Invariant Measures for Path-Dependent Random Diffusions*

Jianhai Bao[b)], **Jinghai Shao**[a)], **Chenggui Yuan**[c)]

[a)]Center of Applied Mathematics, Tianjin University, Tianjin 300072, China

[b)]School of Mathematics and Statistics, Central South University, Changsha 410083, China

[c)]Department of Mathematics, Swansea University, Singleton Park, SA2 8PP, UK

jianhaibao13@gmail.com, shaojh@bnu.edu.cn, C.Yuan@swansea.ac.uk

June 20, 2017

Abstract

In this work, we are concerned with existence and uniqueness of invariant measures for path-dependent random diffusions and their time discretizations. The random diffusion here means a diffusion process living in a random environment characterized by a continuous time Markov chain. Under certain ergodic conditions, we show that the path-dependent random diffusion enjoys a unique invariant probability measure and converges exponentially to its equilibrium under the Wasserstein distance. Also, we demonstrate that the time discretization of the path-dependent random diffusion involved admits a unique invariant probability measure and shares the corresponding ergodic property when the stepsize is sufficiently small. During this procedure, the difficulty arose from the time-discretization of continuous time Markov chain has to be deal with, for which an estimate on its exponential functional is presented.

AMS subject Classification: 37A25 · 60H10 · 60H30 · 60F10 · 60K37.
Keywords: Invariant measure; Path-dependent random diffusion; Ergodicity; Wasserstein distance; Euler-Maruyama scheme

1 Introduction and Main Results

A random diffusion is a Markov process consisting of two components $(X(t), \Lambda(t))$, where the first component $X(t)$ means the underlying continuous dynamics and the second one

*Supported in part by NNSFs of China (Nos. 11301030, 11431014, 11401592) and 985-project.

$\Lambda(t)$ stands for a jump process. Such diffusions have a wide range of emerging and existing applications in, for instance, climate science, material science, molecular biology, ecosystems, econometric modeling, and control and optimization of largescale systems; see, e.g., [6, 14, 17, 18, 20, 29, 36] and references therein. Viewing random diffusions as a number of diffusions with random switching, they may be seemingly not much different from their diffusion counterpart. Nevertheless, the coexistence of continuous dynamics and jump processes results in challenge in dealing with random diffusions $(X(t), \Lambda(t))$ under consideration, even though, in each random temporal environment, $X(t)$ is simple enough for intuitive understanding. [16] revealed that $X(t)$ is exponentially stable in p-th moment in a random temporal environment and algebraically stable in p-th moment in the other scenarios, whereas $X(t)$ is ultimately exponentially stable; [23, 24] constructed several very interesting examples to show that $(X(t), \Lambda(t))$ is recurrent (resp. transient) even if $X(t)$ is transient (resp. recurrent) in each random temporal environment; Unlike Ornstein-Uhlenbeck (OU) process admitting light tail, the random OU process enjoys heavy tail property shown in [4, 10].

Recently, ergodicity of random diffusions with state dependent or state independent jump rates has been investigated extensively; see, for example, [4, 5, 9, 26, 27, 28] for the setting of state independent jump rates, [5, 9, 27] for the setup of bounded state dependent jump rates, [18, 30] for the framework of unbounded and state dependent jump rates. So far, there are several approaches to explore ergodicity for random diffusions; see, for instance, [4, 5, 27] via probabilistic coupling argument, [9, 18, 30] by weak Harris' theorem, [26, 27] based on the theory of M-matrix and Perron-Frobenius theorem. For the ergodicity of random diffusions with infinite regimes, we refer to [26, 27, 32].

More often than not, to understand very well the behavior of numerous real-world systems, one of the better ways is to take the influence of past events on the current and future states of the systems involved into consideration. Such point of view is especially appropriate in the study on population biology, neural networks, viscoelastic materials subjected to heat or mechanical stress, and financial products, to name a few, since predictions on their evolution rely heavily on the knowledge of their past; see, for instance, [2, 8, 19, 21] and references therein for more details. There is vast literature on path-dependent ordinary differential equation, among which the monograph [13] provides an introduction to this subject. Also, there is a sizeable literature on path-dependent stochastic differential equations (SDEs); see, e.g., [12, 15, 22, 25] and references therein. Concerning existence and uniqueness of invariant probability measures for path-dependent SDEs, we refer to [25], where the drift term is semi-linear, [11] with the drift part being superlinear growth and satisfying a dissipativity condition, and [7] under the extended Veretennikov-Khasminski condition.

Under certain Lyapunov condition which is not related to stationary distribution of Markov chain involved, [33, 34] investigated existence and uniqueness of invariant probability measures for a class of random diffusions by exploiting the M-matrix trick, and [35] further discussed the same issue for a range of path-dependent random diffusions. Recently, under ergodic conditions, [3] probed deeply into existence and uniqueness of invariant probability measures for a kind of random diffusions by developing new analytical frameworks.

As described above, there is a natural motivation for considering stochastic dynamical systems, where all three features (i.e. random switching, path dependence and noise) are present. In this work, we are interested in ergodic properties for path-dependent random diffusions. More precisely, as a continuation of [3], in the current work we are concerned with existence and uniqueness of invariant probability measures not only for path-dependent random diffusions but also for their time discretizations. In comparison with [3, 33, 34], the difficulties to deal with existence and uniqueness of (numerical) invariant probability measures for path-dependent random diffusions lie in: (i) the state space of functional solutions $(X_t)_{t\geq 0}$ is an infinite-dimensional space; (ii) Both components of $(X_t, \Lambda(t))$ are discretized, where, in particular, time discretization of continuous time Markov chain causes additional difficulties in analyzing the long-term behavior of numerical scheme; (iii) Our investigation is based on certain ergodic conditions. So, it turns out to be much more challenging to cope with long term (numerical) behavior of path-dependent random diffusions. In the present work, it is worthy to pointing out that $(X_t, \Lambda(t))$ possesses a unique invariant probability measure although the functional solution X_t doesn't admit an invariant probability measure in some fixed environment, which is quite different from the existing results; see, e.g. [7, 11, 25]. For more and precise interpretations on $(X_t)_{t\geq 0}$ and $(\Lambda(t))_{t\geq 0}$, please refer to subsections 1.1-1.4.

Prior to presentation of the setting for this work, we consider and introduce some notation and terminology needed in the rest of the paper. Let $(\mathbb{R}^n, \langle\cdot,\cdot\rangle, |\cdot|)$ be the n-dimensional Euclidean space. For fixed $\tau > 0$, let $\mathscr{C} = C([-\tau, 0]; \mathbb{R}^n)$ denote the family of all continuous functions $f : [-\tau, 0] \to \mathbb{R}^n$, endowed with the uniform norm $\|f\|_\infty := \sup_{-\tau\leq\theta\leq 0} |f(\theta)|$. Let $\mathbf{S} = \{1, 2 \cdots, N\}$ for some integer $N \in [2, \infty)$. Let $(\Lambda(t))$ stand for a continuous–time Markov chain with the state space $\mathbf{S}$, and the transition rules specified by

$$\mathbb{P}(\Lambda(t+\triangle) = j|\Lambda(t) = i) = \begin{cases} q_{ij}\triangle + o(\triangle), & i \neq j \\ 1 + q_{ii}\triangle + o(\triangle), & i = j \end{cases} \tag{1.1}$$

provided $\triangle \downarrow 0$, where $o(\triangle)$ means that $\lim_{\triangle\to 0} \frac{o(\triangle)}{\triangle} = 0$, and $Q = (q_{ij})$ be the Q-matrix associated with the Markov chain $(\Lambda(t))$. Let $(W(t))$ be an m-dimensional Brownian motion. We assume that $(\Lambda(t))$ is irreducible, together with the finiteness of $\mathbf{S}$, which yields the positive recurrence. Let $\pi = (\pi_1, \cdots, \pi_N)$ denote its stationary distribution, which can be solved by $\pi Q = 0$ subject to $\sum_{i\in\mathbf{S}} \pi_i = 1$ with $\pi_i \geq 0$. Assume that $(\Lambda(t))$ is independent of $(W(t))$. Let $\|\cdot\|_{\rm HS}$ means the Hilbert-Schmidt norm. Let $\mathbf{E} = \mathscr{C} \times \mathbf{S}$. For any $\mathbf{x} = (\xi, i) \in \mathbf{E}$ and $\mathbf{y} = (\eta, j) \in \mathbf{E}$, define the distance ρ between $\mathbf{x}$ and $\mathbf{y}$ by

$$\rho(\mathbf{x}, \mathbf{y}) = \|\xi - \eta\|_\infty + \mathbf{1}_{\{i\neq j\}},$$

where, for a set A, $\mathbf{1}_A(x) = 1$ with $x \in A$; otherwise, $\mathbf{1}_A(x) = 0$. Let $\mathcal{P} = \mathcal{P}(\mathbf{E})$ be the space of all probability measures on $\mathbf{E}$. Set

$$\mathcal{P}_0 = \{\nu \in \mathcal{P};\ \int_{\mathbf{E}} \|\xi\|_\infty \nu({\rm d}\xi) < \infty\}.$$

Define the Wasserstein distance W_ρ between two probability measures μ, $\nu \in \mathcal{P}_0$ as follows:

$$W_\rho(\mu, \nu) = \inf_{\pi\in\mathcal{C}(\mu,\nu)} \Big\{ \int_{\mathbf{E}\times\mathbf{E}} \rho(\mathbf{x}, \mathbf{y})\pi({\rm d}\mathbf{x}, {\rm d}\mathbf{y}) \Big\} < \infty,$$

where $\mathcal{C}(\mu,\nu)$ denotes the collection of all probability measures on $\mathbf{E}\times\mathbf{E}$ with marginals μ and ν, respectively. In this work, $c>0$ will stand for a generic constant which might change from occurrence to occurrence.

Next, we present the framework of this work and state our main results.

1.1 Invariant Measures: Additive Noises

In this subsection, we focus on a path-dependent random diffusion with additive noise

$$\mathrm{d}X(t)=b(X_t,\Lambda(t))\mathrm{d}t+\sigma(\Lambda(t))\mathrm{d}W(t),\quad t>0,\quad X_0=\xi\in\mathscr{C},\quad \Lambda(0)=i_0\in\mathbf{S}, \tag{1.2}$$

where $b:\mathscr{C}\times\mathbf{S}\to\mathbb{R}^n$, $\sigma:\mathbf{S}\to\mathbb{R}^n\otimes\mathbb{R}^m$, and, for fixed $t\geq 0$, $X_t(\theta)=X(t+\theta)$, $\theta\in[-\tau,0]$, used the standard notation.

We assume that, for each $i\in\mathbf{S}$ and arbitrary $\xi,\eta\in\mathscr{C}$,

($\mathbf{A}$) There exist $\alpha_i\in\mathbb{R}$ and $\beta_i\in\mathbb{R}_+$ such that

$$2\langle\xi(0)-\eta(0),b(\xi,i)-b(\eta,i)\rangle\leq\alpha_i|\xi(0)-\eta(0)|^2+\beta_i\|\xi-\eta\|_\infty^2.$$

Under ($\mathbf{A}$), in terms of [31, Theorem 2.3], (1.2) admits a unique strong solution $(X(t;\xi,i_0))$ with the initial datum $X_0=\xi\in\mathscr{C}$ and $\Lambda(0)=i_0\in\mathbf{S}$. The segment process (i.e., functional solution) associated with the solution process $(X(t;\xi,i_0))$ is denoted by $(X_t(\xi,i_0))$. The pair $(X_t(\xi,i_0),\Lambda(t))$ is a homogeneous Markov process; see, for instance, [22, Theorem 1.1] & [25, Proposition 3.4].

For (α_i) and (β_i) introduced in ($\mathbf{A}$), set

$$\widehat{\alpha}:=\min_{i\in\mathbf{S}}\alpha_i,\quad \check{\alpha}:=\max_{i\in\mathbf{S}}|\alpha_i|\quad\text{and}\quad\check{\beta}:=\max_{i\in\mathbf{S}}\beta_i. \tag{1.3}$$

Moreover, set

$$Q_1:=Q+\operatorname{diag}\Big(\alpha_1+\mathrm{e}^{-\widehat{\alpha}\tau}\beta_1,\cdots,\alpha_N+\mathrm{e}^{-\widehat{\alpha}\tau}\beta_N\Big),$$

where Q is the Q-matrix of the Markov chain $(\Lambda(t))$, $\tau>0$ is the length of time lag, and $\operatorname{diag}(x_1,\dots,x_N)$ denotes the diagonal matrix generated by the vector $(x_1,\dots,x_N)$. Let

$$\eta_1=-\max_{\gamma\in\operatorname{spec}(Q_1)}\operatorname{Re}(\gamma), \tag{1.4}$$

where $\operatorname{spec}(Q_1)$ and $\operatorname{Re}(\gamma)$ denote respectively the spectrum (i.e., the multiset of its eigenvalues) of Q_1 and the real part of γ. Let $(\Lambda^i(t),\Lambda^j(t))$ be the independent coupling of the Q-process $(\Lambda(t))$ with starting point $(\Lambda^i(0),\Lambda^j(0))=(i,j)$. Let $T=\inf\{t\geq0:\Lambda^i(t)=\Lambda^j(t)\}$ be the coupling time of $(\Lambda^i(t),\Lambda^j(t))$. Since the cardinality of $\mathbf{S}$ is finite and (q_{ij}) is irreducible, there exists a constant $\theta>0$ such that

$$\mathbb{P}(T>t)\leq\mathrm{e}^{-\theta t},\quad t>0. \tag{1.5}$$

Let $P_t((\xi,i),\cdot)$ be the transition kernel of $(X_t(\xi,i),\Lambda^i(t))$. For $\nu\in\mathcal{P}$, νP_t denotes the law of $(X_t(\xi,i),\Lambda^i(t))$ when $(X_0(\xi,i),\Lambda^i(0))$ is distributed according to $\nu\in\mathcal{P}$.

Our first main result in this paper is stated as follows.

Theorem 1.1. Suppose (**A**) holds and $\eta_1 > 0$. Then, it holds that

$$W_\rho(\nu_1 P_t, \nu_2 P_t) \le c\Big(1 + \sum_{i\in\mathbf{S}} \int_{\mathscr{C}} \|\xi\|_\infty \nu_1(\mathrm{d}\xi, i) + \sum_{i\in\mathbf{S}} \int_{\mathscr{C}} \|\eta\|_\infty \nu_2(\mathrm{d}\eta, i)\Big) \mathrm{e}^{-\frac{\theta\eta_1}{2(\theta+\eta_1)}t} \tag{1.6}$$

for any $\nu_1, \nu_2 \in \mathcal{P}_0$, where η_1 is defined in (1.4) and $\theta > 0$ is specified in (1.5). Furthermore, (1.6) implies that $(X_t(\xi, i), \Lambda^i(t))$, determined by (1.2) and (1.1), admits a unique invariant probability measure $\mu \in \mathcal{P}_0$ such that

$$W_\rho(\delta_{(\xi,i)} P_t, \mu P_t) \le c\Big(1 + \|\xi\|_\infty + \sum_{i\in\mathbf{S}} \int_{\mathscr{C}} \|\eta\|_\infty \mu(\mathrm{d}\eta, i)\Big) \mathrm{e}^{-\frac{\theta\eta_1}{2(\theta+\eta_1)}t}, \tag{1.7}$$

where $\delta_{(\xi,i)}$ stands for the Dirac's measure at the point (ξ, i).

Remark 1.1. If the assumption $\eta_1 > 0$ is replaced by

$$\sum_{i\in\mathbf{S}} (\alpha_i + \mathrm{e}^{-\widehat{\alpha}\tau}\beta_i)\pi_i < 0$$

and

$$\min_{i\in\mathbb{S}, \alpha_i + \mathrm{e}^{-\widehat{\alpha}\tau}\beta_i > 0} \Big(-\frac{q_{ii}}{\alpha_i + \mathrm{e}^{-\widehat{\alpha}\tau}\beta_i} \Big) > 1,$$

according to [4, Propositions 4.1 & 4.2], Theorem 1.1 still holds true.

Remark 1.2. From Theorem 1.1 and Remark 1.1, $(X_t, \Lambda(t))$ might have a unique invariant probability measure even though the functional solution X_t does not admit an invariant probability measure in a random temporal environment just as Example 1.3 below shows.

1.2 Invariant Measures: Multiplicative Noises

In this subsection, we move on to consider existence and uniqueness of invariant probability measures under a little bit strong assumptions but for path-dependent random diffusions with multiplicative noises in the form

$$\mathrm{d}X(t) = b(X_t, \Lambda(t))\mathrm{d}t + \sigma(X_t, \Lambda(t))\mathrm{d}W(t), \quad t > 0, \quad X_0 = \xi, \quad \Lambda(0) = i_0 \in \mathbf{S}, \tag{1.8}$$

where $b : \mathscr{C} \times \mathbf{S} \to \mathbb{R}^n$ and $\sigma : \mathscr{C} \times \mathbf{S} \to \mathbb{R}^n \otimes \mathbb{R}^m$.

Let $v(\cdot)$ be a probability measure on $[-\tau, 0]$ and suppose that, for any $\xi, \eta \in \mathscr{C}$ and each $i \in \mathbf{S}$,

(**H1**) There exist $\alpha_i \in \mathbb{R}$ and $\beta_i \in \mathbb{R}_+$ such that

$$\begin{aligned} &2\langle \xi(0) - \eta(0), b(\xi, i) - b(\eta, i)\rangle + \|\sigma(\xi, i) - \sigma(\eta, i)\|_{\mathrm{HS}}^2 \\ &\le \alpha_i |\xi(0) - \eta(0)|^2 + \beta_i \int_{-\tau}^0 |\xi(\theta) - \eta(\theta)|^2 v(\mathrm{d}\theta). \end{aligned}$$

(**H2**) There exists an $L > 0$ such that

$$\|\sigma(\xi, i) - \sigma(\eta, i)\|_{\rm HS}^2 \leq L\Big(|\xi(0) - \eta(0)|^2 + \int_{-\tau}^0 |\xi(\theta) - \eta(\theta)|^2 v(\mathrm{d}\theta)\Big).$$

For (α_i) and (β_i) stipulated in (**H1**), we set

$$Q_2 := Q + \mathrm{diag}\Big(\alpha_1 + \beta_1 \int_{-\tau}^0 \mathrm{e}^{\widehat{\alpha}\theta} v(\mathrm{d}\theta), \cdots, \alpha_N + \beta_N \int_{-\tau}^0 \mathrm{e}^{\widehat{\alpha}\theta} v(\mathrm{d}\theta)\Big),$$

where $\widehat{\alpha}$ is defined as in (1.3). Furthermore, we define

$$\eta_2 = -\max_{\gamma \in \mathrm{spec}(Q_2)} \mathrm{Re}(\gamma). \tag{1.9}$$

Under appropriate assumptions, the semigroup generated by the pair $(X_t(\xi, i), \Lambda^i(t))$ converges exponentially to the equilibrium under the Wasserstein distance as one of the main results below reads.

Theorem 1.2. Let (**H1**)-(**H2**) hold and assume further $\eta_2 > 0$. Then,

$$W_\rho(\nu_1 P_t, \nu_2 P_t) \leq c\left(1 + \sum_{i \in \mathbf{S}} \int_{\mathscr{C}} \|\xi\|_\infty \nu_1(\mathrm{d}\xi, i) + \sum_{i \in \mathbf{S}} \int_{\mathscr{C}} \|\eta\|_\infty \nu_2(\mathrm{d}\eta, i)\right) \mathrm{e}^{-\frac{\theta\eta_2}{2(\theta+\eta_2)}t} \tag{1.10}$$

for any $\nu_1, \nu_2 \in \mathcal{P}_0$, where $\theta > 0$ such that (1.5) holds and $\eta_2 > 0$ is defined in (1.9). Furthermore, (1.10) implies that $(X_t(\xi, i), \Lambda^i(t))$ solving (1.8) and (1.1) and admits a unique invariant probability measure $\mu \in \mathcal{P}_0$ such that

$$W_\rho(\delta_{(\xi,i)} P_t, \mu P_t) \leq c\Big(1 + \|\xi\|_\infty + \sum_{i \in \mathbf{S}} \int_{\mathscr{C}} \|\eta\|_\infty \mu(\mathrm{d}\eta, i)\Big) \mathrm{e}^{-\frac{\theta\eta_2}{2(\theta+\eta_2)}t}. \tag{1.11}$$

Next, we provide an example to demonstrate Theorem 1.2.

Example 1.3. Let $(\Lambda(t))_{t \geq 0}$ be a Markov chain taking values in $\mathbb{S} = \{1, 2\}$ with the generator

$$Q = \begin{pmatrix} -1 & 1 \\ \gamma & -\gamma \end{pmatrix} \tag{1.12}$$

for some constant $\gamma > 0$. Consider a scalar path-dependent OU process

$$\mathrm{d}X(t) = \{a_{\Lambda(t)} X(t) + b_{\Lambda(t)} X(t-1)\}\mathrm{d}t + \sigma_{\Lambda_t} \mathrm{d}W(t),\ t > 0,\ (X_0, \Lambda(0)) = (\xi, 1) \in \mathscr{C} \times \mathbb{S}, \tag{1.13}$$

where $a_1, b_1, b_2 > 0, a_2 < 0$. Set $\alpha := 2a_1 + (1 + \mathrm{e}^{-a_2})b_1,\ \beta := 2a_2 + (1 + \mathrm{e}^{-a_2})b_2$. For $\alpha, \gamma > 0, \beta \in \mathbb{R}$ above, if

$$\begin{cases} \alpha + \beta < 1 + \gamma \\ \beta - \frac{\beta}{\alpha} > \gamma. \end{cases} \tag{1.14}$$

then $(X_t(\xi, i), \Lambda^i(t))$, determined by (1.13) and (1.12), has a unique invariant probability measure, and converges exponentially to the equilibrium.

1.3 Numerical Invariant Measures: Additive Noises

In this subsection, we proceed to discuss existence and uniqueness of invariant probability measures for the time discretization of $(X_t(\xi, i_0), \Lambda^i(t))$, determined by (1.2) and (1.1), respectively, and investigate the exponential ergodicity under the Wasserstein distance.

Without loss of generality, we assume the step size $\delta = \frac{\tau}{M} \in (0, 1)$ for some integer $M > \tau$. Consider the following EM scheme associated with (1.2)

$$\mathrm{d}Y(t) = b(Y_{t_\delta}, \Lambda(t_\delta))\mathrm{d}t + \sigma(\Lambda(t_\delta))\mathrm{d}W(t), \qquad t > 0 \tag{1.15}$$

with the initial condition $Y(\theta) = \xi(\theta)$ for $\theta \in [-\tau, 0]$ and $\Lambda(0) = i_0 \in \mathbf{S}$, where, $t_\delta := \lfloor t/\delta \rfloor \delta$ with $\lfloor t/\delta \rfloor$ being the integer part of t/δ, and $Y_{k\delta} = \{Y_{k\delta}(\theta) : -\tau \leq \theta \leq 0\}$ is a $\mathscr{C}$-valued random variable defined as follows: for any $\theta \in [i\delta, (i+1)\delta]$, $i = -M, -(M-1), \cdots, -1$,

$$Y_{k\delta}(\theta) = Y((k+i)\delta) + \frac{\theta - i\delta}{\delta}\{Y((k+i+1)\delta) - Y((k+i)\delta)\}, \tag{1.16}$$

i.e., $Y_{k\delta}(\cdot)$ is the linear interpolation of $Y((k-M)\delta)$, $Y((k-(M-1))\delta), \cdots, Y((k-1)\delta), Y(k\delta)$. Keep in mind that the $\mathscr{C}$-valued random variables X_t and Y_{t_δ} in (1.8) and (1.15), respectively, are defined in a quite different way. In order to emphasize the initial condition $Y(\theta) = \xi(\theta)$ for $\theta \in [-\tau, 0]$ and $\Lambda(0) = i \in \mathbf{S}$, in some of the occasion, we shall write $Y(t; \xi, i)$ and $Y_{t_\delta}(\xi, i)$ in lieu of $Y(t)$ and Y_{t_δ}, respectively. For latter purpose, we extend the initial value $Y(\theta) = \xi(\theta), \theta \in [-\tau, 0]$, of (1.15) into the interval $[-\tau-1, -\tau)$ by setting $Y(\theta) = \xi(-\tau)$ for any $\theta \in [-\tau-1, -\tau)$. Moreover, the pair $(Y_{t_\delta}(\xi, i), \Lambda(t_\delta))$ enjoys the Markov property as Lemma 5.1 below shows. Let $P^{(\delta)}_{k\delta}((\xi, i), \cdot)$ stand for the transition kernel of $(Y_{k\delta}(\xi, i), \Lambda^i(k\delta))$.

To investigate the long-term behavior of $Y_{k\delta}$ defined by (1.16), besides (**A**), we further assume that there exists an $L_0 > 0$ such that

$$|b(\xi, i) - b(\eta, i)| \leq L_0 \|\xi - \eta\|_\infty, \qquad \xi, \eta \in \mathscr{C}, \quad i \in \mathbf{S}. \tag{1.17}$$

The theorem below shows that the discrete-time semigroup generated by the discretization of $(X_t(\xi, i), \Lambda^i(t))$ admits a unique invariant probability measure and is exponentially convergent to its equilibrium under the Wasserstein distance.

Theorem 1.4. Let the assumptions of Theorem 1.1 be satisfied and suppose further (1.17) holds. Then, there exist $\delta_0 \in (0, 1)$ and $\alpha > 0$ such that for any $k \geq 0$ and $\delta \in (0, \delta_0)$,

$$W_\rho(\nu_1 P^{(\delta)}_{k\delta}, \nu_2 P^{(\delta)}_{k\delta}) \leq c\Big(1 + \sum_{i\in\mathbf{S}} \int_{\mathscr{C}} \|\xi\|_\infty \nu_1(\mathrm{d}\xi, i) + \sum_{i\in\mathbf{S}} \int_{\mathscr{C}} \|\eta\|_\infty \nu_2(\mathrm{d}\eta, i)\Big)\mathrm{e}^{-\alpha k\delta}, \tag{1.18}$$

in which $\nu_1, \nu_2 \in \mathcal{P}_0$. Furthermore, (1.18) implies that $(Y_{k\delta}(\xi, i), \Lambda^i(k\delta))$, ascertained by (1.15) and (1.1), admits a unique invariant probability measure $\mu^{(\delta)} \in \mathcal{P}_0$ such that

$$W_\rho(\delta_{(\xi,i)} P^{(\delta)}_{k\delta}, \mu^{(\delta)} P^{(\delta)}_{k\delta}) \leq c\Big(1 + \|\xi\|_\infty + \sum_{i\in\mathbf{S}} \int_{\mathscr{C}} \|\eta\|_\infty \mu^{(\delta)}(\mathrm{d}\eta, i)\Big)\mathrm{e}^{-\alpha k\delta}.$$

Remark 1.3. By following the argument of [3, Theorem 3.2], it follows that

$$\lim_{\delta\to 0} W_\rho(\mu^{(\delta)},\mu)=0,$$

where $\mu\in\mathcal{P}_0$ is the invariant probability measure of $(X_t(\xi,i),\Lambda^i(t))$, determined by (1.2) and (1.1), and $\mu^{(\delta)}\in\mathcal{P}_0$ is the invariant probability measure of $(Y_{k\delta}(\xi,i),\Lambda^i(k\delta))$ solving (1.15) and (1.1).

1.4 Numerical Invariant Measures: Multiplicative Noises

In this subsection, we move forward to discuss the multiplicative noise case. For this setting, we further assume that there exists an $L_1>0$ such that

$$|b(\xi,i)-b(\eta,i)|^2\le L_1\Big(|\xi(0)-\eta(0)|^2+\int_{-\tau}^0|\xi(\theta)-\eta(\theta)|^2v(\mathrm{d}\theta)\Big) \tag{1.19}$$

for any $\xi,\eta\in\mathscr{C}$ and $i\in\mathbf{S}$. Consider the EM scheme corresponding to (1.8)

$$\mathrm{d}Y(t)=b(Y_{t_\delta},\Lambda(t_\delta))\mathrm{d}t+\sigma(Y_{t_\delta},\Lambda(t_\delta))\mathrm{d}W(t),\qquad t>0, \tag{1.20}$$

with the initial condition $Y(\theta)=\xi(\theta)$ for $\theta\in[-\tau,0]$ and $\Lambda(0)=i_0\in\mathbf{S}$, where Y_{t_δ} is defined exactly as in (1.16). Set

$$Q_3:=Q+\mathrm{diag}\Big(\alpha_1+4\mathrm{e}^{-\widehat{\alpha}\tau}\beta_1,\dots,\alpha_N+4\mathrm{e}^{-\widehat{\alpha}\tau}\beta_N\Big),$$

and

$$\eta_3:=-\max_{\gamma\in\mathrm{spec}(Q_3)}\mathrm{Re}(\gamma). \tag{1.21}$$

Concerning the multiplicative noise case, the time discretization of $(X_t(\xi,i),\Lambda^i(t))$, determined by (1.8) and (1.1), also shares the exponentially ergodic property when the stepsize is sufficiently small, which is presented below as another main result of this paper.

Theorem 1.5. Let (**H1**), (**H2**), and (1.19) hold and assume further $\eta_3>0$. Then, there exist $\delta_0\in(0,1)$ and $\alpha>0$ such that, for any $k\ge 0$ and $\delta\in(0,\delta_0)$,

$$W_\rho(\nu_1P_{k\delta},\nu_2P_{k\delta})\le c\Big(1+\sum_{i\in\mathbf{S}}\int_{\mathscr{C}}\|\xi\|_\infty\nu_1(\mathrm{d}\xi,i)+\sum_{i\in\mathbf{S}}\int_{\mathscr{C}}\|\eta\|_\infty\nu_2(\mathrm{d}\eta,i)\Big)\mathrm{e}^{-\alpha k\delta}, \tag{1.22}$$

where $\nu_1,\nu_2\in\mathcal{P}_0$. Furthermore, (1.22) implies that $(Y_{k\delta}(\xi,i),\Lambda^i(k\delta))$, determined by (1.20) and (1.1), admits a unique invariant probability measure $\mu^{(\delta)}\in\mathcal{P}_0$ such that

$$W_\rho(\delta_{(\xi,i)}P_{k\delta},\mu^{(\delta)}P_{k\delta})\le c\Big(1+\|\xi\|_\infty+\sum_{i\in\mathbf{S}}\int_{\mathscr{C}}\|\eta\|_\infty\mu^{(\delta)}(\mathrm{d}\eta,i)\Big)\mathrm{e}^{-\alpha k\delta}.$$

The remainder of this paper is organized as follows. Section 2 is devoted to the proof of Theorem 1.1; Section 3 is concerned with the proofs of Theorem 1.2 and Example 1.3; In Section 4, we aim to investigate the estimate on exponential functional of the discrete observation for the Markov chain involved and meanwhile finish the proof of Theorem 1.4; At length, we focus on the Markov property of time discretization of $(X_t(\xi,i),\Lambda^i(t))$ and complete the proof of Theorem 1.5.

2 Proof of Theorem 1.1

Let

$$\Omega_1 = \{\omega |\ \omega : [0,\infty) \to \mathbb{R}^m \text{ is continuous with } \omega(0) = 0\},$$

which is endowed with the locally uniform convergence topology and the Wiener measure $\mathbb{P}_1$ so that the coordinate process $W(t,\omega) := \omega(t)$, $t \geq 0$, is a standard m-dimensional Brownian motion. Set

$$\Omega_2 := \Big\{\omega |\ \omega : [0,\infty) \to \mathbf{S} \text{ is right continuous with left limit}\Big\},$$

endowed with Skorokhod topology and a probability measure $\mathbb{P}_2$ so that the coordinate process $\Lambda(t,\omega) = \omega(t)$, $t \geq 0$, is a continuous time Markov chain with Q-matrix (q_{ij}). Let

$$(\Omega, \mathscr{F}, \mathbb{P}) = (\Omega_1 \times \Omega_2, \mathscr{B}(\Omega_1) \times \mathscr{B}(\Omega_2), \mathbb{P}_1 \times \mathbb{P}_2).$$

Then, under $\mathbb{P} := \mathbb{P}_1 \times \mathbb{P}_2$, for $\omega = (\omega_1, \omega_2) \in \Omega$, $\omega_1(\cdot)$ is a Brownian motion, and $\omega_2(\cdot)$ is a continuous time Markov chain with Q-matrix (q_{ij}) on $\mathbf{S}$. Throughout this paper, we shall work on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ constructed above.

The lemma below shows that, under suitable assumptions, the functional solutions starting from different points will close in the L^2-norm sense to each other when time parameter goes to infinity.

Lemma 2.1. Under the assumptions of Theorem 1.1,

$$\mathbb{E}\|X_t(\xi,i) - X_t(\eta,i)\|_\infty^2 \leq c\,\|\xi - \eta\|_\infty^2 \mathrm{e}^{-\eta_1 t} \tag{2.1}$$

for any $\xi, \eta \in \mathscr{C}$ and $i \in \mathbf{S}$, where $\eta_1 > 0$ is defined in (1.4).

Proof. For fixed $\omega_2 \in \Omega_2$, consider the following SDE

$$\mathrm{d}X^{\omega_2}(t) = b(X_t^{\omega_2}, \Lambda^{\omega_2}(t))\mathrm{d}t + \sigma(\Lambda^{\omega_2}(t))\mathrm{d}\omega_1(t), \quad t > 0, \quad X_0^{\omega_2} = \xi \in \mathscr{C}, \quad \Lambda^{\omega_2}(0) = i \in \mathbf{S}.$$

Since $(\Lambda^{\omega_2}(s))_{s\in[0,t]}$ may own finite number of jumps, $t \mapsto \int_0^t \alpha_{\Lambda^{\omega_2}(s)}\mathrm{d}s$ need not to be differentiable. To overcome this drawback, let us introduce a smooth approximation of it. For any $\varepsilon \in (0,1)$, set

$$\alpha^\varepsilon_{\Lambda^{\omega_2}(t)} := \frac{1}{\varepsilon}\int_t^{t+\varepsilon} \alpha_{\Lambda^{\omega_2}(s)}\mathrm{d}s + \varepsilon t = \int_0^1 \alpha_{\Lambda^{\omega_2}(\varepsilon s+t)}\mathrm{d}s + \varepsilon t.$$

Plainly, $t \mapsto \alpha^\varepsilon_{\Lambda^{\omega_2}(t)}$ is continuous and $\alpha^\varepsilon_{\Lambda^{\omega_2}(t)} \to \alpha_{\Lambda^{\omega_2}(t)}$ as $\varepsilon \downarrow 0$ due to the right continuity of the path of $\Lambda^{\omega_2}(\cdot)$. As a consequence, $t \mapsto \int_0^t \alpha^\varepsilon_{\Lambda^{\omega_2}(r)}\mathrm{d}r$ is differentiable by the first fundamental theorem of calculus and $\int_0^t \alpha^\varepsilon_{\Lambda^{\omega_2}(r)}\mathrm{d}r \to \int_0^t \alpha_{\Lambda^{\omega_2}(r)}\mathrm{d}r$ as $\varepsilon \downarrow 0$ according to Lebesgue's dominated convergence theorem. Let

$$\Gamma^{\omega_2}(t) = X^{\omega_2}(t;\xi,i) - X^{\omega_2}(t;\eta,i). \tag{2.2}$$

Applying Itô's formula and taking (**A**) into account ensures that

$$
\begin{aligned}
\mathrm{e}^{-\int_0^t \alpha^{\varepsilon}_{\Lambda^{\omega_2}(s)}\mathrm{d}s}|\Gamma^{\omega_2}(t)|^2 &= |\Gamma^{\omega_2}(0)|^2 + \int_0^t \mathrm{e}^{-\int_0^s \alpha^{\varepsilon}_{\Lambda^{\omega_2}(r)}\mathrm{d}r}\Big\{ -\alpha^{\varepsilon}_{\Lambda^{\omega_2}(s)}|\Gamma^{\omega_2}(s)|^2 \\
&\qquad + 2\langle \Gamma^{\omega_2}(s), b(X_s^{\omega_2}(\xi,i),\Lambda^{\omega_2}(s)) - b(X_s^{\omega_2}(\eta,i),\Lambda^{\omega_2}(s))\rangle\Big\}\mathrm{d}s \\
&\le |\Gamma^{\omega_2}(0)|^2 + \Gamma_1^{\omega_2,\varepsilon}(t) + \int_0^t \beta_{\Lambda^{\omega_2}(s)}\mathrm{e}^{-\int_0^s \alpha^{\varepsilon}_{\Lambda^{\omega_2}(r)}\mathrm{d}r}\|\Gamma_s^{\omega_2}\|_\infty^2\mathrm{d}s,
\end{aligned}
\tag{2.3}
$$

where

$$
\Gamma_1^{\omega_2,\varepsilon}(t) := \int_0^t \mathrm{e}^{-\int_0^s \alpha^{\varepsilon}_{\Lambda^{\omega_2}(r)}\mathrm{d}r}|\alpha_{\Lambda^{\omega_2}(s)} - \alpha^{\varepsilon}_{\Lambda^{\omega_2}(s)}|\cdot|\Gamma^{\omega_2}(s)|^2\mathrm{d}s.
\tag{2.4}
$$

Due to the fact that

$$
\mathrm{e}^{-\int_0^t \alpha^{\varepsilon}_{\Lambda^{\omega_2}(s)}\mathrm{d}s}\|\Gamma_t^{\omega_2}\|_\infty^2 \le \mathrm{e}^{-\widehat{\alpha}\tau}\Big\{ \|\Gamma_0^{\omega_2}\|_\infty^2 + \sup_{(t-\tau)\vee 0\le s\le t}\Big(\mathrm{e}^{-\int_0^s \alpha^{\varepsilon}_{\Lambda^{\omega_2}(r)}\mathrm{d}r}|\Gamma^{\omega_2}(s)|^2\Big)\Big\},
\tag{2.5}
$$

where $\widehat{\alpha}$ is defined in (1.3), we therefore infer from (2.3) that

$$
\mathrm{e}^{-\int_0^t \alpha^{\varepsilon}_{\Lambda^{\omega_2}(s)}\mathrm{d}s}\|\Gamma_t^{\omega_2}\|_\infty^2 \le \mathrm{e}^{-\widehat{\alpha}\tau}\Big\{2\,\|\Gamma_0^{\omega_2}\|_\infty^2 + \Gamma_1^{\omega_2,\varepsilon}(t) + \int_0^t \beta_{\Lambda^{\omega_2}(s)}\mathrm{e}^{-\int_0^s \alpha^{\varepsilon}_{\Lambda^{\omega_2}(r)}\mathrm{d}r}\|\Gamma_s^{\omega_2}\|_\infty^2\mathrm{d}s\Big\}.
$$

Since $\alpha^{\varepsilon}_{\Lambda^{\omega_2}(s)} \to \alpha_{\Lambda^{\omega_2}(s)}$ so that $\Gamma_1^{\omega_2,\varepsilon}(t)\to 0$ as $\varepsilon\to 0$, by taking $\varepsilon\downarrow 0$ one has

$$
\mathrm{e}^{-\int_0^t \alpha_{\Lambda^{\omega_2}(s)}\mathrm{d}s}\|\Gamma_t^{\omega_2}\|_\infty^2 \le \mathrm{e}^{-\widehat{\alpha}\tau}\Big\{2\,\|\Gamma_0^{\omega_2}\|_\infty^2 + \int_0^t \beta_{\Lambda^{\omega_2}(s)}\mathrm{e}^{-\int_0^s \alpha_{\Lambda^{\omega_2}(r)}\mathrm{d}r}\|\Gamma_s^{\omega_2}\|_\infty^2\mathrm{d}s\Big\}.
$$

Thus, employing Gronwall's inequality followed by taking expectation w.r.t. $\mathbb{P}$ yields that

$$
\mathbb{E}\|X_t(\xi,i_0) - X(\eta,i_0)\|_\infty^2 \le 2\,\|\xi-\eta\|_\infty^2\mathbb{E}\,\mathrm{e}^{\int_0^t(\alpha_{\Lambda(s)}+\mathrm{e}^{-\widehat{\alpha}\tau}\beta_{\Lambda(s)})\mathrm{d}s}.
$$

Consequently, the desired assertion follows from [4, Proposition 4.1] at once. □

The following lemma reveals that the functional solution is uniformly bounded in the L^2-norm sense.

Lemma 2.2. Under the assumptions of Theorem 1.1,

$$
\sup_{t\ge 0}\mathbb{E}\|X_t(\xi,i)\|_\infty^2 \le c\,(1+\|\xi\|_\infty^2), \qquad (\xi,i)\in\mathscr{C}\times\mathbf{S}.
\tag{2.6}
$$

Proof. Analogously, we define

$$
\beta^{\varepsilon}_{\Lambda^{\omega_2}(t)} = \frac{1}{\varepsilon}\int_t^{t+\varepsilon}\beta_{\Lambda^{\omega_2}(s)}\mathrm{d}s + \varepsilon t = \int_0^1 \beta_{\Lambda^{\omega_2}(\varepsilon s+t)}\mathrm{d}s + \varepsilon t.
$$

By virtue of (**A**), for any $\gamma>0$, there is a $c_\gamma>0$ such that

$$2\langle\xi(0),b(\xi,i)\rangle+\|\sigma(i)\|_{\rm HS}^2\le c_\gamma+(\alpha_i+\gamma)|\xi(0)|^2+\beta_i\|\xi\|_\infty^2. \tag{2.7}$$

Employing Itô's formula and taking (2.7) into consideration provides that

$$\begin{aligned}\mathrm{e}^{-\int_0^t(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(s)})\mathrm{d}s}|X^{\omega_2}(t)|^2&\le|\xi(0)|^2+c_\gamma\int_0^t\mathrm{e}^{-\int_0^s(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\mathrm{d}s\\&\quad+\int_0^t\beta^\varepsilon_{\Lambda^{\omega_2}(s)}\mathrm{e}^{-\int_0^s(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\|X_s^{\omega_2}\|_\infty^2\mathrm{d}s+\Gamma_2^{\omega_2,\varepsilon}(t)+\Upsilon^{\omega_2,\varepsilon}(t),\end{aligned} \tag{2.8}$$

where

$$\begin{aligned}\Gamma_2^{\omega_2,\varepsilon}(t):=\int_0^t\mathrm{e}^{-\int_0^s(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\Big\{&|\alpha_{\Lambda^{\omega_2}(s)}-\alpha^\varepsilon_{\Lambda^{\omega_2}(s)}|\cdot|X^{\omega_2}(s)|^2\\&+|\beta_{\Lambda^{\omega_2}(s)}-\beta^\varepsilon_{\Lambda^{\omega_2}(s)}|\cdot\|X_s^{\omega_2}\|_\infty^2\Big\}\mathrm{d}s,\end{aligned} \tag{2.9}$$

and

$$\Upsilon^{\omega_2,\varepsilon}(t):=2\int_0^t\mathrm{e}^{-\int_0^s(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\langle X^{\omega_2}(s),\sigma(\Lambda^{\omega_2}(s))\mathrm{d}\omega_1(s)\rangle.$$

For any $0\le s\le t$ with $t-s\le\tau$ and $\kappa\in(0,1)$, exploiting BDG's inequality, we obtain that

$$\begin{aligned}\mathbb{E}_{\mathbb{P}_1}\Big(\sup_{s\le r\le t}\Upsilon^{\omega_2,\varepsilon}(r)\Big)&=\mathbb{E}_{\mathbb{P}_1}\Upsilon^{\omega_2,\varepsilon}(s)+\mathbb{E}_{\mathbb{P}_1}\Big(\sup_{s\le r\le t}(\Upsilon^{\omega_2,\varepsilon}(r)-\Upsilon^{\omega_2,\varepsilon}(s))\Big)\\&\le2\,\mathbb{E}_{\mathbb{P}_1}\Big(\sup_{s\le r\le t}\Big|\int_s^r\mathrm{e}^{-\int_0^u(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\langle X^{\omega_2}(u),\sigma(\Lambda^{\omega_2}(u))\mathrm{d}\omega_1(u)\rangle\Big|\Big)\\&\le c\,\mathbb{E}_{\mathbb{P}_1}\Big(\int_s^t\mathrm{e}^{-2\int_0^u(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}|X^{\omega_2}(u)|^2\cdot\|\sigma(\Lambda^{\omega_2}(u))\|_{\rm HS}^2\mathrm{d}u\Big)^{1/2}\\&\le c\,\mathrm{e}^{-\int_0^t(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\mathbb{E}_{\mathbb{P}_1}\Big(\|X_t^{\omega_2}\|_\infty^2\int_s^t\mathrm{e}^{2\int_u^t(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\mathrm{d}u\Big)^{1/2}\\&\le c\,\mathrm{e}^{-\int_0^t(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\mathbb{E}_{\mathbb{P}_1}\|X_t^{\omega_2}\|_\infty\\&\le\kappa\,\mathrm{e}^{\widehat{\alpha}\tau}\mathrm{e}^{-\int_0^t(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\mathbb{E}_{\mathbb{P}_1}\|X_t^{\omega_2}\|_\infty^2+c\,\mathrm{e}^{-\int_0^t(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}.\end{aligned} \tag{2.10}$$

This, together with (2.5) with Γ^{ω_2} being replaced accordingly by X^{ω_2}, and (2.8), leads to

$$\begin{aligned}\mathrm{e}^{-\int_0^t(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(s)})\mathrm{d}s}\mathbb{E}_{\mathbb{P}_1}\|X_t^{\omega_2}\|_\infty^2&\le\frac{\mathrm{e}^{-\widehat{\alpha}\tau}}{1-\kappa}\Big\{2\,\|\xi\|_\infty^2+c\int_0^t\mathrm{e}^{-\int_0^s(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\mathrm{d}s+c\,\mathrm{e}^{-\int_0^t(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\\&\quad+\int_0^t\beta^\varepsilon_{\Lambda^{\omega_2}(s)}\mathrm{e}^{-\int_0^s(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\mathbb{E}_{\mathbb{P}_1}\|X_s^{\omega_2}\|_\infty^2\mathrm{d}s+\mathbb{E}_{\mathbb{P}_1}\Gamma_2^{\omega_2,\varepsilon}(t)\Big\}.\end{aligned}$$

Then, the application of Gronwall's inequality yields that

$$
\begin{aligned}
&\mathrm{e}^{-\int_0^t(\gamma+\alpha^{\varepsilon}_{\Lambda^{\omega_2}(s)})\mathrm{d}s}\mathbb{E}_{\mathbb{P}_1}\|X_t^{\omega_2}\|_\infty^2\\
&\le c\Big\{\|\xi\|_\infty^2+\int_0^t\mathrm{e}^{-\int_0^s(\gamma+\alpha^{\varepsilon}_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\mathrm{d}s+\mathrm{e}^{-\int_0^t(\gamma+\alpha^{\varepsilon}_{\Lambda^{\omega_2}(r)})\mathrm{d}r}+\mathbb{E}_{\mathbb{P}_1}\Gamma_2^{\omega_2,\varepsilon}(t)\\
&\quad+\|\xi\|_\infty^2\int_0^t\Gamma_3^{\omega_2,\varepsilon}(s)\exp\Big(\int_s^t\Gamma_3^{\omega_2,\varepsilon}(r)\mathrm{d}r\Big)\mathrm{d}s\\
&\quad+\int_0^t\int_0^s\mathrm{e}^{-\int_0^u(\gamma+\alpha^{\varepsilon}_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\Gamma_3^{\omega_2,\varepsilon}(s)\exp\Big(\int_s^t\Gamma_3^{\omega_2,\varepsilon}(r)\mathrm{d}r\Big)\mathrm{d}u\mathrm{d}s\\
&\quad+\int_0^t\mathrm{e}^{-\int_0^s(\gamma+\alpha^{\varepsilon}_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\Gamma_3^{\omega_2,\varepsilon}(s)\exp\Big(\int_s^t\Gamma_3^{\omega_2,\varepsilon}(r)\mathrm{d}r\Big)\mathrm{d}s\\
&\quad+\int_0^t\mathbb{E}_{\mathbb{P}_1}\Gamma_2^{\omega_2,\varepsilon}(t)\Gamma_3^{\omega_2,\varepsilon}(s)\exp\Big(\int_s^t\Gamma_3^{\omega_2,\varepsilon}(r)\mathrm{d}r\Big)\mathrm{d}s\Big\},
\end{aligned}
\tag{2.11}
$$

where

$$
\Gamma_3^{\omega_2,\varepsilon}(t)=\frac{\mathrm{e}^{-\widehat{\alpha}\tau}\beta^{\varepsilon}_{\Lambda^{\omega_2}(t)}}{1-\kappa}.
$$

In the following, we go to estimate the terms in the right hand side of (2.11). By integration by parts, one has

$$
\begin{aligned}
&\|\xi\|_\infty^2\int_0^t\Gamma_3^{\omega_2,\varepsilon}(s)\exp\Big(\int_s^t\Gamma_3^{\omega_2,\varepsilon}(r)\mathrm{d}r\Big)\mathrm{d}s\\
&\quad+\int_0^t\int_0^s\mathrm{e}^{-\int_0^u(\gamma+\alpha^{\varepsilon}_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\Gamma_3^{\omega_2,\varepsilon}(s)\exp\Big(\int_s^t\Gamma_3^{\omega_2,\varepsilon}(r)\mathrm{d}r\Big)\mathrm{d}u\mathrm{d}s\\
&=\|\xi\|_\infty^2\Big(\exp\Big(\int_0^t\Gamma_3^{\omega_2,\varepsilon}(s)\mathrm{d}s\Big)-1\Big)\\
&\quad+\int_0^t\mathrm{e}^{-\int_0^s(\gamma+\alpha^{\varepsilon}_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\Big(\exp\Big(\int_s^t\Gamma_3^{\omega_2,\varepsilon}(r)\mathrm{d}r\Big)-1\Big)\mathrm{d}s\\
&\le\|\xi\|_\infty^2\exp\Big(\int_0^t\Gamma_3^{\omega_2,\varepsilon}(s)\mathrm{d}s\Big)\\
&\quad+\int_0^t\mathrm{e}^{-\int_0^s(\gamma+\alpha^{\varepsilon}_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\exp\Big(\int_s^t\Gamma_3^{\omega_2,\varepsilon}(r)\mathrm{d}r\Big)\mathrm{d}s.
\end{aligned}
\tag{2.12}
$$

On the other hand,

$$
\begin{aligned}
&\int_0^t\mathrm{e}^{-\int_0^s(\gamma+\alpha^{\varepsilon}_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\Gamma_3^{\omega_2,\varepsilon}(s)\exp\Big(\int_s^t\Gamma_3^{\omega_2,\varepsilon}(r)\mathrm{d}r\Big)\mathrm{d}s\\
&\le\frac{\mathrm{e}^{-\widehat{\alpha}\tau}\check{\beta}}{1-\kappa}\int_0^t\mathrm{e}^{-\int_0^s(\gamma+\alpha^{\varepsilon}_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\exp\Big(\int_s^t\Gamma_3^{\omega_2,\varepsilon}(r)\mathrm{d}r\Big)\mathrm{d}s.
\end{aligned}
\tag{2.13}
$$

Under (**A**), it is quite standard to show by using Hölder's inequality and BDG's inequality that

$$
\mathbb{E}_{\mathbb{P}_1}\Big(\sup_{0\le s\le t}\|X_s^{\omega_2}\|_\infty^2\Big)<\infty.
$$

So, the dominated convergence theorem implies that

$$\mathbb{E}_{\mathbb{P}_1}\Gamma_2^{\omega_2,\varepsilon}(t)+\int_0^t\mathbb{E}_{\mathbb{P}_1}\Gamma_2^{\omega_2,\varepsilon}(t)\Gamma_3^{\omega_2,\varepsilon}(s)\exp\Big(\int_s^t\Gamma_3^{\omega_2,\varepsilon}(r)\mathrm{d}r\Big)\mathrm{d}s\to 0,\qquad \text{as } \varepsilon\downarrow 0. \tag{2.14}$$

Whereafter, taking (2.12)-(2.14) into account and keeping in mind that $\alpha^\varepsilon_{\Lambda^{\omega_2}(t)}\to\alpha_{\Lambda^{\omega_2}(t)}$, $\beta^\varepsilon_{\Lambda^{\omega_2}(t)}\to\beta_{\Lambda^{\omega_2}(t)}$ as $\varepsilon\to 0$, we deduce from (2.11) that

$$\begin{aligned}\mathbb{E}\|X_t\|_\infty^2\le c\,\|\xi\|_\infty^2\,\mathbb{E}\exp\Big(\int_0^t\Big(\gamma+\alpha_{\Lambda(s)}+\frac{\mathrm{e}^{-\widehat{\alpha}\tau}}{1-\kappa}\beta_{\Lambda(s)}\Big)\mathrm{d}s\Big)\\ +c\Big(1+\int_0^t\mathbb{E}\exp\Big(\int_s^t\Big(\gamma+\alpha_{\Lambda(u)}+\frac{\mathrm{e}^{-\widehat{\alpha}\tau}}{1-\kappa}\beta_{\Lambda(u)}\Big)\mathrm{d}u\Big)\mathrm{d}s\Big).\end{aligned} \tag{2.15}$$

Accordingly, as $\eta_1>0$, by [4, Proposition 4.1], we obtain that for sufficiently small γ, $\kappa\in(0,1)$,

$$\begin{aligned}\sup_{t\ge0}\mathbb{E}\exp\Big(\int_0^t\Big(\gamma+\alpha_{\Lambda(s)}+\frac{\mathrm{e}^{-\widehat{\alpha}\tau}}{1-\kappa}\beta_{\Lambda(s)}\Big)\mathrm{d}s\Big)<\infty,\\ \sup_{t\ge0}\int_0^t\mathbb{E}\exp\Big(\int_s^t\Big(\gamma+\alpha_{\Lambda(u)}+\frac{\mathrm{e}^{-\widehat{\alpha}\tau}}{1-\kappa}\beta_{\Lambda(u)}\Big)\mathrm{d}u\Big)\mathrm{d}s<\infty,\end{aligned}$$

and hence (2.6) holds. □

We now in position to complete the

Proof of Theorem 1.1. For $\beta\in(0,1)$ to be determined, by Hölder's inequality, it follows that

$$\begin{aligned}W_\rho(\delta_{(\xi,i)}P_t,\delta_{(\eta,j)}P_t)&\le\mathbb{E}\{\|X_t(\xi,i)-X_t(\eta,j)\|_\infty+\mathbf{1}_{\{\Lambda^i(t)\neq\Lambda^j(t)\}}\}\\&=\mathbb{E}\{(\|X_t(\xi,i)-X_t(\eta,j)\|_\infty+\mathbf{1}_{\{\Lambda^i(t)\neq\Lambda^j(t)\}})\mathbf{1}_{\{T\le\beta t\}}\}\\&\quad+\mathbb{E}\{(\|X_t(\xi,i)-X_t(\eta,j)\|_\infty+\mathbf{1}_{\{\Lambda^i(t)\neq\Lambda^j(t)\}})\mathbf{1}_{\{T>\beta t\}}\}\\&\le\mathbb{E}(\mathbf{1}_{\{T\le\beta t\}}\mathbb{E}\{\|X_t(\xi,i)-X_t(\eta,j)\|_\infty\}|\mathscr{F}_T)\\&\quad+2\{1+\sqrt{2(\mathbb{E}\|X_t(\xi,i)\|_\infty^2+\mathbb{E}\|X_t(\eta,j)\|_\infty^2)}\}\sqrt{\mathbb{P}(T>\beta t)}\\&\le c\,\mathbb{E}(\mathbf{1}_{\{T\le\beta t\}}\|X_T(\xi,i)-X_T(\eta,j)\|_\infty\mathrm{e}^{-\frac{\eta_1}{2}(t-T)})\\&\quad+c\,(1+\|\xi\|_\infty+\|\eta\|_\infty)\mathrm{e}^{-\frac{1}{2}\theta\beta t}\\&\le c\,(1+\|\xi\|_\infty+\|\eta\|_\infty)(\mathrm{e}^{-\frac{1}{2}\theta\beta t}+\mathrm{e}^{-\frac{\eta_1}{2}(1-\beta)t}),\end{aligned}$$

where in the last two steps we have used (2.1) and (2.6). Optimizing over β in order to have $\theta\beta=\eta_1(1-\beta)$, i.e., $\beta=\frac{\eta_1}{\theta+\eta_1}$, leads to

$$W_\rho(\delta_{(\xi,i)}P_t,\delta_{(\eta,j)}P_t)\le c\,(1+\|\xi\|_\infty+\|\eta\|_\infty)\mathrm{e}^{-\frac{\theta\eta_1}{2(\theta+\eta_1)}t}. \tag{2.16}$$

Thus, substituting (2.16) into

$$W_\rho(\nu_1P_t,\nu_2P_t)\le\int W_\rho(\delta_{(\xi,i)}P_t,\delta_{(\eta,j)}P_t)\pi(\mathrm{d}\xi\times\mathrm{d}\{i\},\mathrm{d}\eta\times\mathrm{d}\{j\})$$

yields the desired assertion (1.6), where π is a coupling of ν_1 and ν_2.

Fix $\nu \in \mathcal{P}_0$ and observe that $(\nu P_n)_{n\geq 0}$ is a Cauchy sequence under the Wasserstein distance W_ρ due to (1.6) and that $\nu P_{n+1} = \nu P_n P_1$. So, by letting $n \to \infty$, there exists $\nu_\infty \in \mathcal{P}_0$ such that $\nu_\infty P_1 = \nu_\infty$. Set $\mu := \int_0^1 \nu_\infty P_s \mathrm{d}s$. It is easy to check $\mu \in \mathcal{P}_0$. In what follows, we claim that μ is indeed an invariant probability measure. In fact, for any $t > 0$, note that

$$\begin{aligned}\mu P_t &= \int_0^1 \nu_\infty P_{s+t}\mathrm{d}s = \int_t^{t+1} \nu_\infty P_s \mathrm{d}s\\ &= \int_t^0 \nu_\infty P_s \mathrm{d}s + \int_0^1 \nu_\infty P_s \mathrm{d}s + \int_0^t \nu_\infty P_1 P_s \mathrm{d}s\\ &= \mu,\end{aligned}$$

where in the last display we have used $\nu_\infty P_1 = \nu_\infty$. Let $\mu, \widetilde{\mu} \in \mathcal{P}_0$ both be the invariant probability measures of $(X_t(\xi, i), \Lambda^i(t))$. By the invariance, we deduce from (1.6) that

(2.17)
$$W_\rho(\mu, \widetilde{\mu}) = W_\rho(\mu P_t, \widetilde{\mu} P_t) \leq c\Big(1 + \sum_{i\in\mathbf{S}}\int_{\mathscr{C}} \|\xi\|_\infty \mu(\mathrm{d}\xi, i) + \sum_{i\in\mathbf{S}}\int_{\mathscr{C}} \|\eta\|_\infty \widetilde{\mu}(\mathrm{d}\eta, i)\Big)\mathrm{e}^{-\frac{\theta\eta_1}{2(\theta+\eta_1)}t}.$$

Consequently, the uniqueness of invariant measure can be obtained since the right hand side of (2.17) tends to zero as t goes to infinity. Finally, (1.7) follows by just taking $\nu_1 = \delta_{(\xi,i)}$ and $\nu_2 = \mu$ in (1.6).

3 Proof of Theorem 1.2

With the aid of Lemmas 3.1 and 3.2 below, the argument of Theorem 1.2 can be completed by repeating the procedure of Theorem 1.1.

Lemma 3.1. Under the assumptions of Theorem 1.2, it holds that

$$\mathbb{E}\|X_t(\xi, i) - X_t(\eta, i)\|_\infty^2 \leq c\,\|\xi - \eta\|_\infty^2 \mathrm{e}^{-\eta_2 t} \tag{3.1}$$

for any $\xi, \eta \in \mathscr{C}$ and $i \in \mathbf{S}$, where $\eta_2 > 0$ is defined in (1.9).

Proof. Fix $\omega_2 \in \Omega_2$ and let $(X^{\omega_2}(t))$ solve the SDE

$$\mathrm{d}X^{\omega_2}(t) = b(X_t^{\omega_2}, \Lambda^{\omega_2}(t))\mathrm{d}t + \sigma(X_t^{\omega_2}, \Lambda^{\omega_2}(t))\mathrm{d}\omega_1(t),\ t > 0,\ X_0^{\omega_2} = \xi \in \mathscr{C},\ \Lambda^{\omega_2}(0) = i \in \mathbf{S}.$$

Let $\Gamma^{\omega_2}(t)$ and $\Gamma_1^{\omega_2,\varepsilon}(t)$ be defined as in (2.2) and (2.4), respectively. By the Itô formula, we deduce from (**H1**) that

$$
\begin{aligned}
&\mathrm{e}^{-\int_0^t \alpha^{\varepsilon}_{\Lambda^{\omega_2}(s)}\mathrm{d}s}\mathbb{E}_{\mathbb{P}_1}|\Gamma^{\omega_2}(t)|^2\\
&=|\Gamma^{\omega_2}(0)|^2+\int_0^t \mathrm{e}^{-\int_0^s \alpha^{\varepsilon}_{\Lambda^{\omega_2}(r)}\mathrm{d}r}\mathbb{E}_{\mathbb{P}_1}\Big\{-\alpha^{\varepsilon}_{\Lambda^{\omega_2}(s)}|\Gamma^{\omega_2}(s)|^2\\
&\quad+2\langle\Gamma^{\omega_2}(s),b(X^{\omega_2}_s(\xi,i),\Lambda^{\omega_2}(s))-b(X^{\omega_2}_s(\eta,i),\Lambda^{\omega_2}(s))\rangle\\
&\quad+\|\sigma(X^{\omega_2}_s(\xi,i),\Lambda^{\omega_2}(s))-\sigma(X^{\omega_2}_s(\eta,i),\Lambda^{\omega_2}(s))\|^2_{\mathrm{HS}}\Big\}\mathrm{d}s \qquad (3.2)\\
&\le|\Gamma^{\omega_2}(0)|^2+\mathbb{E}_{\mathbb{P}_1}\Gamma^{\omega_2,\varepsilon}_1(t)+\int_0^t\beta_{\Lambda^{\omega_2}(s)}\mathrm{e}^{-\int_0^s \alpha^{\varepsilon}_{\Lambda^{\omega_2}(r)}\mathrm{d}r}\int_{-\tau}^0\mathbb{E}_{\mathbb{P}_1}|\Gamma^{\omega_2}(s+\theta)|^2v(\mathrm{d}\theta)\mathrm{d}s\\
&\le c\,\|\Gamma^{\omega_2}_0\|^2_\infty+\mathbb{E}_{\mathbb{P}_1}\Gamma^{\omega_2,\varepsilon}_1(t)+\int_0^t\Gamma^{\omega_2,\varepsilon}_4(s)\mathrm{e}^{-\int_0^s \alpha^{\varepsilon}_{\Lambda^{\omega_2}(r)}\mathrm{d}r}\mathbb{E}_{\mathbb{P}_1}|\Gamma^{\omega_2}(s)|^2\mathrm{d}s,
\end{aligned}
$$

where in the last step we have used the fact that

$$
\begin{aligned}
&\int_0^t\beta_{\Lambda^{\omega_2}(s)}\mathrm{e}^{-\int_0^s \alpha^{\varepsilon}_{\Lambda^{\omega_2}(r)}\mathrm{d}r}\int_{-\tau}^0\mathbb{E}_{\mathbb{P}_1}|\Gamma^{\omega_2}(s+\theta)|^2v(\mathrm{d}\theta)\mathrm{d}s\\
&=\int_{-\tau}^0\int_\theta^{t+\theta}\beta_{\Lambda^{\omega_2}(s-\theta)}\mathrm{e}^{-\int_0^{s-\theta} \alpha^{\varepsilon}_{\Lambda^{\omega_2}(r)}\mathrm{d}r}\mathbb{E}_{\mathbb{P}_1}|\Gamma^{\omega_2}(s)|^2\mathrm{d}sv(\mathrm{d}\theta)\\
&\le c\,\|\Gamma^{\omega_2}_0\|^2_\infty+\int_0^t\Gamma^{\omega_2,\varepsilon}_4(s)\mathrm{e}^{-\int_0^s \alpha^{\varepsilon}_{\Lambda^{\omega_2}(r)}\mathrm{d}r}\mathbb{E}_{\mathbb{P}_1}|\Gamma^{\omega_2}(s)|^2\mathrm{d}s
\end{aligned}
$$

with

$$\Gamma^{\omega_2,\varepsilon}_4(t):=\int_{-\tau}^0\beta_{\Lambda^{\omega_2}(t-\theta)}\mathrm{e}^{-\int_t^{t-\theta} \alpha^{\varepsilon}_{\Lambda^{\omega_2}(r)}\mathrm{d}r}v(\mathrm{d}\theta).$$

Then, applying Gronwall's inequality yields that

$$\mathbb{E}_{\mathbb{P}_1}|\Gamma^{\omega_2}(t)|^2\le\{c\,\|\Gamma^{\omega_2}_0\|^2_\infty+\mathbb{E}_{\mathbb{P}_1}\Gamma^{\omega_2,\varepsilon}_1(t)\}\mathrm{e}^{\int_0^t(\alpha^{\varepsilon}_{\Lambda^{\omega_2}(s)}+\Gamma^{\omega_2,\varepsilon}_4(s))\mathrm{d}s}. \qquad (3.3)$$

Letting $\varepsilon\to0$ followed by taking expectation w.r.t. $\mathbb{P}_2$ on both sides of (3.3), together with $\int_0^t\alpha^{\varepsilon}_{\Lambda^{\omega_2}(r)}\mathrm{d}r\to\int_0^t\alpha_{\Lambda^{\omega_2}(r)}\mathrm{d}r$ and $\mathbb{E}_{\mathbb{P}_1}\Gamma^{\omega_2,\varepsilon}_1(t)\to0$ as $\varepsilon\downarrow0$, gives that

$$\mathbb{E}|\Gamma(t)|^2\le c\,\|\Gamma_0\|^2_\infty\mathbb{E}\exp\Big(\int_0^t\Big(\alpha_{\Lambda(s)}+\int_{-\tau}^0\beta_{\Lambda(s-\theta)}\mathrm{e}^{-\int_s^{s-\theta}\alpha_{\Lambda(r)}\mathrm{d}r}v(\mathrm{d}\theta)\Big)\mathrm{d}s\Big), \qquad (3.4)$$

where $\Gamma(t):=X(t;\xi,i)-X(t;\eta,i)$. It is readily to see that

$$
\begin{aligned}
\int_0^t\int_{-\tau}^0\beta_{\Lambda(s-\theta)}\mathrm{e}^{-\int_s^{s-\theta}\alpha_{\Lambda(r)}\mathrm{d}r}v(\mathrm{d}\theta)\mathrm{d}s&\le\int_{-\tau}^0\mathrm{e}^{\widehat{\alpha}\theta}\int_{-\theta}^{t-\theta}\beta_{\Lambda(s)}\mathrm{d}sv(\mathrm{d}\theta)\\
&\le c+\int_{-\tau}^0\mathrm{e}^{\widehat{\alpha}\theta}v(\mathrm{d}\theta)\int_0^t\beta_{\Lambda(s)}\mathrm{d}s.
\end{aligned}
$$

Inserting this into (3.4), one has

$$\mathbb{E}|\Gamma(t)|^2 \le c\,\|\Gamma_0\|_\infty^2 \mathbb{E}\exp\Big(\int_0^t \Big(\alpha_{\Lambda(s)} + \int_{-\tau}^0 \mathrm{e}^{\widehat{\alpha}\theta} v(\mathrm{d}\theta)\beta_{\Lambda(s)}\Big)\mathrm{d}s\Big).$$

According to [4, Proposition 4.1], we derive from $\eta_2 > 0$ that

$$\mathbb{E}|\Gamma(t)|^2 \le c\,\mathrm{e}^{-\eta_2 t}\,\|\Gamma_0\|_\infty^2. \tag{3.5}$$

Next, for any $0 \le s \le t$, applying Itô's formula and BDG's inequality and making advantage of (**H1**) and (**H2**), we find that

$$\begin{aligned}\mathbb{E}\Big(\sup_{s\le r\le t}|\Gamma(r)|^2\Big) &\le \mathbb{E}|\Gamma(s)|^2 + (\breve{\alpha}+\breve{\beta})\int_{s-\tau}^t \mathbb{E}|\Gamma(r)|^2\mathrm{d}r \\ &\quad + 8\sqrt{2}\Big(\mathbb{E}\Big(\int_s^t |\Gamma(r)|^2\cdot\|\sigma(X_r(\xi,i),\Lambda(r)) - \sigma(X_r(\eta,i),\Lambda(r))\|_{\mathrm{HS}}^2\mathrm{d}r\Big)^{1/2}\Big) \\ &\le \mathbb{E}|\Gamma(s)|^2 + c\int_{s-\tau}^t \mathbb{E}|\Gamma(r)|^2\mathrm{d}r + \frac{1}{2}\mathbb{E}\Big(\sup_{s\le r\le t}|\Gamma(r)|^2\Big),\end{aligned}$$

which further implies that

$$\mathbb{E}\Big(\sup_{s\le r\le t}|\Gamma(r)|^2\Big) \le c\Big\{\mathbb{E}|\Gamma(s)|^2 + \int_{s-\tau}^t \mathbb{E}|\Gamma(r)|^2\mathrm{d}r\Big\}. \tag{3.6}$$

This leads to (3.1) by using (3.5) and noting that

$$\mathbb{E}\|\Gamma_t\|_\infty^2 = \mathbb{E}\Big(\sup_{t-\tau\le s\le t}|\Gamma(s)|^2\Big) \le \|\xi\|_\infty^2 + \mathbb{E}\Big(\sup_{(t-\tau)\vee 0\le s\le t}|\Gamma(s)|^2\Big).$$

□

Lemma 3.2. Under the assumptions of Theorem 1.2,

$$\mathbb{E}\|X_t(\xi,i)\|_\infty^2 \le c\,(1+\|\xi\|_\infty^2), \qquad (\xi,i)\in\mathscr{C}\times\mathbf{S}. \tag{3.7}$$

Proof. By virtue of (**H1**), for any $\gamma > 0$, there exists a $c_\gamma > 0$ such that

$$2\langle\xi(0),b(\xi,i)\rangle + \|\sigma(\xi,i)\|_{\mathrm{HS}}^2 \le c_\gamma + (\gamma+\alpha_i)|\xi(0)|^2 + (\gamma+\beta_i)\int_{-\tau}^0 |\xi(\theta)|^2 v(\mathrm{d}\theta) \tag{3.8}$$

holds. Next, following the argument to derive (3.2) and making use of (3.8), we infer that

$$\begin{aligned}\mathrm{e}^{-\int_0^t(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(s)})\mathrm{d}s}\mathbb{E}_{\mathbb{P}_1}|X^{\omega_2}(t)|^2 &\le c\,\|\xi\|_\infty^2 + \mathbb{E}_{\mathbb{P}_1}\Gamma_2^{\omega_2,\varepsilon}(t) + c\int_0^t \mathrm{e}^{-\int_0^s(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\mathrm{d}s \\ &\quad + \int_0^t \Gamma_5^{\omega_2,\varepsilon}(s)\mathrm{e}^{-\int_0^s(\delta+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\mathbb{E}_{\mathbb{P}_1}|X^{\omega_2}(s)|^2\mathrm{d}s,\end{aligned}$$

where $\Gamma_2^{\omega_2,\varepsilon}(t)$ is defined as in (2.9) with writing $\int_{[-\tau,0]}|X^{\omega_2}(s+\theta)|^2v(\mathrm{d}\theta)$ in lieu of $X_s^{\omega_2}$, and

$$\Gamma_5^{\omega_2,\varepsilon}(t):=\int_{-\tau}^0(\gamma+\beta^\varepsilon_{\Lambda^{\omega_2}(t-\theta)})\mathrm{e}^{-\int_t^{t-\theta}(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}v(\mathrm{d}\theta).$$

Subsequently, an application of Gronwall's inequality yields that

$$\begin{aligned}\mathbb{E}_{\mathbb{P}_1}|X^{\omega_2}(t)|^2\le c\,\|\xi\|_\infty^2+\mathbb{E}_{\mathbb{P}_1}\Gamma_2^{\omega_2,\varepsilon}(t)+c\int_0^t\mathrm{e}^{-\int_0^s(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\mathrm{d}s\\ +\int_0^t\Big(c\,\|\xi\|_\infty^2+\mathbb{E}_{\mathbb{P}_1}\Gamma_2^{\omega_2,\varepsilon}(s)+c\int_0^s\mathrm{e}^{-\int_0^u(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\mathrm{d}u\Big)\\ \times\Gamma_5^{\omega_2,\varepsilon}(s)\exp\Big(\int_s^t\Gamma_5^{\omega_2,\varepsilon}(r)\mathrm{d}r\Big)\mathrm{d}s.\end{aligned}$$

Thus, following the lines to derive (2.15), we arrive at

$$\mathbb{E}|X(t)|^2\le c\,\|\xi\|_\infty^2\mathbb{E}\,\mathrm{e}^{\int_0^t(\gamma+\alpha_{\Lambda(s)}+\Gamma_4(s))\mathrm{d}s}+c\int_0^t\mathbb{E}\mathrm{e}^{\int_s^t(\gamma+\alpha_{\Lambda(u)}+\Gamma_4(u))\mathrm{d}u}\mathrm{d}s, \tag{3.9}$$

where

$$\Gamma_4(t):=\int_{-\tau}^0(\gamma+\beta_{\Lambda(t-\theta)})\mathrm{e}^{-\int_t^{t-\theta}(\gamma+\alpha_{\Lambda(r)})\mathrm{d}r}v(\mathrm{d}\theta),\qquad t>0.$$

Plugging the fact that

$$\int_s^t\Gamma_4(r)\mathrm{d}r\le c+\int_{-\tau}^0\mathrm{e}^{\widehat{\alpha}\theta}v(\mathrm{d}\theta)\int_s^t(\gamma+\beta_{\Lambda^{\omega_2}(r)})\mathrm{d}r$$

into (3.9) means that

$$\begin{aligned}\mathbb{E}|X(t)|^2\le c\,\|\xi\|_\infty^2\mathbb{E}\exp\Big(\int_0^t\Big(C_\gamma+\alpha_{\Lambda(s)}+\int_{-\tau}^0\mathrm{e}^{\widehat{\alpha}\theta}v(\mathrm{d}\theta)\beta_{\Lambda(s)}\Big)\mathrm{d}s\Big)\\ +c\int_0^t\mathbb{E}\exp\Big(\int_s^t\Big(C_\gamma+\alpha_{\Lambda(r)}+\int_{-\tau}^0\mathrm{e}^{\widehat{\alpha}\theta}v(\mathrm{d}\theta)\beta_{\Lambda(r)}\Big)\mathrm{d}r\Big)\mathrm{d}s,\end{aligned}$$

where $C_\gamma:=\gamma\Big(1+\int_{-\tau}^0\mathrm{e}^{\widehat{\alpha}\theta}v(\mathrm{d}\theta)\Big)$. Thus, with the aid of [4, Proposition 4.1] and by choosing $\gamma>0$ such that $C_\gamma=\eta_2/2$, we obtain from $\eta_2>0$ that

$$\mathbb{E}|X(t)|^2\le c\,\|\xi\|_\infty^2\mathrm{e}^{-\eta_2t/2}+c\int_0^t\mathrm{e}^{-\frac{\eta_2}{2}s}\mathrm{d}s\le c\,(1+\|\xi\|_\infty^2). \tag{3.10}$$

Carrying out an analogous manner to derive (3.6), we have

$$\mathbb{E}\Big(\sup_{s\le r\le t}|X(r)|^2\Big)\le c\Big\{1+\|\xi\|_\infty^2+\mathbb{E}|X(s)|^2+\int_{s-\tau}^t\mathbb{E}|X(r)|^2\mathrm{d}r\Big\}. \tag{3.11}$$

Thereby, (3.7) is now available from (3.10) and (3.11). □

Proof of Example 1.3. (1.13) can be regarded as the interactions between the following path-dependent diffusion processes

$$\text{(3.12)}\quad \mathrm{d}X^{(i)}(t) = \{a_iX^{(i)}(t) + b_iX^{(i)}(t-1)\}\mathrm{d}t + \sigma_i\mathrm{d}W(t), \quad t>0, \quad X_0^{(i)} = \xi \in \mathscr{C}, \quad i=1,2.$$

The characteristic equation associated with the deterministic counterpart (i.e., $\sigma_i = 0$) of (3.12) is

$$\lambda_i - \int_{-1}^{0} \mathrm{e}^{\lambda_i s}\mu_i(\mathrm{d}s) =: \triangle_{\mu_i}(\lambda_i) = 0, \qquad i=1,2,$$

where $\mu_i(\cdot) := a_i\delta_0(\cdot) + b_i\delta_{-1}(\cdot)$, where $\delta_x(\cdot)$ signifies Dirac's delta measure or unit mass at the point x. By the variation-of-constants formula (see, e.g., [1, Theorem 1]), (3.12) can be expressed respectively as

$$X^{(i)}(t) = \Gamma_i(t)\xi(0) + b_i\int_{-1}^{0}\Gamma_i(t-1-s)\xi(s)\mathrm{d}s + \int_0^t \Gamma_i(t-s)\sigma_i\mathrm{d}W(s), \quad t>0, \quad i=1,2.$$

Herein, $\Gamma_i(t)$ is the solution to the delay equation

$$\text{(3.13)}\qquad \mathrm{d}Z^{(i)}(t) = \{a_iZ^{(i)}(t) + b_iZ^{(i)}(t-1)\}\mathrm{d}t, \quad t>0$$

with the initial value $Z^{(i)}(0)=1$ and $Z^{(i)}(\theta)=0, \theta\in[-1,0)$. In general, $\Gamma_i(t)$ is called the fundamental solution of (3.12) with $\sigma_i = 0$. It is readily to see that $\triangle_{\mu_1}(\lambda)=0$ has a unique positive root. Thus $\Gamma_0(t)\to\infty$ as $t\uparrow\infty$ so that $\mathbb{E}|X^{(0)}(t)|\to\infty$; see, e.g., [25]. Hence, $(X_t^{(0)})$ does not admit an invariant probability measure. The invariant probability measure of $(\Lambda(t))_{t\geq 0}$ is

$$\pi = (\pi_0, \pi_1) = \Big(\frac{\gamma}{1+\gamma}, \frac{1}{1+\gamma}\Big).$$

Observe that

$$\begin{aligned}|Q_2 - \lambda E| &= \begin{vmatrix} \alpha-1-\lambda & 1 \\ \gamma & \beta-\gamma-\lambda \end{vmatrix} \\ &= (\alpha-1-\lambda)(p\beta-\gamma-\lambda)-\gamma \\ &= \lambda^2 - (\alpha+\beta-1-\gamma)\lambda + \alpha\beta - (\alpha\gamma+\beta).\end{aligned}$$

As we know, the characteristic equation $|Q_2-\lambda E|=0$ has two negative roots, λ_1 and λ_2, if and only if

$$\begin{cases}\lambda_1+\lambda_2 = \alpha+\beta-1-\gamma < 0 \\ \lambda_1\lambda_2 = \alpha\beta-(\alpha\gamma+\beta) > 0.\end{cases}$$

Nevertheless, the inequalities above hold under (1.14).

4 Proof of Theorem 1.4

Before proving Theorem 1.4, we present an estimate on the exponential functional of the discrete-time observations of the Markov chain. This lemma plays a crucial role in the analyzing the long-time behavior of the time discretization for $(X_t(\xi, i_0), \Lambda(t))$ and is of interest by itself.

Lemma 4.1. Let $K : \mathbf{S} \to \mathbb{R}$, and $Q_K = Q + \mathrm{diag}(K_1, \cdots, K_N)$. Set

$$\eta_K = -\max_{\gamma \in \mathrm{spec}(Q_K)} \mathrm{Re}(\gamma).$$

Then there exist $\delta_0 \in (0, 1)$ and $c > 0$ such that, for any $\delta \in (0, \delta_0)$,

$$\mathbb{E}\, \mathrm{e}^{\int_0^t K_{\Lambda(s_\delta)} \mathrm{d}s} \le c\, \mathrm{e}^{-\eta_K t/2}, \quad \forall\, t > 0. \tag{4.1}$$

Proof. By Hölder's inequality, it follows that

$$\begin{aligned} \mathbb{E}\, \mathrm{e}^{\int_0^t K_{\Lambda(s_\delta)} \mathrm{d}s} &= \mathbb{E}\, \mathrm{e}^{\int_0^t K_{\Lambda(s)} \mathrm{d}s + \int_0^t (K_{\Lambda(s_\delta)} - K_{\Lambda(s)}) \mathrm{d}s} \\ &\le \Big(\mathbb{E}\, \mathrm{e}^{(1+\varepsilon) \int_0^t K_{\Lambda(s)} \mathrm{d}s} \Big)^{\frac{1}{1+\varepsilon}} \Big(\mathbb{E}\, \mathrm{e}^{\frac{1+\varepsilon}{\varepsilon} \int_0^t (K_{\Lambda(s_\delta)} - K_{\Lambda(s)}) \mathrm{d}s} \Big)^{\frac{\varepsilon}{1+\varepsilon}}, \qquad \varepsilon > 0. \end{aligned} \tag{4.2}$$

Observe from (1.1) that there exists $\delta_1 \in (0, 1)$ such that for any $\triangle \in (0, \delta_1)$,

$$\mathbb{P}(\Lambda(t + \triangle) = i | \Lambda(t) = i) = 1 + q_{ii} \triangle + o(\triangle), \tag{4.3}$$

and that

$$\mathbb{P}(\Lambda(t + \triangle) \ne i | \Lambda(t) = i) = \sum_{j \ne i} (q_{ij} \triangle + o(\triangle)) \le \max_{i \in \mathbf{S}} (-q_{ii}) \triangle + o(\triangle). \tag{4.4}$$

Utilizing Jensen's inequality and taking advantage of (4.3) and (4.4), we derive that for any

$\delta\in(0,\delta_1)$,

$$
\begin{aligned}
&\mathbb{E}\left(\mathrm{e}^{\frac{1+\varepsilon}{\varepsilon}\int_{i\delta}^{(i+1)\delta\wedge t}(K_{\Lambda(i\delta)}-K_{\Lambda(s)})\mathrm{d}s}\Big|\Lambda(i\delta)\right)\\
&\le \frac{1}{(i+1)\delta\wedge t-i\delta}\int_{i\delta}^{(i+1)\delta\wedge t}\mathbb{E}\left(\mathrm{e}^{\frac{1+\varepsilon}{\varepsilon}((i+1)\delta\wedge t-i\delta)(K_{\Lambda(i\delta)}-K_{\Lambda(s)})}\Big|\Lambda(i\delta)\right)\mathrm{d}s\\
&=\frac{\sum_{j\in\mathbf{S}}\mathbf{1}_{\{\Lambda(i\delta)=j\}}}{(i+1)\delta\wedge t-i\delta}\int_{i\delta}^{(i+1)\delta\wedge t}\mathbb{E}\left(\mathrm{e}^{\frac{1+\varepsilon}{\varepsilon}((i+1)\delta\wedge t-i\delta)(K_j-K_{\Lambda(s)})}\Big|\Lambda(i\delta)=j\right)\mathrm{d}s\\
&=\frac{\sum_{j\in\mathbf{S}}\mathbf{1}_{\{\Lambda(i\delta)=j\}}}{(i+1)\delta\wedge t-i\delta}\int_{i\delta}^{(i+1)\delta\wedge t}\mathbb{E}\left(\mathbf{1}_{\{\Lambda(s)=j\}}|\Lambda(i\delta)=j\right)\mathrm{d}s\\
&\quad+\frac{\sum_{j\in\mathbf{S}}\mathbf{1}_{\{\Lambda(i\delta)=j\}}}{(i+1)\delta\wedge t-i\delta}\int_{i\delta}^{(i+1)\delta\wedge t}\mathbb{E}\left(\mathrm{e}^{\frac{1+\varepsilon}{\varepsilon}((i+1)\delta\wedge t-i\delta)(K_j-K_{\Lambda(s)})}\mathbf{1}_{\{\Lambda(s)\neq j\}}\Big|\Lambda(i\delta)=j\right)\mathrm{d}s\\
&\le\frac{\sum_{j\in\mathbf{S}}\mathbf{1}_{\{\Lambda(i\delta)=j\}}}{(i+1)\delta\wedge t-i\delta}\int_{i\delta}^{(i+1)\delta\wedge t}\mathbb{E}\left(\mathbf{1}_{\{\Lambda(s)=j\}}|\Lambda(i\delta)=j\right)\mathrm{d}s\\
&\quad+\mathrm{e}^{\frac{2(1+\varepsilon)\check{K}\delta}{\varepsilon}}\frac{\sum_{j\in\mathbf{S}}\mathbf{1}_{\{\Lambda(i\delta)=j\}}}{(i+1)\delta\wedge t-i\delta}\int_{i\delta}^{(i+1)\delta\wedge t}\mathbb{P}\left(\mathbf{1}_{\{\Lambda(s)\neq j\}}|\Lambda(i\delta)=j\right)\mathrm{d}s\\
&\le\frac{\sum_{j\in\mathbf{S}}\mathbf{1}_{\{\Lambda(i\delta)=j\}}}{(i+1)\delta\wedge t-i\delta}\int_{i\delta}^{(i+1)\delta\wedge t}(1+q_{jj}(s-i\delta)+o(s-i\delta))\mathrm{d}s\\
&\quad+\mathrm{e}^{\frac{2(1+\varepsilon)\check{K}\delta}{\varepsilon}}\frac{\sum_{j\in\mathbf{S}}\mathbf{1}_{\{\Lambda(i\delta)=j\}}}{(i+1)\delta\wedge t-i\delta}\int_{i\delta}^{(i+1)\delta\wedge t}\Big(\max_{i\in\mathbf{S}}(-q_{ii})(s-i\delta)+o(s-i\delta)\Big)\mathrm{d}s\\
&\le 1+\frac{\max_{i\in\mathbf{S}}(-q_{ii})}{2}\,\delta\,\mathrm{e}^{\frac{2(1+\varepsilon)\check{K}\delta}{\varepsilon}}+o(\delta),
\end{aligned}
\tag{4.5}
$$

where $\check{K}:=\max_{i\in\mathbf{S}}|K_i|$. By the property of conditional expectation, we deduce from (4.5) that

$$
\begin{aligned}
&\mathbb{E}\,\mathrm{e}^{\frac{1+\varepsilon}{\varepsilon}\int_0^t(K_{\Lambda(s_\delta)}-K_{\Lambda(s)})\mathrm{d}s}\\
&=\mathbb{E}\,\mathrm{e}^{\frac{1+\varepsilon}{\varepsilon}\sum_{i=0}^{\lfloor t/\delta\rfloor}\int_{i\delta}^{(i+1)\delta\wedge t}(K_{\Lambda(i\delta)}-K_{\Lambda(s)})\mathrm{d}s}\\
&=\mathbb{E}\Big(\mathbb{E}\left(\mathrm{e}^{\frac{1+\varepsilon}{\varepsilon}\sum_{i=0}^{\lfloor t/\delta\rfloor}\int_{i\delta}^{(i+1)\delta\wedge t}(K_{\Lambda(i\delta)}-K_{\Lambda(s)})\mathrm{d}s}\Big|\Lambda(t_\delta)\right)\Big)\\
&=\mathbb{E}\Big(\mathrm{e}^{\frac{1+\varepsilon}{\varepsilon}\sum_{i=0}^{\lfloor t/\delta\rfloor-1}\int_{i\delta}^{(i+1)\delta}(K_{\Lambda(i\delta)}-K_{\Lambda(s)})\mathrm{d}s}\mathbb{E}\left(\mathrm{e}^{\frac{1+\varepsilon}{\varepsilon}\int_{t_\delta}^{(t_\delta+\delta)\wedge t}(K_{\Lambda(t_\delta)}-K_{\Lambda(s)})\mathrm{d}s}\Big|\Lambda(t_\delta)\right)\Big)\\
&\le\Big(1+\frac{\max_{i\in\mathbf{S}}(-q_{ii})}{2}\,\delta\,\mathrm{e}^{\frac{2(1+\varepsilon)\check{K}\delta}{\varepsilon}}+o(\delta)\Big)\mathbb{E}\Big(\mathrm{e}^{\frac{1+\varepsilon}{\varepsilon}\sum_{i=0}^{\lfloor t/\delta\rfloor-1}\int_{i\delta}^{(i+1)\delta}(K_{\Lambda(i\delta)}-K_{\Lambda(s)})\mathrm{d}s}\Big)\\
&\le\Big(1+\frac{\max_{i\in\mathbf{S}}(-q_{ii})}{2}\,\delta\,\mathrm{e}^{\frac{2(1+\varepsilon)\check{K}\delta}{\varepsilon}}+o(\delta)\Big)^{\lfloor t/\delta\rfloor+1},\qquad\delta\in(0,\delta_1),
\end{aligned}
\tag{4.6}
$$

where $t_\delta:=\lfloor t/\delta\rfloor\delta$. For any $c_1,c_2>0$,

$$\lim_{\delta\to0}\frac{1}{\delta}\ln(1+c_1\delta\,\mathrm{e}^{c_2\delta}+o(\delta))=c_1.$$

So, there exists $\delta_2 = \delta_2(c_1, c_2) \in (0, 1)$ such that

$$\frac{1}{\delta}\ln(1 + c_1\delta\,\mathrm{e}^{c_2\delta}) \le 2\,c_1, \qquad \delta \in (0, \delta_2). \tag{4.7}$$

Note that δ_2 depends on c_2, and δ_2 decreases as c_2 increasing.

According to (4.7), there exists $\delta_2 = \delta_2(\varepsilon)$ so that for any $\delta \in (0, \delta_1 \wedge \delta_2)$,

$$\begin{aligned}
&\Big(1 + \frac{\max_{i\in\mathbf{S}}(-q_{ii})}{2}\,\delta\,\mathrm{e}^{\frac{2(1+\varepsilon)\check{K}\delta}{\varepsilon}} + o(\delta)\Big)^{\frac{\varepsilon(\lfloor t/\delta\rfloor+1)}{1+\varepsilon}}\\
&= \exp\Big(\frac{\varepsilon(\lfloor t/\delta\rfloor + 1)}{1+\varepsilon}\ln\Big(1 + \frac{\max_{i\in\mathbf{S}}(-q_{ii})}{2}\,\delta\,\mathrm{e}^{\frac{2(1+\varepsilon)\check{K}\delta}{\varepsilon}} + o(\delta)\Big)\Big)\\
&\le \exp\Big(\frac{\varepsilon(t+\delta)}{1+\varepsilon}\frac{1}{\delta}\ln\Big(1 + \frac{\max_{i\in\mathbf{S}}(-q_{ii})}{2}\,\delta\,\mathrm{e}^{\frac{2(1+\varepsilon)\check{K}\delta}{\varepsilon}} + o(\delta)\Big)\Big)\\
&\le \exp\Big(\frac{\varepsilon}{1+\varepsilon}\max_{i\in\mathbf{S}}(-q_{ii})\Big)\exp\Big(\frac{\varepsilon}{1+\varepsilon}\max_{i\in\mathbf{S}}(-q_{ii})t\Big).
\end{aligned} \tag{4.8}$$

Taking (4.6) and (4.8) into consideration, we deduce from (4.2) that

$$\mathbb{E}\,\mathrm{e}^{\int_0^t K_{\Lambda(s_\delta)}\mathrm{d}s} \le \exp\Big(\frac{\varepsilon}{1+\varepsilon}\max_{i\in\mathbf{S}}(-q_{ii})\Big)\Big(\mathbb{E}\,\mathrm{e}^{(1+\varepsilon)\int_0^t K_{\Lambda(s)}\mathrm{d}s}\Big)^{\frac{1}{1+\varepsilon}}\exp\Big(\frac{\varepsilon}{1+\varepsilon}\max_{i\in\mathbf{S}}(-q_{ii})t\Big). \tag{4.9}$$

By virtue of [4, Theorem 1.5 & Proposition 4.1], there exist $\varepsilon_0 \in (0, 1)$ sufficiently small and $c > 0$ such that

$$\mathbb{E}\,\mathrm{e}^{(1+\varepsilon)\int_0^t K_{\Lambda(s)}\mathrm{d}s} \le c\,\mathrm{e}^{-2\eta_K t/3}, \qquad \varepsilon \in (0, \varepsilon_0),\ t > 0.$$

Inserting this into (4.9) yields that

$$\mathbb{E}\,\mathrm{e}^{\int_0^t K_{\Lambda(s_\delta)}\mathrm{d}s} \le c^{\frac{1}{1+\varepsilon}}\exp\Big(\frac{\varepsilon}{1+\varepsilon}\max_{i\in\mathbf{S}}(-q_{ii})\Big)\exp\Big(-\frac{2\eta_K/3 - \max_{i\in\mathbf{S}}(-q_{ii})\varepsilon}{1+\varepsilon}t\Big), \quad t > 0.$$

Thus, the desired assertion follows by taking first ε small enough and then $\delta \in (0, \delta_1 \wedge \delta_2(\varepsilon))$. □

Remark 4.1. The crucial point of lemma 4.1 is that the choice of δ_0 is independent of time t. Otherwise, it is easy to obtain a similar estimate for a time-dependent δ_0 by using the dominated convergence theorem. Indeed,

$$\lim_{\delta\to 0}\mathbb{E}\mathrm{e}^{\int_0^t K_{\Lambda(s(\delta))}\mathrm{d}s} = \mathbb{E}\mathrm{e}^{\int_0^t K_{\Lambda(s-)}\mathrm{d}s} = \mathbb{E}\mathrm{e}^{\int_0^t K_{\Lambda(s)}\mathrm{d}s}.$$

Then, applying [4, Theorem 1.5 & Proposition 4.1] will yield the estimate for a given time t.

Next, we provide two crucial lemmas on distance between the $\mathscr{C}$-valued stochastic processes Y_{t_δ} starting from different points and its uniform boundedness under some appropriate assumptions.

Lemma 4.2. Let the assumptions of Theorem 1.1 be satisfied and suppose further (1.17) holds. Then, there exist $\delta_0 \in (0, 1)$ and $\alpha > 0$ such that

$$\mathbb{E}\|Y_{t_\delta}(\xi, i) - Y_{t_\delta}(\eta, i)\|_\infty^2 \le c\,\mathrm{e}^{-\alpha t}\|\xi - \eta\|_\infty^2, \quad t \ge \tau + 1,\ \ \delta \in (0, \delta_0) \tag{4.10}$$

for any $\xi, \eta \in \mathscr{C}$ and $i \in \mathbf{S}$.

Proof. Hereinafter, we assume that $t \geq \tau+1$. Fix $\omega_2 \in \Omega_2$ and let $(Y^{\omega_2}(t))$ solve the following SDE

$$\mathrm{d}Y^{\omega_2}(t) = b(Y^{\omega_2}_{t_\delta}, \Lambda^{\omega_2}(t_\delta))\mathrm{d}t + \sigma(\Lambda^{\omega_2}(t_\delta))\mathrm{d}\omega_1(t)$$

with the initial value $Y^{\omega_2}(s) = \xi(s), s \in [-\tau, 0]$, and $\Lambda^{\omega_2}(0) = i \in \mathbf{S}$. For notational simplicity, set

$$\Upsilon^{\omega_2}(t) := Y^{\omega_2}(t; \xi, i) - Y^{\omega_2}(t; \eta, i). \tag{4.11}$$

First of all, we claim that

$$\begin{aligned}\mathrm{e}^{-\int_0^t \alpha_{\Lambda^{\omega_2}(s_\delta)}\mathrm{d}s}|\Upsilon^{\omega_2}(t)|^2 &= |\Upsilon^{\omega_2}(0)|^2 + \int_0^t \mathrm{e}^{-\int_0^s \alpha_{\Lambda^{\omega_2}(r_\delta)}\mathrm{d}r}\Big\{ - \alpha_{\Lambda^{\omega_2}(s_\delta)}|\Upsilon^{\omega_2}(s)|^2 \\ &\quad + 2\langle \Upsilon^{\omega_2}(s), b(Y^{\omega_2}_{s_\delta}(\xi, i), \Lambda^{\omega_2}(s_\delta)) - b(Y^{\omega_2}_{s_\delta}(\eta, i), \Lambda^{\omega_2}(s_\delta))\rangle\Big\}\mathrm{d}s.\end{aligned} \tag{4.12}$$

For any $t \in (0, \delta)$, by Itô's formula, we have

$$\begin{aligned}\mathrm{e}^{-\int_0^t \alpha_{\Lambda^{\omega_2}(s_\delta)}\mathrm{d}s}|\Upsilon^{\omega_2}(t)|^2 &= \mathrm{e}^{-\alpha_{\Lambda^{\omega_2}(0)}t}|\Upsilon^{\omega_2}(t)|^2 \\ &= |\Upsilon^{\omega_2}(0)|^2 + \int_0^t \mathrm{e}^{-\alpha_{\Lambda^{\omega_2}(0)}s}\Big\{ - \alpha_{\Lambda^{\omega_2}(0)}|\Upsilon^{\omega_2}(s)|^2 \\ &\quad + 2\langle \Upsilon^{\omega_2}(s), b(Y^{\omega_2}_0(\xi, i), \Lambda^{\omega_2}(0)) - b(Y^{\omega_2}_0(\eta, i), \Lambda^{\omega_2}(0))\rangle\Big\}\mathrm{d}s \\ &= |\Upsilon^{\omega_2}(0)|^2 + \int_0^t \mathrm{e}^{-\int_0^s \alpha_{\Lambda^{\omega_2}(r_\delta)}\mathrm{d}r}\Big\{ - \alpha_{\Lambda^{\omega_2}(s_\delta)}|\Upsilon^{\omega_2}(s)|^2 \\ &\quad + 2\langle \Upsilon^{\omega_2}(s), b(Y^{\omega_2}_{s_\delta}(\xi, i), \Lambda^{\omega_2}(s_\delta)) - b(Y^{\omega_2}_{s_\delta}(\eta, i), \Lambda^{\omega_2}(s_\delta))\rangle\Big\}\mathrm{d}s.\end{aligned}$$

Accordingly, (4.12) holds for any $t \in [0, \delta]$. Next, we assume that (4.12) is true for any $t \in [(k-1)\delta, k\delta)$. For any $t \in [k\delta, (k+1)\delta)$, Itô's formula yields that

$$\begin{aligned}\mathrm{e}^{-\alpha_{\Lambda^{\omega_2}(k\delta)}(t-k\delta)}|\Upsilon^{\omega_2}(t)|^2 &= |\Upsilon^{\omega_2}(k\delta)|^2 + \int_{k\delta}^t \mathrm{e}^{-\alpha_{\Lambda^{\omega_2}(k\delta)}(s-k\delta)}\Big\{ - \alpha_{\Lambda^{\omega_2}(k\delta)}|\Upsilon^{\omega_2}(s)|^2\mathrm{d}s \\ &\quad + 2\langle \Upsilon^{\omega_2}(s), b(Y^{\omega_2}_{k\delta}(\xi, i), \Lambda^{\omega_2}(k\delta)) - b(Y^{\omega_2}_{k\delta}(\eta, i), \Lambda^{\omega_2}(k\delta))\rangle\Big\}\mathrm{d}s.\end{aligned}$$

Multiplying both sides by $\mathrm{e}^{-\int_0^{k\delta}(\gamma+\alpha_{\Lambda^{\omega_2}(s_\delta)})\mathrm{d}s}$ and applying (4.12) with $t = k\delta$ leads to

$$\begin{aligned}\mathrm{e}^{-\int_0^t \alpha_{\Lambda^{\omega_2}(s_\delta)}\mathrm{d}s}|\Upsilon^{\omega_2}(t)|^2 &= \mathrm{e}^{-\int_0^{k\delta}(\gamma+\alpha_{\Lambda^{\omega_2}(s_\delta)})\mathrm{d}s}|\Upsilon^{\omega_2}(k\delta)|^2 + \int_{k\delta}^t \mathrm{e}^{-\int_0^s \alpha_{\Lambda^{\omega_2}(r_\delta)}\mathrm{d}r}\Big\{ - \alpha_{\Lambda^{\omega_2}(s_\delta)}|\Upsilon^{\omega_2}(s)|^2\mathrm{d}s \\ &\quad + 2\langle \Upsilon^{\omega_2}(s), b(Y^{\omega_2}_{s_\delta}(\xi, i), \Lambda^{\omega_2}(s_\delta)) - b(Y^{\omega_2}_{s_\delta}(\eta, i), \Lambda^{\omega_2}(s_\delta))\rangle\Big\}\mathrm{d}s \\ &= |\Upsilon^{\omega_2}(0)|^2 + \int_0^t \mathrm{e}^{-\int_0^s \alpha_{\Lambda^{\omega_2}(r_\delta)}\mathrm{d}r}\Big\{ - \alpha_{\Lambda^{\omega_2}(s_\delta)}|\Upsilon^{\omega_2}(s)|^2 \\ &\quad + 2\langle \Upsilon^{\omega_2}(s), b(Y^{\omega_2}_{s_\delta}(\xi, i), \Lambda^{\omega_2}(s_\delta)) - b(Y^{\omega_2}_{s_\delta}(\eta, i), \Lambda^{\omega_2}(s_\delta))\rangle\Big\}\mathrm{d}s.\end{aligned}$$

Thereby, (4.12) follows immediately. It is readily to see from (1.17) that

$$
\begin{aligned}
|\Upsilon^{\omega_2}(t)-\Upsilon^{\omega_2}(t_\delta)| &= |b(Y^{\omega_2}_{t_\delta}(\xi,i),\Lambda^{\omega_2}(t_\delta))-b(Y^{\omega_2}_{t_\delta}(\eta,i),\Lambda^{\omega_2}(t_\delta))|\delta \\
&\le L_0\|\Upsilon^{\omega_2}_{t_\delta}\|_\infty\delta.
\end{aligned}\tag{4.13}
$$

By virtue of (4.12) and (**A**), it follows that

$$
\begin{aligned}
&\mathrm{e}^{-\int_0^t\alpha_{\Lambda^{\omega_2}(s_\delta)}\mathrm{d}s}|\Upsilon^{\omega_2}(t)|^2\\
&=|\Upsilon^{\omega_2}(0)|^2+\int_0^t\mathrm{e}^{-\int_0^s\alpha_{\Lambda^{\omega_2}(r_\delta)}\mathrm{d}r}\Big\{-\alpha_{\Lambda^{\omega_2}(s_\delta)}|\Upsilon^{\omega_2}(s)|^2\\
&\quad+2\langle\Upsilon^{\omega_2}(s),b(Y^{\omega_2}_{s_\delta}(\xi,i),\Lambda^{\omega_2}(s_\delta))-b(Y^{\omega_2}_{s_\delta}(\eta,i),\Lambda^{\omega_2}(s_\delta))\rangle\Big\}\mathrm{d}s\\
&\le|\Upsilon^{\omega_2}(0)|^2+\int_0^t\mathrm{e}^{-\int_0^s\alpha_{\Lambda^{\omega_2}(r_\delta)}\mathrm{d}r}\Big\{\alpha_{\Lambda^{\omega_2}(s_\delta)}(|\Upsilon^{\omega_2}(s_\delta)|^2-|\Upsilon^{\omega_2}(s)|^2)+\beta_{\Lambda^{\omega_2}(s_\delta)}\|\Upsilon^{\omega_2}_{s_\delta}\|^2_\infty\\
&\quad+2\langle\Upsilon^{\omega_2}(s)-\Upsilon^{\omega_2}(s_\delta),b(Y^{\omega_2}_{s_\delta}(\xi,i),\Lambda^{\omega_2}(s_\delta))-b(Y^{\omega_2}_{s_\delta}(\eta,i),\Lambda^{\omega_2}(s_\delta))\rangle\Big\}\mathrm{d}s\\
&\le|\Upsilon^{\omega_2}(0)|^2+\int_0^t\mathrm{e}^{-\int_0^s\alpha_{\Lambda^{\omega_2}(r_\delta)}\mathrm{d}r}\Big\{\frac{1+2\check\alpha}{\sqrt\delta}|\Upsilon^{\omega_2}(s)-\Upsilon^{\omega_2}(s_\delta)|^2\\
&\quad+(\check\alpha\sqrt\delta+L_0^2\sqrt\delta+\beta_{\Lambda^{\omega_2}(s_\delta)})\|\Upsilon^{\omega_2}_{s_\delta}\|^2_\infty\Big\}\mathrm{d}s\\
&\le|\Upsilon^{\omega_2}(0)|^2+\int_0^t(\rho\sqrt\delta+\beta_{\Lambda^{\omega_2}(s_\delta)})\mathrm{e}^{-\int_0^s\alpha_{\Lambda^{\omega_2}(r_\delta)}\mathrm{d}r}\|\Upsilon^{\omega_2}_{s_\delta}\|^2_\infty\mathrm{d}s,
\end{aligned}\tag{4.14}
$$

where $\rho:=2(1+\check a)L_0^2+\check\alpha$, and in the penultimate display we have used (4.13). Observe that

$$
\begin{aligned}
\Pi^{\omega_2}(t):&=\mathrm{e}^{-\int_0^t\alpha_{\Lambda^{\omega_2}(s_\delta)}\mathrm{d}s}\Big(\sup_{t-\tau-\delta\le s\le t}|\Upsilon^{\omega_2}(s)|^2\Big)\\
&\le\mathrm{e}^{-\widehat\alpha(\tau+\delta)}\Big(\sup_{t-\tau-\delta\le s\le t}\Big(\mathrm{e}^{-\int_0^s\alpha_{\Lambda^{\omega_2}(r_\delta)}\mathrm{d}r}|\Upsilon^{\omega_2}(s)|^2\Big)\Big)
\end{aligned}
$$

and that

$$
\|\Upsilon^{\omega_2}_{t_\delta}\|_\infty=\sup_{-\tau\le\theta\le0}|\Upsilon^{\omega_2}_{t_\delta}(\theta;\xi,i)|=\max_{-N\le i\le0}|\Upsilon^{\omega_2}(t_\delta+i\delta;\xi,i)|\le\sup_{t-\tau-\delta\le s\le t}|\Upsilon^{\omega_2}(s)|,
$$

by (1.16) and $Y(t)=\xi(-\tau)$ for any $t\in[-\tau-1,-\tau)$. We therefore obtain from (4.14) that

$$
\Pi^{\omega_2}(t)\le c\,\|\Upsilon^{\omega_2}_0\|^2_\infty+\mathrm{e}^{-\widehat\alpha(\tau+\delta)}\int_0^t(\rho\sqrt\delta+\beta_{\Lambda^{\omega_2}(s_\delta)})\Pi^{\omega_2}(s)\mathrm{d}s.
$$

This, together with Gronwall's inequality, implies that

$$
\mathbb{E}\|Y_{t_\delta}(\xi,i_0)-Y_{t_\delta}(\eta,i_0)\|^2_\infty\le c\,\|\xi-\eta\|^2_\infty\mathrm{e}^{\mathrm{e}^{-\widehat\alpha(\tau+\delta)}\rho\sqrt\delta t}\mathbb{E}\,\mathrm{e}^{\int_0^t(\alpha_{\Lambda(s_\delta)}+\mathrm{e}^{-\widehat\alpha(\tau+\delta)}\beta_{\Lambda(s_\delta)})\mathrm{d}s}.
$$

Thus, according to Lemma 4.1, it holds

$$
\mathbb{E}\|Y_{t_\delta}(\xi,i_0)-Y_{t_\delta}(\eta,i_0)\|^2_\infty\le c\,\|\xi-\eta\|^2_\infty\mathrm{e}^{\mathrm{e}^{-\widehat\alpha(\tau+\delta)}\rho\sqrt\delta t}\mathrm{e}^{-2\eta_1t/3}
$$

for $\delta\in(0,\delta_1)$ with some $\delta_1\in(0,1)$ sufficiently small. As $\eta_1>0$, there exists $\delta_2\in(0,\delta_1)$ such that $\mathrm{e}^{-\widehat\alpha(\tau+\delta)}\rho\sqrt\delta<\eta_1$ for $\delta\in(0,\delta_2)$. As a consequence, (4.10) holds for $\delta\in(0,\delta_2)$. □

Lemma 4.3. Under the assumptions of Lemma 4.2, there exists some $\delta_0 \in (0,1)$ such that

$$\mathbb{E}\|Y_{t_\delta}(\xi,i)\|_\infty^2 \le c\,(1+\|\xi\|_\infty^2), \qquad t\ge \tau+1, \quad \delta\in(0,\delta_0) \tag{4.15}$$

for any $(\xi,i)\in\mathscr{C}\times\mathbf{S}$.

Proof. Below, we assume $t\ge\tau+1$. Carrying out the procedure to gain (4.12), we have

$$\begin{aligned}
&\mathrm{e}^{-\int_0^t(\gamma+\alpha_{\Lambda^{\omega_2}(s_\delta)})\mathrm{d}s}|Y^{\omega_2}(t)|^2\\
&=|\xi(0)|^2+\int_0^t \mathrm{e}^{-\int_0^s(\gamma+\alpha_{\Lambda^{\omega_2}(r_\delta)})\mathrm{d}r}\Big\{-(\gamma+\alpha_{\Lambda^{\omega_2}(s_\delta)})|Y^{\omega_2}(s)|^2\\
&\quad+2\langle Y^{\omega_2}(s), b(Y^{\omega_2}_{s_\delta},\Lambda^{\omega_2}(s_\delta))\rangle+\|\sigma(\Lambda^{\omega_2}(s_\delta))\|_{\mathrm{HS}}^2\Big\}\mathrm{d}s\\
&\quad+2\int_0^t \mathrm{e}^{-\int_0^s(\gamma+\alpha_{\Lambda^{\omega_2}(r_\delta)})\mathrm{d}r}\langle Y^{\omega_2}(s),\sigma(\Lambda^{\omega_2}(s_\delta))\mathrm{d}\omega_1(s)\rangle,
\end{aligned} \tag{4.16}$$

where $\gamma>0$ is introduced in (2.7) and is to be determined. Thanks to (1.17), it holds that

$$|Y^{\omega_2}(t)-Y^{\omega_2}(t_\delta)|^2\le 2L_0^2\delta\|Y^{\omega_2}_{t_\delta}\|_\infty^2+c\,|\omega_1(t)-\omega_1(t_\delta)|^2. \tag{4.17}$$

and, from (1.16) and $Y(t)=\xi(-\tau)$ for any $t\in[-\tau-1,-\tau)$, that

$$\|Y^{\omega_2}_{t_\delta}\|_\infty\le \sup_{t-\tau-\delta\le s\le t}|Y^{\omega_2}(s)|. \tag{4.18}$$

Thus, by combining (2.7) with (4.16)-(4.18), it follows that

$$\begin{aligned}
&\mathrm{e}^{-\int_0^t(\gamma+\alpha_{\Lambda^{\omega_2}(s_\delta)})\mathrm{d}s}|Y^{\omega_2}(t)|^2\\
&\le|\xi(0)|^2+\int_0^t \mathrm{e}^{-\int_0^s(\gamma+\alpha_{\Lambda^{\omega_2}(r_\delta)})\mathrm{d}r}\Big\{c+((\gamma+\breve{a})\sqrt{\delta}+\beta_{\Lambda^{\omega_2}(s_\delta)})\|Y^{\omega_2}_{s_\delta}\|_\infty^2\\
&\quad+\frac{1+\gamma+2\breve{a}}{\sqrt{\delta}}|Y^{\omega_2}(s)-Y^{\omega_2}(s_\delta)|^2+\sqrt{\delta}|b(Y^{\omega_2}_{s_\delta},\Lambda^{\omega_2}(s_\delta))|^2\Big\}\mathrm{d}s+\Theta^{\omega_2}(t)\\
&\le|\xi(0)|^2+\int_0^t \mathrm{e}^{-\int_0^s(\gamma+\alpha^{\varepsilon}_{\Lambda^{\omega_2}(r_\delta)})\mathrm{d}r}\Big\{c+\frac{c}{\sqrt{\delta}}\,|\omega_1(t)-\omega_1(t_\delta)|^2\\
&\quad+(\vartheta\sqrt{\delta}+\beta_{\Lambda^{\omega_2}(s_\delta)})\sup_{s-\tau-\delta\le r\le s}|Y(r)|^2\Big\}\mathrm{d}s+\Theta^{\omega_2}(t)
\end{aligned} \tag{4.19}$$

where $\vartheta:=\gamma+\breve{a}+2L_0^2(2+\gamma+2\breve{a})$, and

$$\Theta^{\omega_2}(t):=2\int_0^t \mathrm{e}^{-\int_0^s(\gamma+\alpha_{\Lambda^{\omega_2}(r_\delta)})\mathrm{d}r}\langle Y^{\omega_2}(s),\sigma(\Lambda^{\omega_2}(s_\delta))\mathrm{d}\omega_1(s)\rangle.$$

Following the argument to derive (2.10), for $0\le s\le t$ with $t-s\in[0,\tau+\delta]$ and $\kappa\in(0,1)$, which is also to be determined, we have

$$\mathbb{E}_{\mathbb{P}_1}\Big(\sup_{s\le r\le t}\Theta^{\omega_2}(r)\Big)\le\kappa\,\mathrm{e}^{\widehat{\alpha}(\tau+\delta)}\Pi^{\omega_2}(t)+c\,\mathrm{e}^{-\int_0^t(\delta+\alpha_{\Lambda^{\omega_2}(r_\delta)})\mathrm{d}r}, \tag{4.20}$$

and observe that

$$\Pi^{\omega_2}(t) := \mathrm{e}^{-\int_0^t(\gamma+\alpha_{\Lambda^{\omega_2}(s_\delta)})\mathrm{d}s}\mathbb{E}_{\mathbb{P}_1}\Big(\sup_{t-\tau-\delta\le s\le t}|Y^{\omega_2}(s)|^2\Big)$$
$$\le \mathrm{e}^{-\widehat{\alpha}(\tau+\delta)}\mathbb{E}_{\mathbb{P}_1}\Big(\sup_{t-\tau-\delta\le s\le t}\Big(\mathrm{e}^{-\int_0^s(\gamma+\alpha_{\Lambda^{\omega_2}(r)})\mathrm{d}r}|Y^{\omega_2}(s)|^2\Big)\Big).$$

Hence, we deduce from (4.19) and (4.20) that

$$\Pi^{\omega_2}(t) \le \frac{\mathrm{e}^{-\widehat{\alpha}(\tau+\delta)}}{1-\kappa}\Big\{c\,\|\xi\|_\infty^2 + c\,\mathrm{e}^{-\int_0^t(\gamma+\alpha_{\Lambda^{\omega_2}(r_\delta)})\mathrm{d}r} + c\int_0^t \mathrm{e}^{-\int_0^s(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r_\delta)})\mathrm{d}r}\mathrm{d}s$$
$$+\int_0^t(\vartheta\sqrt{\delta}+\beta_{\Lambda^{\omega_2}(s_\delta)})\Pi^{\omega_2}(s)\mathrm{d}s\Big\}.$$

Thus, an application of Gronwall's inequality enables us to get

$$\begin{aligned}\Pi^{\omega_2}(t) \le{}& \frac{\mathrm{e}^{-\widehat{\alpha}(\tau+\delta)}}{1-\kappa}\Big\{c\,\|\xi\|_\infty^2 + c\,\mathrm{e}^{-\int_0^t(\gamma+\alpha_{\Lambda^{\omega_2}(r_\delta)})\mathrm{d}r} + c\int_0^t \mathrm{e}^{-\int_0^s(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r_\delta)})\mathrm{d}r}\mathrm{d}s\Big\}\\ &+\frac{\mathrm{e}^{-\widehat{\alpha}(\tau+\delta)}}{1-\kappa}\int_0^t\Big\{c\,\|\xi\|_\infty^2 + c\,\mathrm{e}^{-\int_0^s(\gamma+\alpha_{\Lambda^{\omega_2}(r_\delta)})\mathrm{d}r} + c\int_0^s \mathrm{e}^{-\int_0^u(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r_\delta)})\mathrm{d}r}\mathrm{d}u\Big\}\\ &\times\Phi^{\omega_2}(s_\delta)\exp\Big(\int_s^t\Phi^{\omega_2}(r_\delta)\mathrm{d}r\Big)\mathrm{d}s,\end{aligned}\tag{4.21}$$

in which

$$\Phi^{\omega_2}(t_\delta) := \frac{\mathrm{e}^{-\widehat{\alpha}(\tau+\delta)}(\vartheta\delta+\beta_{\Lambda^{\omega_2}(t_\delta)})}{1-\kappa}.$$

For any $0\le s\le t$, set

$$\Upsilon^{\omega_2}(s,t) := \int_s^t\Phi^{\omega_2}(u_\delta)\exp\Big(\int_u^t\Phi^{\omega_2}(r_\delta)\mathrm{d}r\Big)\mathrm{d}u.$$

For any $0\le s\le t$, there exist integers $j,k>0$ such that $s\in[j\delta,(j+1)\delta)$ and $t\in[k\delta,(k+1)\delta)$. If $j=k$, then we obtain that

$$\Upsilon^{\omega_2}(s,t) = \int_s^t\Phi^{\omega_2}(k\delta)\mathrm{e}^{\Phi^{\omega_2}(k\delta)(t-u)}\mathrm{d}u = \mathrm{e}^{\Phi^{\omega_2}(k\delta)(t-s)}-1 = \exp\Big(\int_s^t\Phi^{\omega_2}(r_\delta)\mathrm{d}r\Big)-1.\tag{4.22}$$

In the sequel, we assume that $j<k$. Observe that

(4.23)

$$\begin{aligned}\Upsilon^{\omega_2}(s,t) &= \mathrm{e}^{\Phi^{\omega_2}(k\delta)(t-k\delta)}\Upsilon(s,k\delta)+\mathrm{e}^{\Phi^{\omega_2}(k\delta)(t-k\delta)}-1\\ &= \mathrm{e}^{\Phi^{\omega_2}(k\delta)(t-k\delta)}\Big\{\mathrm{e}^{\Phi^{\omega_2}((k-1)\delta)\delta}\Upsilon(s,(k-1)\delta)+\mathrm{e}^{\Phi^{\omega_2}((k-1)\delta)\delta}-1\Big\}+\mathrm{e}^{\Phi^{\omega_2}(k\delta)(t-k\delta)}-1\\ &= \mathrm{e}^{\Phi^{\omega_2}(k\delta)(t-k\delta)+\Phi^{\omega_2}((k-1)\delta)\delta}\Upsilon(s,(k-1)\delta)+\mathrm{e}^{\Phi^{\omega_2}(k\delta)(t-k\delta)+\Phi^{\omega_2}((k-1)\delta)\delta}-1\\ &= \cdots\\ &= \mathrm{e}^{\Phi^{\omega_2}(k\delta)(t-k\delta)+\Phi^{\omega_2}((k-1)\delta)\delta+\cdots+\Phi^{\omega_2}(j\delta)((j+1)\delta-s)}-1\\ &= \exp\Big(\int_s^t\Phi^{\omega_2}(r_\delta)\mathrm{d}r\Big)-1.\end{aligned}$$

Subsequently, taking (4.21)-(4.23) and Fubini's theorem into account, we deduce that

$$\mathbb{E}\Big(\sup_{t-\tau-\delta\le s\le t}|Y(s)|^2\Big)\le c\,\mathbb{E}\Big\{1+\mathrm{e}^{\int_0^t(\gamma+\alpha_{\Lambda(s_\delta)}+\Phi(s_\delta))\mathrm{d}s}+\int_0^t\mathrm{e}^{\int_s^t(\gamma+\alpha_{\Lambda(r_\delta)}+\Phi(r_\delta))\mathrm{d}r}\mathrm{d}s\Big\},$$

where

$$\Phi(t_\delta):=\frac{\mathrm{e}^{-\widehat{\alpha}(\tau+\delta)}(\vartheta\sqrt{\delta}+\beta_{\Lambda(t_\delta)})}{1-\kappa}.$$

Thus, with the help of $\eta_1>0$, (4.15) follows from Lemma 4.1 and by taking $\gamma,\delta,\kappa\in(0,1)$ sufficiently small. □

Now we are ready to finish the proof of Theorem 1.4.

Proof. With Lemmas 4.2 and 4.3 in hand, we can complete the argument of Theorem 1.4 by mimicking the proof of Theorem 1.2. □

5 Proof of Theorem 1.5

Before we complete the proof of Theorem 1.5, let's make some preparations. For any $t\ge0$, let $\mathcal{F}_t=\sigma((W(u),\Lambda(u)),0\le u\le t)\vee\mathcal{N}$, where $\mathcal{N}$ stands for the set of all $\mathbb{P}$-null sets in $\mathscr{F}$.

Lemma 5.1. $(Y_{k\delta},\Lambda(k\delta))$ is a homogeneous Markov chain, i.e.,

(5.1)
$$\begin{aligned}&\mathbb{P}((Y_{(k+1)\delta},\Lambda((k+1)\delta))\in A\times\{j\}|(Y_{k\delta},\Lambda(k\delta))=(\xi,i))\\&=\mathbb{P}((Y_\delta,\Lambda(\delta))\in A\times\{j\}|(Y_0,\Lambda(0))=(\xi,i))\end{aligned}$$

and

(5.2)
$$\mathbb{P}((Y_{(k+1)\delta},\Lambda((k+1)\delta))\in A\times\{i\}|\mathscr{F}_{k\delta})=\mathbb{P}((Y_{(k+1)\delta},\Lambda((k+1)\delta))\in A\times\{i\}|(Y_{k\delta},\Lambda(k\delta)))$$

for any $A\in\mathscr{B}(\mathscr{C})$ and $(\xi,i)\in\mathscr{C}\times\mathbf{S}$.

Proof. We shall verify (5.1) and (5.2) one-by-one. To begin, we show that (5.1) holds. It is easy to see from (1.16) that

(5.3)
$$Y_{k\delta}(i\delta)=Y((k+i)\delta),\qquad i=-M,\cdots,-1.$$

Observe from (1.20) and (5.3) that

(5.4)
$$\begin{aligned}Y_\delta(\theta)&=Y((1+i)\delta)+\frac{\theta-i\delta}{\delta}\{Y((2+i)\delta)-Y((1+i)\delta)\}\\&=\begin{cases}Y(0)+\frac{\theta+\delta}{\delta}\{Y(\delta)-Y(0)\}, & \theta\in[-\delta,0]\\ Y((1+i)\delta)+\frac{\theta-i\delta}{\delta}\{Y((2+i)\delta)-Y((1+i)\delta)\}, & \theta\in[i\delta,(i+1)\delta],\quad i\neq-1\end{cases}\\&=\begin{cases}Y(0)+\frac{\theta+\delta}{\delta}\{b(Y_0,\Lambda(0))\delta+\sigma(Y_0,\Lambda(0))W(\delta)\}, & \theta\in[-\delta,0],\\ Y((1+i)\delta)+\frac{\theta-i\delta}{\delta}\{Y((2+i)\delta)-Y((1+i)\delta)\}, & \theta\in[i\delta,(i+1)\delta],\quad i\neq-1\end{cases}\end{aligned}$$

and that
(5.5)

$$Y_{(k+1)\delta}(\theta) = Y((k+1+i)\delta) + \frac{\theta - i\delta}{\delta}\{Y((k+2+i)\delta) - Y((k+1+i)\delta)\}$$

$$= \begin{cases} Y_{k\delta}(0)+\frac{\theta+\delta}{\delta}\{Y((k+1)\delta)-Y_{k\delta}(0)\}, & \theta\in[-\delta,0] \\ Y_{k\delta}((1+i)\delta)+\frac{\theta-i\delta}{\delta}\{Y_{k\delta}((2+i)\delta)-Y_{k\delta}((1+i)\delta)\}, & \theta\in[i\delta,(i+1)\delta],\ i\neq -1 \end{cases}$$

$$= \begin{cases} Y_{k\delta}(0)+\frac{\theta+\delta}{\delta}\{b(Y_{k\delta},\Lambda(k\delta))\delta \\ \qquad +\sigma(Y_{k\delta},\Lambda(k\delta))(W((k+1)\delta)-W(k\delta))\}, & \theta\in[-\delta,0] \\ Y_{k\delta}((1+i)\delta)+\frac{\theta-i\delta}{\delta}\{Y_{k\delta}((2+i)\delta)-Y_{k\delta}((1+i)\delta)\}, & \theta\in[i\delta,(i+1)\delta],\ i\neq -1. \end{cases}$$

Thus, comparing (1.16) with (5.5) and noting that $W((k+1)\delta) - W(k\delta)$ and $W(\delta)$ are identical in distribution, we infer that $(Y_{(k+1)\delta}, \Lambda((k+1)\delta))$ and $(Y_\delta, \Lambda(\delta))$ are equal in distribution given $(Y_{k\delta}, \Lambda(k\delta)) = (\xi, i)$ and $(Y_0, \Lambda(0)) = (\xi, i)$, respectively. Therefore, (5.1) holds immediately.

Next, we demonstrate that (5.2) is fulfilled. Set

$$\chi^{\xi,j}_{(k+1)\delta}(\theta) := \begin{cases} \xi(0) + \frac{\theta+\delta}{\delta}\{b(\xi,j)\delta + \sigma(\xi,j)(W((k+1)\delta) - W(k\delta))\}, & \theta\in[-\delta,0] \\ \xi((1+i)\delta) + \frac{\theta-i\delta}{\delta}[\xi((2+i)\delta) - \xi((1+i)\delta)], & \theta\in[i\delta,(i+1)\delta],\ i\neq -1. \end{cases}$$

and $\Lambda^{j,\delta}_{k+1} := j + \Lambda((k+1)\delta) - \Lambda(k\delta)$. Thus, it is easy to see that

$$\Lambda((k+1)\delta) = \Lambda^{\Lambda(k\delta),\delta}_{k+1} \qquad \text{and} \qquad Y_{(k+1)\delta} = \chi^{Y_{k\delta},\Lambda(k\delta)}_{(k+1)\delta}. \tag{5.6}$$

For any $0 \le s \le t$, let $\mathcal{G}_{t,s} = \sigma(W(u) - W(s), s \le u \le t) \vee \mathcal{N}$. Plainly, $\mathcal{G}_{(k+1)\delta,k\delta}$ is independent of $\mathscr{F}_{k\delta}$. Moreover, $\chi^{\xi,j}_{(k+1)\delta}$ depends completely on the increment $W((k+1)\delta) - W(k\delta)$ so is $\mathcal{G}_{(k+1)\delta,k\delta}$-measurable. Hence, $\chi^{\xi,j}_{(k+1)\delta}$ is independent of $\mathcal{F}_{k\delta}$. Noting that $\chi^{Y_{k\delta},\Lambda(k\delta)}_{(k+1)\delta}$ and $\Lambda^{\Lambda(k\delta),\delta}_{(k+1)\delta}$ are conditionally independent given $(Y_{k\delta}, \Lambda(k\delta))$. Therefore,

$$\begin{aligned}
&\mathbb{P}((Y_{(k+1)\delta}, \Lambda((k+1)\delta) \in A \times \{j\} | \mathcal{F}_{k\delta}) \\
&= \mathbb{E}(I_{A\times\{j\}}(\chi^{Y_{k\delta},\Lambda(k\delta)}_{(k+1)\delta}, \Lambda^{\Lambda(k\delta)}_{k+1}) | \mathcal{F}_{k\delta}) \\
&= \mathbb{E}(I_A(\chi^{Y_{k\delta},\Lambda(k\delta)}_{(k+1)\delta}) | \mathcal{F}_{k\delta}) \mathbb{E}(I_{\{j\}}(\Lambda^{\Lambda(k\delta)}_{k+1}) | \mathcal{F}_{k\delta}) \\
&= \mathbb{E}(I_A(\chi^{\xi,i}_{(k+1)\delta}))|_{\xi=Y_{k\delta}, i=\Lambda(k\delta)} \mathbb{E}(I_{\{j\}}(\Lambda^i_{k+1}))|_{i=\Lambda(k\delta)} \\
&= \mathbb{P}(\chi^{\xi,i}_{(k+1)\delta} \in A)|_{\xi=Y_{k\delta}, i=\Lambda(k\delta)} \mathbb{P}(\Lambda^i_{k+1} \in \{j\})|_{i=\Lambda(k\delta)} \\
&= \mathbb{P}((\chi^{\xi,i}_{(k+1)\delta}, \Lambda^i_{k+1}) \in A \times \{j\})|_{\xi=Y_{k\delta}, i=\Lambda(k\delta)} \\
&= \mathbb{P}((Y_{(k+1)\delta}, \Lambda((k+1)\delta)) \in A \times \{i\} | (Y_{k\delta}, \Lambda(k\delta))).
\end{aligned}$$

So (5.2) holds, and then $(Y_{k\delta}, \Lambda(k\delta))$ is a homogeneous Markov chain. □

Lemma 5.2. Under the assumptions of Theorem 1.5, there exist $\delta_0 \in (0,1)$ and $\alpha > 0$ such that

$$\mathbb{E}\|Y_{t_\delta}(\xi,i) - Y_{t_\delta}(\eta,i)\|^2_\infty \le c\,\mathrm{e}^{-\alpha t}\|\xi - \eta\|^2_\infty, \qquad t \ge \tau, \quad \delta \in (0,\delta_0) \tag{5.7}$$

for any $\xi, \eta \in \mathscr{C}$ and $i \in \mathbf{S}$.

Proof. For fixed ω_2, we focus on the following SDE

$$\mathrm{d}Y^{\omega_2}(t)=b(Y^{\omega_2}_{t_\delta},\Lambda^{\omega_2}(t_\delta))\mathrm{d}t+\sigma(Y^{\omega_2}_{t_\delta},\Lambda^{\omega_2}(t_\delta))\mathrm{d}\omega_1(t)$$

with the initial value $Y^{\omega_2}(\theta)=\xi(\theta),\theta\in[-\tau,0]$, and $\Lambda^{\omega_2}(0)=i\in\mathbf{S}$. Let $\Upsilon^{\omega_2}(t)$ be defined as in (4.11). By (1.20), it is easy to see that

$$\mathbb{E}_{\mathbb{P}_1}|\Upsilon^{\omega_2}(t)-\Upsilon^{\omega_2}(t_\delta)|^2\le(L_0+L)\delta\Big\{\mathbb{E}_{\mathbb{P}_1}|\Upsilon^{\omega_2}(t_\delta)|^2+\int_{-\tau}^0\mathbb{E}_{\mathbb{P}_1}|\Upsilon^{\omega_2}_{t_\delta}(\theta)|^2v(\mathrm{d}\theta)\Big\}. \tag{5.8}$$

Following the procedure to derive (4.12), we obtain from (**H1**), (1.19) and (5.8) that

$$\begin{aligned}&\mathrm{e}^{-\int_0^t\alpha_{\Lambda^{\omega_2}(s_\delta)}\mathrm{d}s}\mathbb{E}_{\mathbb{P}_1}|\Upsilon^{\omega_2}(t)|^2\\&=|\Upsilon^{\omega_2}(0)|^2+\int_0^t\mathrm{e}^{-\int_0^s\alpha_{\Lambda^{\omega_2}(r_\delta)}\mathrm{d}r}\mathbb{E}_{\mathbb{P}_1}\Big\{-\alpha_{\Lambda^{\omega_2}(s_\delta)}|\Upsilon^{\omega_2}(s)|^2\\&\quad+2\langle\Upsilon^{\omega_2}(s),b(Y^{\omega_2}_{s_\delta}(\xi,i_0),\Lambda^{\omega_2}(s_\delta))-b(Y^{\omega_2}_{s_\delta}(\eta,i_0),\Lambda^{\omega_2}(s_\delta))\rangle\\&\quad+\|\sigma(Y^{\omega_2}_{s_\delta}(\xi,i),\Lambda^{\omega_2}(s_\delta))-\sigma(Y^{\omega_2}_{s_\delta}(\eta,i),\Lambda^{\omega_2}(s_\delta))\|_{\mathrm{HS}}^2\Big\}\mathrm{d}s\\&\le|\Upsilon^{\omega_2}(0)|^2+\int_0^t\mathrm{e}^{-\int_0^s\alpha_{\Lambda^{\omega_2}(r_\delta)}\mathrm{d}r}\Big\{(\breve\alpha+L_0)\sqrt\delta\mathbb{E}_{\mathbb{P}_1}|\Upsilon^{\omega_2}(s_\delta)|^2\\&\quad+(L_0\sqrt\delta+\beta_{\Lambda^{\omega_2}(s_\delta)})\int_{-\tau}^0\mathbb{E}_{\mathbb{P}_1}|\Upsilon^{\omega_2}_{s_\delta}(\theta)|^2v(\mathrm{d}\theta)+\frac{1+2\breve\alpha}{\sqrt\delta}\mathbb{E}_{\mathbb{P}_1}|\Upsilon^{\omega_2}(s)-\Upsilon^{\omega_2}(s_\delta)|^2\Big\}\mathrm{d}s\\&\le|\Upsilon^{\omega_2}(0)|^2+\nu\sqrt\delta\int_0^t\mathrm{e}^{-\int_0^s\alpha_{\Lambda^{\omega_2}(r_\delta)}\mathrm{d}r}\mathbb{E}_{\mathbb{P}_1}|\Upsilon^{\omega_2}(s_\delta)|^2\mathrm{d}s+\Psi^{\omega_2}(t),\end{aligned} \tag{5.9}$$

where $\nu:=\breve\alpha+L_0+(1+2\breve\alpha)(L_0+L)$ and

$$\Psi^{\omega_2}(t):=\int_0^t(\nu\sqrt\delta+\beta_{\Lambda^{\omega_2}(s_\delta)})\mathrm{e}^{-\int_0^s\alpha_{\Lambda^{\omega_2}(r_\delta)}\mathrm{d}r}\int_{-\tau}^0\mathbb{E}_{\mathbb{P}_1}|\Upsilon^{\omega_2}_{s_\delta}(\theta)|^2v(\mathrm{d}\theta)\mathrm{d}s.$$

By virtue of (1.16), we deduce that

$$\begin{aligned}\Psi^{\omega_2}(t)&=\int_{-\tau}^0\int_0^t(\nu\sqrt\delta+\beta_{\Lambda^{\omega_2}(s_\delta)})\mathrm{e}^{-\int_0^s\alpha_{\Lambda^{\omega_2}(r_\delta)}\mathrm{d}r}\mathbb{E}_{\mathbb{P}_1}|\Upsilon^{\omega_2}_{s_\delta}(\theta)|^2\mathrm{d}sv(\mathrm{d}\theta)\\&\le2\sum_{i=-N}^{-1}\int_{i\delta}^{(i+1)\delta}\int_0^t(\nu\sqrt\delta+\beta_{\Lambda^{\omega_2}(s_\delta)})\mathrm{e}^{-\int_0^s\alpha_{\Lambda^{\omega_2}(r_\delta)}\mathrm{d}r}\mathbb{E}_{\mathbb{P}_1}|\Upsilon^{\omega_2}(s_\delta+i\delta)|^2\mathrm{d}sv(\mathrm{d}\theta)\\&\quad+2\sum_{i=-N}^{-1}\int_{i\delta}^{(i+1)\delta}\int_0^t(\nu\sqrt\delta+\beta_{\Lambda^{\omega_2}(s_\delta)})\mathrm{e}^{-\int_0^s\alpha_{\Lambda^{\omega_2}(r_\delta)}\mathrm{d}r}\mathbb{E}_{\mathbb{P}_1}|\Upsilon^{\omega_2}(s_\delta+(i+1)\delta)|^2\mathrm{d}sv(\mathrm{d}\theta)\\&\le c\,\|\xi-\eta\|_\infty^2+\int_0^t\Theta^{\omega_2}_\nu(s)\mathrm{e}^{-\int_0^s\alpha_{\Lambda^{\omega_2}(r_\delta)}\mathrm{d}r}\mathbb{E}_{\mathbb{P}_1}|\Upsilon^{\omega_2}(s_\delta)|^2\mathrm{d}s,\end{aligned} \tag{5.10}$$

where

$$\Theta^{\omega_2}_\nu(t):=2\mathrm{e}^{-\widehat\alpha\tau}\sum_{i=-N}^{-1}\int_{i\delta}^{(i+1)\delta}\{2\nu\sqrt\delta+\beta_{\Lambda^{\omega_2}(t_\delta-i\delta)}+\beta_{\Lambda^{\omega_2}(t_\delta-(i+1)\delta)}\}v(\mathrm{d}\theta). \tag{5.11}$$

Inserting (5.10) into (5.9), we arrive at

$$\mathrm{e}^{-\int_0^t \alpha_{\Lambda^{\omega_2}(s_\delta)}\mathrm{d}s}\mathbb{E}_{\mathbb{P}_1}|\Upsilon^{\omega_2}(t)|^2 \le c\,\|\xi-\eta\|_\infty^2 + \int_0^t (\nu\sqrt{\delta}+\Theta_\nu^{\omega_2}(s))\mathrm{e}^{-\int_0^s \alpha_{\Lambda^{\omega_2}(r_\delta)}\mathrm{d}r}\mathbb{E}_{\mathbb{P}_1}|\Upsilon^{\omega_2}(s_\delta)|^2\mathrm{d}s.$$

This, together with the fact that

$$\Pi^{\omega_2}(t) := \mathrm{e}^{-\int_0^t \alpha_{\Lambda^{\omega_2}(s_\delta)}\mathrm{d}s}\sup_{t-\delta\le s\le t}\mathbb{E}_{\mathbb{P}_1}|\Upsilon^{\omega_2}(s)|^2 \le \mathrm{e}^{-\widehat{a}\delta}\sup_{t-\delta\le s\le t}\Big(\mathrm{e}^{-\int_0^s \alpha_{\Lambda^{\omega_2}(r_\delta)}\mathrm{d}r}\mathbb{E}_{\mathbb{P}_1}|\Upsilon^{\omega_2}(s)|^2\Big), \tag{5.12}$$

implies that

$$\Pi^{\omega_2}(t) \le c\,\|\xi-\eta\|_\infty^2 + \mathrm{e}^{-\widehat{a}\delta}\int_0^t (\nu\sqrt{\delta}+\Theta_\nu^{\omega_2}(s))\Pi^{\omega_2}(s)\mathrm{d}s.$$

Thus, an application of Gronwall inequality leads to

$$\Pi^{\omega_2}(t) \le c\,\|\xi-\eta\|_\infty^2 \mathrm{e}^{\mathrm{e}^{-\widehat{a}\delta}\int_0^t(\nu\sqrt{\delta}+\Theta_\nu^{\omega_2}(s))\mathrm{d}s}. \tag{5.13}$$

Furthermore, observe that

$$\begin{aligned}\int_0^t \Theta_\nu^{\omega_2}(s)\mathrm{d}s &= 2\mathrm{e}^{-\widehat{\alpha}\tau}\sum_{i=-N}^{-1}\int_{i\delta}^{(i+1)\delta}\int_{-i\delta}^{t-i\delta}\{2\nu\sqrt{\delta}+\beta_{\Lambda^{\omega_2}(s_\delta)}\}\mathrm{d}s v(\mathrm{d}\theta)\\ &\quad + 2\mathrm{e}^{-\widehat{\alpha}\tau}\sum_{i=-N}^{-1}\int_{i\delta}^{(i+1)\delta}\int_{-(i+1)\delta}^{t-(i+1)\delta}\beta_{\Lambda^{\omega_2}(s_\delta)}\mathrm{d}s v(\mathrm{d}\theta)\\ &\le c + 4\mathrm{e}^{-\widehat{\alpha}\tau}\int_0^t\{\nu\sqrt{\delta}+\beta_{\Lambda^{\omega_2}(s_\delta)}\}\mathrm{d}s.\end{aligned} \tag{5.14}$$

Hence, we infer from (5.13) and (5.14) that

$$\mathbb{E}|Y(t;\xi,i)-Y(t;\eta,i)|^2 \le c\,\|\xi-\eta\|_\infty^2\mathbb{E}\,\mathrm{e}^{\mathrm{e}^{-\widehat{a}\delta}(1+4\mathrm{e}^{-\widehat{a}\tau})\nu\sqrt{\delta}t+\int_0^t\{\alpha_{\Lambda(s_\delta)}+4\mathrm{e}^{-\widehat{a}(\tau+\delta)}\}\beta_{\Lambda(s_\delta)}\mathrm{d}s}.$$

By applying Lemma 4.1 and combining with $\eta_3>0$, there exist $\delta_0\in(0,1)$ and $\alpha>0$ such that

$$\mathbb{E}|Y(t;\xi,i)-Y(t;\eta,i)|^2 \le c\,\mathrm{e}^{-\alpha t}\|\xi-\eta\|_\infty^2, \qquad t\ge\tau,\quad \delta\in(0,\delta_0). \tag{5.15}$$

With (**H2**) and (5.15) in hand, (5.7) can be obtained by a standard procedure. □

Lemma 5.3. Let the assumptions of Lemma 5.2 hold. Then, there exist some $\delta_0\in(0,1)$ such that

$$\mathbb{E}\|Y_{t_\delta}(\xi,i)\|_\infty^2 \le c\,(1+\|\xi\|_\infty^2), \qquad t\ge\tau,\quad \delta\in(0,\delta_0) \tag{5.16}$$

for any $(\xi,i)\in\mathscr{C}\times\mathbf{S}$.

Proof. Mimicking the procedure to derive (4.16), we have

$$
\begin{aligned}
&\mathrm{e}^{-\int_0^t(\gamma+\alpha_{\Lambda^{\omega_2}(s_\delta)})\mathrm{d}s}\mathbb{E}_{\mathbb{P}_1}|Y^{\omega_2}(t)|^2\\
&=|\xi(0)|^2+\int_0^t\mathrm{e}^{-\int_0^s(\gamma+\alpha_{\Lambda^{\omega_2}(r_\delta)})\mathrm{d}r}\mathbb{E}_{\mathbb{P}_1}\Big\{-(\gamma+\alpha_{\Lambda^{\omega_2}(s_\delta)})|Y^{\omega_2}(s)|^2\\
&\quad+2\langle Y^{\omega_2}(s),b(Y^{\omega_2}_{s_\delta},\Lambda^{\omega_2}(s_\delta))\rangle+\|\sigma(Y^{\omega_2}_{s_\delta},\Lambda^{\omega_2}(s_\delta))\|_{\mathrm{HS}}^2\Big\}\mathrm{d}s,
\end{aligned}\tag{5.17}
$$

where $\gamma>0$ is introduced in (3.8). By (**H2**) and (1.19), it follows that

$$
\begin{aligned}
\mathbb{E}_{\mathbb{P}_1}|Y^{\omega_2}(t)-Y^{\omega_2}(t_\delta)|^2&=\mathbb{E}_{\mathbb{P}_1}|b(Y^{\omega_2}_{t_\delta},\Lambda^{\omega_2}(t_\delta))|^2\delta^2+\mathbb{E}_{\mathbb{P}_1}\|\sigma(Y^{\omega_2}_{t_\delta},\Lambda^{\omega_2}(t_\delta))\|_{\mathrm{HS}}^2\delta\\
&\le c+2(L_0+L)\delta\Big\{\mathbb{E}_{\mathbb{P}_1}|Y^{\omega_2}(t_\delta)|^2+\int_{-\tau}^0\mathbb{E}_{\mathbb{P}_1}|Y^{\omega_2}_{t_\delta}(\theta)|^2v(\mathrm{d}\theta)\Big\}.
\end{aligned}\tag{5.18}
$$

Then, taking (5.17) and (5.18) into consideration, we deduce that

$$
\begin{aligned}
&\mathrm{e}^{-\int_0^t(\gamma+\alpha_{\Lambda^{\omega_2}(s_\delta)})\mathrm{d}s}\mathbb{E}_{\mathbb{P}_1}|Y^{\omega_2}(t)|^2\\
&\le|\xi(0)|^2+\int_0^t\mathrm{e}^{-\int_0^s(\gamma+\alpha_{\Lambda^{\omega_2}(r_\delta)})\mathrm{d}r}\Big\{c+\frac{1+2(\gamma+\breve{\alpha})}{\sqrt{\delta}}\mathbb{E}_{\mathbb{P}_1}|Y^{\omega_2}(s)-Y^{\omega_2}(s_\delta)|^2\\
&\quad+(\gamma+\breve{\alpha})\sqrt{\delta}\mathbb{E}_{\mathbb{P}_1}|Y^{\omega_2}(s_\delta)|^2+(\gamma+\beta_{\Lambda^{\omega_2}(s_\delta)})\int_{-\tau}^0\mathbb{E}_{\mathbb{P}_1}|Y^{\omega_2}_{s_\delta}(\theta)|^2v(\mathrm{d}\theta)\\
&\quad+\sqrt{\delta}\mathbb{E}_{\mathbb{P}_1}|b(Y^{\omega_2}_{s_\delta},\Lambda^{\omega_2}(s_\delta))|^2\Big\}\mathrm{d}s\\
&\le|\xi(0)|^2+\int_0^t\mathrm{e}^{-\int_0^s(\gamma+\alpha_{\Lambda^{\omega_2}(r_\delta)})\mathrm{d}r}\Big\{c+\rho\sqrt{\delta}\,\mathbb{E}_{\mathbb{P}_1}|Y^{\omega_2}(s_\delta)|^2\Big\}\mathrm{d}s+\Psi_0^{\omega_2}(t),
\end{aligned}\tag{5.19}
$$

where $\rho:=4(1+\gamma+\breve{a})(L_0+L)$ and

$$
\Psi_0^{\omega_2}(t):=\int_0^t(\rho\sqrt{\delta}+\beta_{\Lambda^{\omega_2}(s_\delta)})\mathrm{e}^{-\int_0^s(\gamma+\alpha_{\Lambda^{\omega_2}(r_\delta)})\mathrm{d}r}\int_{-\tau}^0\mathbb{E}_{\mathbb{P}_1}|Y^{\omega_2}_{s_\delta}(\theta)|^2v(\mathrm{d}\theta)\mathrm{d}s.
$$

Following the argument to deduce (5.10), we find that

$$
\Psi_0^{\omega_2}(t)\le c\,\|\xi\|_\infty^2+\int_0^t\Theta_\rho^{\omega_2}(s)\mathrm{e}^{-\int_0^s(\gamma+\alpha_{\Lambda^{\omega_2}(r_\delta)})\mathrm{d}r}\mathbb{E}_{\mathbb{P}_1}|Y^{\omega_2}(s_\delta)|^2\mathrm{d}s,\tag{5.20}
$$

where $\Theta_\rho^{\omega_2}>0$ is defined as in (5.11). Substituting (5.20) into (5.19) leads to

$$
\begin{aligned}
&\mathrm{e}^{-\int_0^t(\gamma+\alpha_{\Lambda^{\omega_2}(s_\delta)})\mathrm{d}s}\mathbb{E}_{\mathbb{P}_1}|Y^{\omega_2}(t)|^2\\
&\le c\,\|\xi\|_\infty^2+\int_0^t\mathrm{e}^{-\int_0^s(\gamma+\alpha_{\Lambda^{\omega_2}(r_\delta)})\mathrm{d}r}\Big\{c+(\rho\sqrt{\delta}+\Theta_\rho^{\omega_2}(s))\mathbb{E}_{\mathbb{P}_1}|Y^{\omega_2}(s_\delta)|^2\Big\}\mathrm{d}s.
\end{aligned}
$$

This, applying the Gronwall inequality and utilizing (5.12) with Υ^{ω_2} being replaced by Y^{ω_2} enables us to obtain that

$$
\begin{aligned}
\Pi^{\omega_2}(t)&\le c\,\|\xi\|_\infty^2+c\int_0^t\mathrm{e}^{-\int_0^s(\gamma+\alpha_{\Lambda^{\omega_2}(r_\delta)})\mathrm{d}r}\mathrm{d}s+\mathrm{e}^{-\widehat{a}\delta}\int_0^t\Big\{c\,\|\xi\|_\infty^2+c\int_0^s\mathrm{e}^{-\int_0^u(\gamma+\alpha_{\Lambda^{\omega_2}(r_\delta)})\mathrm{d}r}\mathrm{d}u\Big\}\\
&\quad\times(\rho\sqrt{\delta}+\Theta^{\omega_2}(s))\mathrm{e}^{\int_s^t\mathrm{e}^{-\widehat{a}\delta}(\rho\sqrt{\delta}+\Theta^{\omega_2}(r))\mathrm{d}r}\mathrm{d}s.
\end{aligned}
$$

Subsequently, the desired assertion follows from Fubini's theorem and Lemma 4.1 and by taking $\gamma, \delta \in (0,1)$ sufficiently small. □

So far, the proof of Theorem 1.5 can be available.

Proof. With the help of Lemmas 5.1-5.3, we can finish the proof by following the argument of Theorem 1.1. □

References

[1] Appleby, J.A.D., Wu, H., Mao, X, On the almost sure running maxima of solutions of affine neutral stochastic functional diffenrential equations, arXiv:1310.2349v1.

[2] Arriojas, M., Hu, Y., Mohammed, S.-E., Pap, G., A delayed Black and Scholes formula, *Stoch. Anal. Appl.*, **25** (2007), 471–492.

[3] Bao, J., Shao, J., Yuan, C., Approximation of invariant measures for regime-switching diffusions, *Potential Anal.*, **44** (2016), 707–727.

[4] Bardet, J.-B., Guérin, H., Malrieu, F., Long time behavior of diffusions with Markov switching, *ALEA Lat. Am. J. Probab. Math. Stat.*, **7** (2010), 151–170.

[5] Benaim, M., Le Borgne, S., Malrieu, F., Zitt, P.-A., Quantitative ergodicity for some switched dynamical systems, *Electron. Commun. Probab.*, **17** (2012), 14 pp.

[6] Buckwar, E., Riedler, M., An exact stochastic hybrid model of excitable membranes including spatio-temporal evolution, *J. Math. Biol.*, **63** (2011), 1051–1093.

[7] Butkovsky, O., Scheutzow M., Invariant measures for stochastic functional differential equations, arXiv:1703.05120.

[8] Caraballo, T., Chueshov, I. D., Marin-Rubio, P., Real, J., Existence and asymptotic behaviour for stochastic heat equations with multiplicative noise in materials with memory, *Discrete Contin. Dyn. Syst.*, **18** (2007), 253–270.

[9] Cloez, B., Hairer, M., Exponential ergodicity for Markov processes with random switching, *Bernoulli*, **21** (2015), 505–536.

[10] de Saporta, B., Yao, J., Tail of a linear diffusion with Markov switching, *Ann. Appl. Probab.*, **15** (2005), 992–1018.

[11] Es-Sarhir, A., Scheutzow, M., van Gaans, O., Invariant measures for stochastic functional differential equations with superlinear drift term, *Differential Integral Equations*, **23** (2010), 189–200.

[12] Hairer, M., Mattingly, J. C., Scheutzow, M., Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations, *Probab. Theory Related Fields*, **149** (2011), 223–259.

[13] Hale, J. K., Lunel, S. M., *Introduction to Functional Differential Equations*, Applied Mathematical Sciences. Springer-Verlag, New York, 1993.

[14] Katsoulakis, M. A., Majda, A. J., Sopasakis, A., Intermittency, metastability and coarse graining for coupled deterministic-stochastic lattice systems, *Nonlinearity* , **19** (2006), 1021–1323.

[15] Kinnally, M. S., Williams, R. J., On existence and uniqueness of stationary distributions for stochastic delay differential equations with positivity constraints, *Electron. J. Probab.*, **15** (2010), 409–451.

[16] Li, J., Shao, J., Algebraic stability of non-homogeneous regime-switching diffusion processes, arXiv:1509.04768v2.

[17] Lygeros, J., Mao, X., Yuan, C., Stochastic hybrid delay population dynamics. Hybrid systems: computation and control, 436–450, Lecture Notes in Comput. Sci., 3927, Springer, Berlin, 2006.

[18] Majda, A. J., Tong, X., Geometric ergodicity for piecewise contracting processes with applications for tropical stochastic lattice models, *Comm. Pure Appl. Math.*, **69** (2016), 1110–1153.

[19] Mao, X., Sabanis, S., Delay geometric Brownian motion in financial option valuation, *Stochastics*, **85** (2013), 295–320.

[20] Mao, X., Truman, A., Yuan, C., Euler-Maruyama approximations in mean-reverting stochastic volatility model under regime-switching, *J. Appl. Math. Stoch. Anal.*, 2006, Art. ID 80967, 20 pp.

[21] Mao, X., Delay population dynamics and environmental noise, *Stoch. Dyn.*,**5** (2005), 149–162.

[22] Mohammed, S-E. A., Stochastic Functional Differential Equations, Pitman, Boston, 1984.

[23] Pinsky, M., Pinsky, R., Transience recurrence and central limit theorem behavior for diffusions in random temporal enrivoments, *Ann. Probab.*, **21** (1993), 433–452.

[24] Pinsky, M., Scheutzow, M., Some remarks and examples concerning the transience and recurrence of random diffusions, *Ann. Inst. Henri. Poincaré*, **28** (1992), 519–536.

[25] Reiß, M., Riedle, M., van Gaans, O., Delay differential equations driven by Lévy processes: stationarity and Feller properties, *Stochastic Process. Appl.*, **116** (2006), 1409–1432.

[26] Shao, J., Ergodicity of regime-switching diffusions in Wasserstein distances, *Stochastic Process. Appl.*, **125** (2015), 739–758.

[27] Shao, J., Criteria for transience and recurrence of regime-switching diffusion processes, *Electron. J. Probab.*, **20** (2015), no. 63, 1–15.

[28] Shao, J., Xi, F., Strong ergodicity of the regime-switching diffusion processes, *Stochastic Process. Appl.*, **123** (2013), 3903–3918.

[29] Song, Q., Yin, G., Zhu, C., Optimal switching with constraints and utility maximization of an indivisible market, *SIAM J. Control Optim.*, **50** (2012), 629–651.

[30] Tong, X., Majda, A. J., Moment bounds and geometric ergodicity of diffusions with random switching and unbounded transition rates, *Res. Math. Sci.*, **3** (2016), Paper No. 41, 33 pp.

[31] von Renesse, Max-K., Scheutzow, M., Existence and uniqueness of solutions of stochastic functional differential equations, *Random Oper. Stoch. Equ.*, **18** (2010), 267–284.

[32] Xi, F., Zhu, C., On Feller and Strong Feller Properties and Exponential Ergodicity of Regime-Switching Jump Diffusion Processes with Countable Regimes, arXiv:1702.01048.

[33] Yuan, C., Mao, X., Asymptotic stability in distribution of stochastic differential equations with Markovian switching, *Stochastic Process. Appl.*, **103** (2003), 277–291.

[34] Yuan, C., Mao, X., Stationary distributions of Euler-Maruyama-type stochastic difference equations with Markovian switching and their convergence, *J. Difference Equ. Appl.*, **11** (2005), 29–48.

[35] Yuan, C., Zou, J., Mao, X., Stability in distribution of stochastic differential delay equations with Markovian switching, *Systems Control Lett.*, **50** (2003), 195–207.

[36] Zhu, C., Yin, G., On competitive Lotka-Volterra model in random environments, *J. Math. Anal. Appl.*, **357** (2009), 154–170.