{L}_{X_{(n+1)\delta}} \circ \pi_{\widetilde{F}}^{-1} = \big(\mathscr{L}_{X_{(n+1)\delta}} \circ \pi_F^{-1}\big) \circ \pi_{F,\widetilde{F}}^{-1}.$$

Invoking the convergence of (3.15), this yields that

$$\Gamma_{\widetilde{F}} \circ \pi_{\widetilde{F}}^{-1} = \big(\Gamma_F \circ \pi_F^{-1}\big) \circ \pi_{F,\widetilde{F}}^{-1}.$$

So, the family of probability measures $\{\Gamma_F;\ F \in \mathcal{D}\}$ on $\mathscr{C}([-\delta, 0]; \mathbb{R}^d) \times \mathcal{S}$ is Kolmogorov consistent. Then, by the Kolmogorov extension theorem, there is a unique probability measure $\tilde{\mu}$ on $\mathscr{C}([-\delta, 0]; \mathbb{R}^d) \times \mathcal{S}$ such that

$$\tilde{\mu}\big(\pi_F^{-1}(A) \times B\big) = \Gamma_F(A \times B), \quad \forall\, A \in \mathscr{B}(\mathbb{R}^F),\ B \in \mathscr{B}(\mathcal{S}),\ \forall\, F \in \mathcal{D}. \tag{3.16}$$

Let μ be a probability measure on $\mathscr{C}([-\delta, 0]; \mathbb{R}^d) \times \mathscr{D}([-\delta, 0]; \mathcal{S})$ given by

$$\mu(A \times G) = \tilde{\mu}(A \times \pi_\Lambda(G)), \quad \forall\, A \in \mathscr{B}\big(\mathscr{C}([-\delta, 0]; \mathbb{R}^d)\big),\ G \in \mathscr{B}\big(\mathscr{D}([-\delta, 0]; \mathcal{S})\big).$$

Using (3.15) again, we have

$$\|\mathscr{L}_{(X(n\delta+r_m),\dots,X(n\delta+r_1),\Lambda(n\delta))} - \mu \circ (\pi_F \times \pi_\Lambda)^{-1}\|_{\mathrm{var}} \to 0$$

as $n \to \infty$. By Definition 2.2, $(X(t), \Lambda(t))$ is weakly stable in distribution. □

Corollary 3.4. *Under the conditions of Theorem 3.3, the obtained measure $\tilde{\mu}$ in (3.16) on $\mathscr{C}([-\delta, 0]; \mathbb{R}^d) \times \mathcal{S}$ is also a stationary probability measure for the process $(X_{(n+1)\delta}, \Lambda(n\delta))_{n\ge 0}$.*

Proof. For $F \in \mathcal{D}$, denote $\mathscr{C}_b(\mathbb{R}^F \times \mathcal{S})$ the set of bounded continuous functions on $\mathbb{R}^F \times \mathcal{S}$. As an application of the Kolmogorov extension theorem, if for two probability measures ν and $\tilde{\nu}$ on $\mathscr{C}([-\delta, 0]; \mathbb{R}^d) \times \mathcal{S}$, if

$$\nu \circ \pi_F^{-1}(f) = \tilde{\nu} \circ \pi_F^{-1}(f), \quad \forall\, f \in \mathscr{C}_b(\mathbb{R}^F \times \mathcal{S}),\ F \in \mathcal{D},$$

then $\nu = \tilde{\nu}$.

According to the argument of **Step 1** in Theorem 3.3, $\Gamma_{r_k,\dots,r_1,r_0,i}$ is the unique stationary distribution of the Markov processes $\big(X(n\delta + r_k), \dots, X(n\delta + r_1), X(n\delta), \Lambda(n\delta)\big)_{n\ge 0}$, which yields that if $\big(X(r_k), \dots, X(r_1), X(0), \Lambda(0)\big)$ is distributed as $\Gamma_{r_k,\dots,r_1,r_0,i}$, then $\big(X(n\delta+r_k), \dots, X(n\delta+r_1), X(n\delta), \Lambda(n\delta)\big)$ for all $n \in \mathbb{N}$ is also distributed as $\Gamma_{r_k,\dots,r_1,r_0,i}$. Thus, for every $f \in \mathscr{C}_b(\mathbb{R}^F \times \mathcal{S})$

$$\mathbb{E}_{\Gamma_{r_k,\dots,r_1,r_0,i}}\big[f\big(X(n\delta + r_k), \dots, X(n\delta + r_1), X(n\delta), \Lambda(n\delta)\big)\big] = (\Gamma_{r_k,\dots,r_1,r_0,i})(f).$$

Invoking (3.16), we can rewrite above equation into

$$\mathbb{E}_{\tilde{\mu}}\big[(f \circ (\pi_F \times \mathrm{Id}_{\mathcal{S}}))(X_{(n+1)\delta}, \Lambda(n\delta))\big] = (\tilde{\mu} \circ (\pi_F \times \mathrm{Id}_{\mathcal{S}})^{-1})(f), \tag{3.17}$$

where $\mathrm{Id}_{\mathcal{S}}$ stands for the identity map on $\mathcal{S}$. Together with the arbitrariness of $F \in \mathcal{D}$ and $f \in \mathscr{C}_b(\mathbb{R}^F \times \mathcal{S})$, this means that if $(X_\delta, \Lambda(0))$ is distributed as $\tilde{\mu}$ on $\mathscr{C}([-\delta, 0]; \mathbb{R}^d) \times \mathcal{S}$, the distribution of $(X_{(n+1)\delta}, \Lambda(n\delta))$ will coincide with μ for $n \ge 1$. Namely, $\tilde{\mu}$ is a stationary probability measure for the process $(X_{(n+1)\delta}, \Lambda(n\delta))_{n\ge 0}$. □

Remark 3.5. Notice that the weak convergence of probability measures on the path space $\mathscr{C}([-\delta, 0]; \mathbb{R}^d)$ implies the convergence of corresponding finite dimension projection measures. But the converse fails. See [4, Sec. 1.5]. So, the convergence of Theorem 3.3 is weaker than that of Theorem 2.8. Certainly, the conditions of Theorem 3.3 are weaker than those of Theorem 2.8 as well.

Funding

This work is supported in part by National Key R&D Program of China (No. 2022YFA1006000) and NNSF of China (No. 12271397).

Declaration of competing interest

The authors declare that they have no known competing interests.

References

[1] Y. Bakhtin, J. Mattingly, Stationary solutions of stochastic differential equations with memory and stochastic partial differential equations, Commun. Contemp. Math. 7 (2005) 553–582.
[2] J. Bao, J. Shao, C. Yuan, Invariant probability measures for path-dependent random diffusions, Nonlinear Anal. 228 (2023) 113201.
[3] G. Basak, R. Bhattacharya, Stability in distribution for a class of singular diffusions, Ann. Probab. 20 (1992) 312–321.
[4] P. Billinsley, Convergence of Probability Measures, John Wiley, New York, 1968.
[5] O. Butkovsky, M. Scheutzow, Invariant measures for stochastic functional differential equations, Electron. J. Probab. 22 (2017) 1–23.
[6] X. Chen, X. Li, X. Mao, Stabilization of hybrid stochastic differential equations by delay feedback control based on discrete-time observations, Numer. Algebra Control Optim. (2024), https://doi.org/10.3934/naco.2024009.
[7] P. Chow, R. Khasminskii, Stationary solutions of nonlinear stochastic evolution equations, Stoch. Anal. Appl. 15 (1997) 671–699.
[8] G. Da Prato, J. Zabczyk, Ergodicity for Infinite Dimensional Systems, London Mathematical Society Lecture Notes, vol. 229, Cambridge University Press, 1996.
[9] A. Es-Sarhir, O. van Gaans, M. Scheutzow, Invariant measures for stochastic functional differential equations with superlinear drift term, Differ. Integral Equ. 23 (2010) 189–200.
[10] S. Ethier, T. Kurtz, Markov Processes Characterization and Convergence, Wiley, New York, 1986.
[11] L. Hu, Y. Ren, Q. Zhang, Periodically intermittent discrete observations control for stabilization of highly nonlinear hybrid stochastic differential equations, Nonlinear Anal. Hybrid Syst. 50 (2023) 101400.
[12] Y. Li, J. Lu, X. Mao, Q. Qiu, Stabilization of hybrid systems by feedback control based on discrete-time state and mode observations, Asian J. Control 19 (2017) 1943–1953.
[13] X. Li, X. Mao, D. Mukama, C. Yuan, Delay feedback control for switching diffusion systems based on discrete time observations, SIAM J. Control Optim. 58 (2020) 2900–2926.
[14] J. Mattingly, A. Stuat, D. Higham, Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise, Stoch. Process. Appl. 101 (2002) 185–232.
[15] S. Meyn, R. Tweedie, Markov Chains and Stochastic Stability, Springer Verlag, 1993.
[16] X. Mao, Stabilization of continuous-time hybrid stochastic differential equations by discrete-time feedback control, Automatica 49 (2013) 3677–3681.
[17] X. Mao, W. Liu, L. Hu, Q. Luo, J. Lu, Stabilization of hybrid stochastic differential equations by feedback control based on discrete-time state observations, Syst. Control Lett. 73 (2014) 88–95.
[18] S. Mohammed, Stochastic Functional Differential Equations, Pitman Publishing Pty Ltd, Melbourne, 1984.
[19] B. Øksendal, Stochastic Differential Equations, an Introduction with Applications, 6th ed., Springer, 2003.
[20] J. Shao, Stabilization of regime-switching processes by feedback control based on discrete time observations, SIAM J. Control Optim. 55 (2017) 724–740.
[21] J. Shao, Comparison theorem and stability under perturbation of transition rate matrices for regime-switching processes, J. Appl. Probab. 61 (2024) 540–557.
[22] J. Shao, F. Xi, Stabilization of regime-switching processes by feedback control based on discrete time observations II: state-dependent case, SIAM J. Control Optim. 57 (2019) 1413–1439.
[23] G. Song, B. Zheng, Q. Luo, X. Mao, Stabilisation of hybrid stochastic differential equations by feedback control based on discrete-time observations of state and mode, IET Control Theory Appl. 11 (2017) 301–307.
[24] F. Xi, Feller property and exponential ergodicity of diffusion processes with state-dependent switching, Sci. China Ser. A, Math. 51 (2008) 329–342.
[25] G. Yin, C. Zhu, Hybrid Switching Diffusions: Properties and Applications, Stoch. Model. Appl. Prob., vol. 63, Springer, New York, 2010.

[26] S. You, W. Liu, X. Mao, Q. Qiu, Stabilization of hybrid systems by feedback control based on discrete-time state observations, SIAM J. Control Optim. 53 (2015) 905–925.
[27] C. Yuan, X. Mao, Asymptotic stability in distribution of stochastic differential equations with Markovian switching, Stoch. Process. Appl. 103 (2003) 277–291.
[28] C. Yuan, J. Zou, X. Mao, Stability in distribution of stochastic differential delay equations with Markovian switching, Syst. Control Lett. 50 (2003) 195–207.