Nonlinear Analysis 228 (2023) 113201

Contents lists available at ScienceDirect

Nonlinear Analysis

www.elsevier.com/locate/na

# Invariant probability measures for path-dependent random diffusions☆

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

[a] *Center of Applied Mathematics, Tianjin University, Tianjin 300072, China*
[b] *Department of Mathematics, Swansea University, Bay Campus, Swansea SA1 8EN, UK*

A R T I C L E   I N F O

*Article history:*
Received 20 August 2021
Accepted 16 December 2022
Communicated by Francesco Maggi

*MSC:*
37A25
60H10
60H30
60K37

*Keywords:*
Invariant probability measure
Path-dependent random diffusion
Ergodicity
Wasserstein distance
Euler–Maruyama scheme

A B S T R A C T

In this work, we are concerned with path-dependent random diffusions. Under certain ergodic condition, we show that the path-dependent random diffusion under consideration has 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 (numerical) invariant probability measure and preserves the corresponding ergodic property when the step size is sufficiently small. Moreover, we provide an estimate on the exponential functional of the discrete observation for a Markov chain, which may be interesting by itself.


## 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 $\Lambda(t)$ stands for a jump process. Such diffusions have a wide range of emerging and existing applications in e.g. climate science, material science, molecular biology, ecosystems, econometric modeling, and control and optimization of large scale systems; see e.g. [21,29,37] and references therein. Viewing random diffusions as a number of diffusions with random switching, they may be seemingly not much more different from the classical diffusions. Nevertheless, the coexistence of continuous dynamics and jump processes results in challenges in handling a random diffusion $(X(t), \Lambda(t))$ under consideration, even though, in each random temporal environment, $(X(t))$ is simple enough for intuitive understanding. [19] reveals that $(X(t))$ is exponentially

☆ Supported in part by NNSFC, China (11831014, 12071340).
* Corresponding author.
*E-mail addresses:* jianhaibao@tju.edu.cn (J. Bao), shaojh@tju.edu.cn (J. Shao), C.Yuan@swansea.ac.uk (C. Yuan).

https://doi.org/10.1016/j.na.2022.113201

stable in the $p$th moment in a random temporal environment but algebraically stable in the $p$th moment in the other scenarios, whereas $(X(t))$ is ultimately exponentially stable; [24,25] provide several 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; Under certain ergodicity conditions, [4,7,19] show that the random Ornstein–Uhlenbeck (OU) process (with jumps) admits the heavy tail property.

Recently, ergodicity of random diffusions with constant or non-constant jump rates has been investigated extensively; see, for example, [4–6,27,28] for the setting of constant jump rates, [5,6] as for the setup of bounded non-constant jump rates, [21,30] concerning the framework of unbounded and non-constant jump rates. So far, there are several approaches to explore ergodicity for random diffusions; see, for instance, [4,5,28] via probabilistic coupling arguments, [6,21,30] by the weak Harris' theorem, [27] based on the theory of M-matrix and Perron–Frobenius theorem. For ergodicity of random diffusions with infinite regimes, we refer to [27,33] among others.

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. There is a sizeable literature concerning path-dependent stochastic differential equations (SDEs) upon e.g. wellposedness, existence and uniqueness of stationary solutions, and ergodicity; see e.g. [11,12,14,15,17,23,26] and references therein. In terminology, a path-dependent SDE is also called a functional SDE or an SDE with memory.

Under certain Lyapunov condition (which is not related to the stationary distribution of the Markov chain involved), [34,35] investigate existence and uniqueness of invariant probability measures (IPMs for short) for random diffusions without memory by exploiting the M-matrix trick, and [36] tackles the same issue but for path-dependent random diffusions. Recently, under an ergodic condition, [3] probes deeply into existence and uniqueness of IPMs 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], under ergodic conditions, we are concerned with existence and uniqueness of IPMs not only for path-dependent random diffusions but also for their time discretization versions. In comparison with [3,34,35], the difficulties to treat existence and uniqueness of (numerical) IPMs for path-dependent random diffusions lie in (i) the state space of functional solutions $(X_t)_{t\geq 0}$ is infinite-dimensional; (ii) The couple $(X_t, \Lambda(t))$ is discretized; (iii) The investigation is based on certain ergodic condition. Based on the points above, it is much more challengeable to deal with long term (numerical) behavior of path-dependent random diffusions.

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

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

provided $\triangle \downarrow 0$, where $o(\triangle)$ means $\lim_{\triangle\to 0} o(\triangle)/\triangle = 0$; $Q = (q_{ij})$ is the $Q$-matrix associated with the Markov chain $(\Lambda(t))$. We assume that $(\Lambda(t))$ is irreducible with the stationary distribution $\pi = (\pi_1, \ldots, \pi_N)$, which is determined by solving the algebraic equation $\pi Q = 0$ subject to the constraint condition $\sum_{i\in\mathbf{S}} \pi_i = 1$ with $\pi_i \geq 0$. Let $\mathbf{E} = \mathscr{C} \times \mathbf{S}$ for notation brevity. Define the metric $d$ on $\mathbf{E}$ by

$$d\big((\xi, i), (\eta, j)\big) = \|\xi - \eta\|_\infty + \mathbf{1}_{\{i\neq j\}}, \quad (\xi, i), (\eta, j) \in \mathbf{E},$$

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

$$\mathcal{P}_{p,d}(\mathbf{E}) := \Big\{\nu \in \mathcal{P}(\mathbf{E})\Big| \int_{\mathbf{E}} d\big((\xi,i),(o,i_0)\big)^p \nu\big(\mathrm{d}\xi, \mathrm{d}\{i\}\big) < \infty\Big\}, \quad p > 0$$

for some point $(o, i_0) \in \mathbf{E}$. For $\nu_1, \nu_2 \in \mathcal{P}_{p,d}(\mathbf{E})$, define the Wasserstein distance $\mathbb{W}_{d,p}$ induced by the transportation cost function $d$ by

$$\mathbb{W}_{p,d}(\nu_1,\nu_2) = \inf_{\pi\in\mathcal{C}(\nu_1,\nu_2)} \Big\{\int_{\mathbf{E}\times\mathbf{E}} d\big((\xi,i),(\eta,j)\big)^p \pi\big((\mathrm{d}\xi,\mathrm{d}\{i\}),(\mathrm{d}\eta,\mathrm{d}\{j\})\big)\Big\}^{\frac{1}{1\vee p}}, \quad p > 0,$$

where $\mathcal{C}(\mu,\nu)$ denotes the collection of all probability measures on $\mathbf{E}\times\mathbf{E}$ with marginals $\mu$ and $\nu$, respectively (i.e., $\pi \in \mathcal{C}(\nu_1,\nu_2)$ if and only if $\pi(\cdot,\mathbf{E}) = \nu_1(\cdot)$ and $\pi(\mathbf{E},\cdot) = \nu_2(\cdot)$). Let

$$\Omega_1 = \big\{\omega|\omega : \mathbb{R}_+ \to \mathbb{R}^m \text{ is continuous with } \omega(0) = \mathbf{0}\big\},$$

which is endowed with the locally uniform convergence topology and the Wiener measure $\mathbb{P}_1$ so that the coordinate process $W(t,\omega) := \omega(t)$ is a standard $m$-dimensional Brownian motion. Let $(\Omega_2, \mathscr{B}(\Omega_2), \mathbb{P}_2)$ be the Poisson space with the intensity measure $\mathrm{d}t\mathrm{d}u$, where

$$\Omega_2 = \Big\{\omega\Big|\omega = \sum_{i=1}^n \delta_{t_i,u_i} : n \in \mathbb{N}\cup\{\infty\}, (t_i,u_i) \in \mathbb{R}_+\times\mathbb{R}_+\Big\},$$

the space of configurations (i.e., the realizations of a random point measure). Under the probability measure $\mathbb{P}_2$, $N(\mathrm{d}t,\mathrm{d}u,\omega) := \omega(\mathrm{d}t,\mathrm{d}u)$ is a Poisson random measure with the intensity measure $\mathrm{d}t\mathrm{d}u$. Set

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

Then, under the probability measure $\mathbb{P} := \mathbb{P}_1 \times \mathbb{P}_2$, for $\omega = (\omega_1,\omega_2) \in \Omega$, $\omega_1(\cdot)$ is an $m$-dimensional Brownian motion, $\omega_2(\cdot)$ is a Poisson random measure with the intensity measure $\mathrm{d}t\mathrm{d}u$, and they are mutually independent. Throughout this paper, we shall work on the probability space $(\Omega,\mathscr{F},\mathbb{P})$ constructed above.

Hereinafter, $c > 0$ refers to a generic constant which might change from occurrence to occurrence. Below, we present the framework of our work and state the main results we derive.

### 1.1. *Ergodicity: the additive noise*

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

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

where $b : \mathbf{E} \to \mathbb{R}^n$, $\sigma : \mathbf{S} \to \mathbb{R}^n \otimes \mathbb{R}^m$, and, for each fixed $t \geq 0$, $X_t \in \mathscr{C}$ is defined by

$$X_t(\theta) = X(t+\theta), \quad \theta \in [-\tau, 0].$$

In literature, $(X_t)_{t\geq 0}$ is called the segment (or window) process associated with $(X(t))_{t\geq -\tau}$.

Assume that for $(\xi,j),(\eta,j) \in \mathbf{E}$,

(**A**) There exist $\alpha_j \in \mathbb{R}$ and $\beta_j \in \mathbb{R}_+$ such that

$$2\langle\xi(0)-\eta(0), b(\xi,j)-b(\eta,j)\rangle \leq \alpha_j|\xi(0)-\eta(0)|^2 + \beta_j\|\xi-\eta\|_\infty^2.$$

Under (**A**), by following a more or less standard argument (see e.g. [22]), (1.2) admits from [32, Theorem 2.3] a unique strong solution $(X^{\xi,i}(t))$ with the initial datum $(\xi,i)\in\mathbf{E}$. The segment process (which is also called functional solution in terminology) associated with the solution process $(X^{\xi,i}(t))$ is written as $(X_t^{\xi,i})$ to highlight the initial value $(\xi,i)\in\mathbf{E}$. The pair $(X_t^{\xi,i},\Lambda^i(t))$ is a homogeneous Markov process; see, for instance, [23, Theorem 1.1] and [26, Proposition 3.4] for more details.

For a map $\kappa:\mathbf{S}\to\mathbb{R}_+$, let

$$\widehat{\kappa}=\min_{j\in\mathbf{S}}\kappa_j,\quad \check{\kappa}=\max_{j\in\mathbf{S}}\kappa_j.$$

For any $p\geq 0$, set

$$Q_p:=Q+p\,\mathrm{diag}\big(\alpha_1+\mathrm{e}^{-\widehat{\alpha}\tau}\beta_1,\ldots,\alpha_N+\mathrm{e}^{-\widehat{\alpha}\tau}\beta_N\big)\in\mathbb{R}^N\otimes\mathbb{R}^N,$$

where $Q$ is the $Q$-matrix of the Markov chain $(\Lambda(t))$ and $\tau>0$ is the length of time lag. Let

$$\eta_p=-\max_{\gamma\in\mathrm{spec}(Q_p)}\mathrm{Re}(\gamma),\quad p\geq 0;\quad \kappa^*=\sup\{p\geq 0:\eta_p>0\}\in(0,+\infty],\tag{1.3}$$

where $\mathrm{spec}(Q_p)$ and $\mathrm{Re}(\gamma)$ denote the spectrum (i.e., the multiset of its eigenvalues) of $Q_p$ and the real part of $\gamma$, respectively. Let

$$T=\inf\big\{t\geq 0:\Lambda^i(t)=\Lambda^j(t)\big\}$$

be the coupling time of $(\Lambda^i(t),\Lambda^j(t))$. Since $\mathbf{S}$ is a finite set and $Q$ is irreducible, there exists a constant $\theta>0$ such that

$$\mathbb{P}(T>t)\leq\mathrm{e}^{-\theta t},\qquad t\geq 0,\tag{1.4}$$

see e.g. [1] for more details. Let $P_t((\xi,i),\cdot)$ be the transition kernel associated with the Markov process $(X_t^{\xi,i},\Lambda^i(t))$. For $\nu\in\mathcal{P}(\mathbf{E})$, let $\nu P_t$ denote the law of $(X_t,\Lambda(t))$ when $(X_0,\Lambda(0))$ is distributed according to $\nu\in\mathcal{P}(\mathbf{E})$.

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

**Theorem 1.1.** *Assume* (**A**) *and* $\kappa^*>1$. *Then, for any* $\nu_1,\nu_2\in\mathcal{P}_{p,d}(\mathbf{E})$ *with* $p\in(0,1]$,

$$\mathbb{W}_{p,d}\big(\nu_1P_t,\nu_2P_t\big)\leq c\Big\{1+\int_{\mathscr{C}}\|\xi\|_\infty^p\nu_1(d\xi,\mathbf{S})+\int_{\mathscr{C}}\|\xi\|_\infty^p\nu_2(d\xi,\mathbf{S})\Big\}e^{-\frac{p\theta\eta_1 t}{2(\theta+p\eta_1)}}\tag{1.5}$$

*where* $\eta_1>0$ *is defined in* (1.3) *with* $p=1$ *and* $\theta>0$ *is specified in* (1.4). *Furthermore,* (1.5) *implies that* $(X_t^{\xi,i},\Lambda^i(t))$, *determined by* (1.2) *and* (1.1), *admits a unique IPM* $\nu\in\mathcal{P}_{p,d}(\mathbf{E})$ *such that*

$$\mathbb{W}_{p,d}\big(\delta_{(\xi,i)}P_t,\nu\big)\leq c\Big\{1+\|\xi\|_\infty^p+\int_{\mathscr{C}}\|\eta\|_\infty^p\nu(d\eta,\mathbf{S})\Big\}e^{-\frac{p\theta\eta_1 t}{2(\theta+p\eta_1)}},\quad(\xi,i)\in\mathbf{E},\tag{1.6}$$

*where* $\delta_{(\xi,i)}$ *means Dirac's delta measure (or unit mass) at the point* $(\xi,i)$.

Below, we provide an example to demonstrate Theorem 1.1.

**Example 1.2.** Let $(\Lambda(t))_{t\geq 0}$ be a right-continuous Markov chain taking values in $\mathbb{S}=\{0,1\}$ with the generator

$$Q=\begin{pmatrix}-1 & 1\\ \gamma & -\gamma\end{pmatrix}\tag{1.7}$$

for some constant $\gamma>0$. For $\theta>0$, consider a path-dependent random diffusion on $\mathbb{R}^n$

$$\mathrm{d}X(t)=\big\{a_{\Lambda(t)}X(t)+c_{\Lambda(t)}|X(t)|^\theta X(t)+b_{\Lambda(t)}X(t-1)\big\}\mathrm{d}t+\sigma_{\Lambda_t}\mathrm{d}W(t),\ t>0,\tag{1.8}$$

where $a_0, b_0, b_1 > 0, a_1 < 0$ with $\alpha_1 := 2a_1 + b_1 < 2a_0 + b_0 =: \alpha_0$, $c_0 = 0, c_1 = -1$, $\sigma_0, \sigma_1 \in \mathbb{R}^n \otimes \mathbb{R}^m$, and $(W(t))$ is an $m$-dimensional Brownian motion. Set $a := \alpha_0 + b_0 \mathrm{e}^{-\alpha_1}$, $b := \alpha_1 + b_1 \mathrm{e}^{-\alpha_1}$. If

$$b < 0, \quad a + b < 1 + \gamma, \quad \gamma/b + 1/a > 1, \tag{1.9}$$

then $(X_t^{\xi,i}, \Lambda^i(t))$, determined by (1.8) and (1.7), has a unique IPM, and converges exponentially to the equilibrium under the Wasserstein distance $\mathbb{W}_{1,d}$.

In the following, we are going to explain this example. For $j = 0, 1$, let

$$b(\xi, j) = a_j \xi(0) + c_j |\xi|^\theta \xi(0) + b_j \xi(-1), \qquad \xi \in \mathscr{C}.$$

Then, (1.8) can be regarded as the interactions between the following diffusion processes with point delays for $j = 0, 1$,

$$\mathrm{d}X^{(j)}(t) = b(X_t^{(j)}, j)\mathrm{d}t + \sigma_j \mathrm{d}W(t), \quad X_0^{(j)} = \xi \in \mathscr{C}.$$

Let $f(x) = |x|^\theta x, x \in \mathbb{R}^n$. A direct calculation shows

$$\nabla f(x) = \theta |x|^{\theta-2} x x^* + |x|^\theta I_{n\times n}, \quad x \in \mathbb{R}^n,$$

where $x^*$ is the transpose of $x$, and $I_{n\times n}$ is the $n \times n$ identity matrix. Whence, the chain rule yields

$$\begin{aligned}
\langle x - y, f(x) - f(y)\rangle &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}s} \langle x - y, f(y + s(x - y))\rangle \mathrm{d}s \\
&= \int_0^1 \langle x - y, (\nabla_{x-y} f)(y + s(x - y))\rangle \mathrm{d}s \\
&= \int_0^1 \left\{ |y + s(x - y)|^\theta |x - y|^2 + \theta |y + s(x - y)|^{\theta-2} \langle x - y, y + s(x - y)\rangle^2 \right\} \mathrm{d}s \geq 0.
\end{aligned}$$

This, together with $c_0 = 0, c_1 = -1$, implies that for all $\xi, \eta \in \mathscr{C}$,

$$\begin{aligned}
&2\langle \xi(0) - \eta(0), b(\xi, j) - b(\eta, j)\rangle \\
&= 2\langle \xi(0) - \eta(0), a_j(\xi(0) - \eta(0)) + c_j(f(\xi(0)) - f(\eta(0))) + b_j(\xi(-1) - \eta(-1))\rangle \\
&\leq (2a_j + b_j)|\xi(0) - \eta(0)|^2 + b_j |\xi(-1) - \eta(-1)|^2, \qquad j = 0, 1.
\end{aligned}$$

Hence, we have $\alpha_j = 2a_j + b_j$ and $\beta_j = b_j$ in $(\mathbf{H_1})$. Subsequently, it is easy to see that

$$Q_p = \begin{pmatrix} -1 + pa & 1 \\ \gamma & -\gamma + pb \end{pmatrix},$$

where $a := \alpha_0 + b_0 \mathrm{e}^{-\alpha_1}$, $b := \alpha_1 + b_1 \mathrm{e}^{-\alpha_1}$. Observe that the determinant of $Q_p - \lambda I_{2\times 2}$ is given by

$$|Q_p - \lambda I_{2\times 2}| = \lambda^2 - (p(a + b) - 1 - \gamma)\lambda + p^2 ab - p(a\gamma + b).$$

Under (1.9), we take $p > 1$ such that

$$p(a + b) < 1 + \gamma, \qquad p^2 ab - p(a\gamma + b) > 0.$$

Therefore, the characteristic equation $|Q_p - \lambda I_{2\times 2}| = 0$ has two roots with negative real parts. Moreover, due to $p > 1$, we immediately obtain $\kappa^* > 1$.

Consequently, all the assumptions in Theorem 1.1 are satisfied so that the assertions in Example 1.2 hold true.

**Remark 1.1.** In the past few years, ergodicity for stochastic systems with non-uniformly dissipative conditions (e.g., the drift is contractive only outside of a compact set and the drift is allowed to be repulsive on a compact set) has been extensively investigated; see e.g. [8,10,31] and references within. In contrast to the literature mentioned above, the condition imposed in Theorem 1.1 seems to be a little bit strong. Whereas our conditions allow that some subsystems need not to be dissipative even outside of some compact set. The coupling (e.g., reflection coupling and refined basic coupling) approach is one of potential ways to study ergodic property of SDEs with non-dissipative condition, where the crucial points are to choose a suitable coupling metric function and to solve the corresponding second order differential inequality; see, for instance, [9,20] and references therein. Nevertheless, as for a path-dependent diffusion, it is in general a very hard task to solve the associated path-dependent second order differential inequality. Whence, as for path-dependent diffusion processes with non-uniformly dissipative drifts, some new approaches need to be invoked to investigate ergodicity. Once the diffusion coefficient is elliptic, the condition imposed on the drift term might be weaken in a certain sense, see, for example, [2] for more details.

**Remark 1.2.** To demonstrate that the prerequisite $\eta_1 > 0$ cannot be dropped, we take

$$b(\xi, i) = b(\xi(0), \xi(-\tau), i), \qquad \sigma(i) \equiv \sigma, \qquad \xi \in \mathscr{C}$$

as a toy example. For such setting, for all $x, y, z_1, z_2 \in \mathbb{R}^d$ and $i \in \mathbf{S}$, assume that there exist $\alpha_i \in \mathbb{R}$ and $\beta_i \in \mathbb{R}_+$ such that

$$\langle x - y, b(x, z_1, i) - b(y, z_2, i)\rangle \le \alpha_i |x - y|^2 + \beta_i |z_1 - z_2|^2. \tag{1.10}$$

To show the continuous dependence of initial values in the $p$th moment sense, it is crucial to show that (2.1) below holds true with the power 2 therein replaced by the power $p > 0$. To achieve this, via Itô's formula, one of the ingredients is to show that there exist $\alpha_{p,i} \in \mathbb{R}$ and $\beta_{p,i} \in \mathbb{R}_+$ such that for all $x, y, z_1, z_2 \in \mathbb{R}^d$,

$$p|x - y|^{p-2}\langle x - y, b(x, z_1, i) - b(y, z_2, i)\rangle \le \alpha_{p,i}|x - y|^p + \beta_{p,i}|z_1 - z_2|^p \tag{1.11}$$

once (1.10) is valid. It is easy to see from (1.10) that

$$p|x - y|^{p-2}\langle x - y, b(x, z_1, i) - b(y, z_2, i)\rangle \le p\alpha_i|x - y|^p + p\beta_i|x - y|^{p-2}|z_1 - z_2|^2.$$

Whence, to ensure that (1.11) holds true, it is sufficient to verify that for all $x, y, z_1, z_2 \in \mathbb{R}^d$ and $i \in \mathbf{S}$,

$$p\alpha_i|x - y|^p + p\beta_i|x - y|^{p-2}|z_1 - z_2|^2 \le \alpha_{p,i}|x - y|^p + \beta_{p,i}|z_1 - z_2|^p. \tag{1.12}$$

Note that $|x - y|$ is not comparable with $|z_1 - z_2|$ totally. In particular, when $|x - y|$ goes to zero, for the case $p \in (0, 2)$ the left hand side of (1.12) tends to infinity however the right hand side remains finite. Hence, (1.11) is invalid for the setting $p \in (0, 2)$. Thus, to guarantee that (1.11) is true, we particularly take $p = 2$ and therefore we need to assume the associated $\eta_1 > 0$ rather than $\eta_p > 0$ for arbitrary $p > 0$.

For $\beta_i \equiv 0$ (which corresponds to the SDE without memory), (1.10) implies definitely (1.11) for any $p > 0$. Therefore, in this setting, we can take the power $p > 0$ sufficiently small.

### *1.2. Ergodicity: the multiplicative noise*

In the previous subsection, we discuss ergodicity under the Wasserstein distance for a class of path-dependent random diffusions with additive noises. In this subsection, we proceed to consider the same issue under a little bit strong assumptions but for path-dependent random diffusions with multiplicative noises in the form

$$\mathrm{d}X(t) = b(X_t, \Lambda(t))\mathrm{d}t + \sigma(X_t, \Lambda(t))\mathrm{d}W(t), \quad t > 0, \quad (X_0, \Lambda(0)) = (\xi, i) \in \mathbf{E}, \tag{1.13}$$

where $b : \mathbf{E} \to \mathbb{R}^n$ and $\sigma : \mathbf{E} \to \mathbb{R}^n \otimes \mathbb{R}^m$.

Assume that for $(\xi, j), (\eta, j) \in \mathbf{E}$,

($\mathbf{H_1}$) There exist $\alpha_j \in \mathbb{R}$ and $\beta_j \in \mathbb{R}_+$ such that

$$2\langle \xi(0)-\eta(0), b(\xi,j)-b(\eta,j)\rangle \le \alpha_j |\xi(0)-\eta(0)|^2 + \beta_j \int_{-\tau}^0 |\xi(s)-\eta(s)|^2 \mu(\mathrm{d}s),$$

where $\mu(\cdot)$ is a probability measure on $[-\tau, 0]$.

($\mathbf{H_2}$) There exist $\theta_j, \gamma_j \in \mathbb{R}_+$ such that

$$\|\sigma(\xi,j)-\sigma(\eta,j)\|_{\mathrm{HS}}^2 \le \theta_j |\xi(0)-\eta(0)|^2 + \gamma_j \int_{-\tau}^0 |\xi(s)-\eta(s)|^2 \mu(\mathrm{d}s),$$

where $\|\cdot\|_{\mathrm{HS}}$ means the Hilbert–Schmidt norm.

For $p \ge 0$, set

$$Q_p := Q + p\,\mathrm{diag}\Big(\lambda_1 + \zeta_1 \int_{-\tau}^0 \mathrm{e}^{\widehat{\lambda} s}\mu(\mathrm{d}s), \ldots, \zeta_N \int_{-\tau}^0 \mathrm{e}^{\widehat{\lambda} s}\mu(\mathrm{d}s)\Big),$$

where $\lambda_i := \alpha_i + \theta_i, \zeta_i := \beta_i + \gamma_i, i \in \mathbf{S}$. Define

$$\eta_p = -\max_{\gamma \in \mathrm{spec}(Q_p)} \mathrm{Re}(\gamma), \qquad p \ge 0; \quad \kappa^{**} = \sup\{p \ge 0 : \eta_p > 0\} \in (0, +\infty]. \tag{1.14}$$

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

**Theorem 1.3.** *Assume* ($\mathbf{H_1}$), ($\mathbf{H_2}$) *and* $\kappa^{**} > 1$. *Then, for any* $\nu_1, \nu_2 \in \mathcal{P}_{p,d}(\mathbf{E})$ *with* $p \in (0, 1]$,

$$\mathbb{W}_{p,d}\big(\nu_1 P_t, \nu_2 P_t\big) \le c\Big\{1 + \int_{\mathscr{C}} \|\xi\|_\infty^p \nu_1(d\xi, \mathbf{S}) + \int_{\mathscr{C}} \|\xi\|_\infty^p \nu_2(d\xi, \mathbf{S})\Big\} e^{-\frac{p\theta\eta_1 t}{2(\theta + p\eta_1)}} \tag{1.15}$$

*where* $\theta > 0$ *such that* (1.4) *holds and* $\eta_1 > 0$ *is defined in* (1.14) *with* $p = 1$. *Furthermore,* (1.15) *implies that* $(X_t^{\xi,i}, \Lambda^i(t))$ *solving* (1.13) *and* (1.1) *admits a unique IPM* $\nu \in \mathcal{P}_{p,d}(\mathbf{E})$ *such that*

$$\mathbb{W}_{p,d}\big(\delta_{(\xi,i)} P_t, \nu\big) \le c\Big\{1 + \|\xi\|_\infty + \int_{\mathscr{C}} \|\xi\|_\infty^p \nu(d\xi, \mathbf{S})\Big\} e^{-\frac{p\theta\eta_1 t}{2(\theta + p\eta_1)}}. \tag{1.16}$$

### *1.3. Ergodicity preservation: the additive noise*

In this subsection, we aim to discuss exponential ergodicity under the Wasserstein distance for the time discretization version of $(X_t^{\xi,i}, \Lambda^i(t))$, determined by (1.2) and (1.1).

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

$$\mathrm{d}Y(t) = b(Y_{t_\delta}, \Lambda(t_\delta))\mathrm{d}t + \sigma(\Lambda(t_\delta))\mathrm{d}W(t), \qquad t > 0 \tag{1.17}$$

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

$$Y_{k\delta}(\theta) = Y((k+j)\delta) + \frac{\theta - j\delta}{\delta}\{Y((k+j+1)\delta) - Y((k+j)\delta)\}, \tag{1.18}$$

i.e., $Y_{k\delta}$ is the linear interpolation of $(Y(k\delta))_{k \ge -M}$ at the gridpoints. Below, please keep in mind that $Y_{t_\delta} \in \mathscr{C}$ above is defined in a quite different way although it shares the same extrinsic feature as $X_t$. In order to emphasize the initial condition $(\xi, i)$, we shall write $Y^{\xi,i}(t)$ and $Y^{\xi,i}_{t_\delta}$ in lieu of $Y(t)$ and $Y_{t_\delta}$, respectively. The pair $(Y^{\xi,i}_{t_\delta}, \Lambda^i(t_\delta))$ admits the Markov property as Lemma 5.1 below shows. Let $P^{(\delta)}_{k\delta}((\xi, i), \cdot)$ stand for the Markov transition kernel corresponding to the Markov chain $(Y^{\xi,i}_{k\delta}, \Lambda^i(k\delta))$.

To discretize (1.2), in this work we adopt the simple EM scheme (i.e., (1.17)). Since the EM scheme is not stable whenever the drift coefficients of SDEs involved is non-globally Lipschitz (see e.g. [13]), besides (**A**), we further assume that there is an $L_0 > 0$ such that

$$|b(\xi, j) - b(\eta, j)| \le L_0\|\xi - \eta\|_\infty, \qquad (\xi, j), (\eta, j) \in \mathbf{E}. \tag{1.19}$$

The theorem below shows that the discrete-time semigroup generated by $(Y^{\xi,i}_{k\delta}, \Lambda^i(k\delta))$ admits a unique IPM and is exponentially convergent to its equilibrium under the Wasserstein distance.

**Theorem 1.4.** *Assume the assumptions of Theorem 1.1 and suppose further (1.19). Then, for $\nu_1, \nu_2 \in \mathcal{P}_{p,d}(\mathbf{E})$ with $p \in (0, 1]$, there exists $\delta_0 \in (0, 1)$ and $\alpha > 0$ such that*

$$\mathbb{W}_{p,d}\big(\nu_1 P^{(\delta)}_{k\delta}, \nu_2 P^{(\delta)}_{k\delta}\big) \le c\bigg\{1 + \int_{\mathscr{C}} \|\xi\|^p_\infty \nu_1(d\xi, \mathbf{S}) + \int_{\mathscr{C}} \|\xi\|^p_\infty \nu_2(d\xi, \mathbf{S})\bigg\} e^{-\alpha k\delta} \tag{1.20}$$

*for any $k \ge 0$ and $\delta \in (0, \delta_0)$. Furthermore, (1.20) implies that $(Y^{\xi,i}_{k\delta}, \Lambda^i(k\delta))$ admits a unique IPM $\nu^{(\delta)} \in \mathcal{P}_{p,d}(\mathbf{E})$ such that for all $k \ge 0, \delta \in (0, \delta_0), (\xi, i) \in \mathbf{E}$,*

$$\mathbb{W}_{p,d}\big(\delta_{(\xi,i)} P^{(\delta)}_{k\delta}, \nu^{(\delta)}\big) \le c\left\{1 + \|\xi\|^p_\infty + \int_{\mathscr{C}} \|\xi\|^p_\infty \nu^{(\delta)}(d\xi, \mathbf{S})\right\} e^{-\alpha k\delta}.$$

**Remark 1.3.** Recently, under the dissipative condition on the drift term, [18] investigated existence of numerical IPMs for semigroups generated by backward EM scheme associated with path-independent random diffusions, which improves our paper [3]. Another motivation for us to employ the EM scheme is due to the fact that we herein show merely the idea of our work although the assumption (1.19) can be further relaxed by making use of backward EM scheme.

**Corollary 1.5.** *Under the assumptions of Theorem 1.4,*

$$\lim_{\delta \to 0} \mathbb{W}_{p,d}\big(\nu, \nu^{(\delta)}\big) = 0,$$

*where $\nu \in \mathcal{P}_{p,d}(\mathbf{E})$ is the IPM of $(X^{\xi,i}_t, \Lambda^i(t))$ solving (1.2) and (1.1) and $\nu^{(\delta)} \in \mathcal{P}_{p,d}(\mathbf{E})$ is the IPM of $(Y^{\xi,i}_{k\delta}, \Lambda^i(k\delta))$, determined by (1.17) and (1.1).*

### *1.4. Ergodicity preservation: the multiplicative noise*

In this subsection, we move forward to discuss the setup with multiplicative noises. As we stated, the EM scheme is unstable provided that the drift coefficient of an SDE under consideration is locally Lipschitz (see e.g. [13]). So, for the present setting, we further need to strengthen ($\mathbf{H_1}$) as follows: there exists an $L_1 > 0$ such that for all $(\xi, j), (\eta, j) \in \mathbf{E}$,

$$|b(\xi, j) - b(\eta, j)|^2 \le L_1\Big\{|\xi(0) - \eta(0)|^2 + \int_{\mathscr{C}} |\xi(s) - \eta(s)|^2 \mu(\mathrm{d}s)\Big\}. \tag{1.21}$$

Consider the EM scheme corresponding to (1.13):

$$\mathrm{d}Y(t) = b(Y_{t_\delta}, \Lambda(t_\delta))\mathrm{d}t + \sigma(Y_{t_\delta}, \Lambda(t_\delta))\mathrm{d}W(t), \qquad t > 0 \tag{1.22}$$

with the initial condition $Y(\theta) = X(\theta) = \xi(\theta)$ for $\theta \in [-\tau, 0]$ and $\Lambda(0) = i \in \mathbf{S}$, where $Y_{t_\delta}$ is defined as in (5.4). Let

$$\kappa^\star = \sup\{p \geq 0 : \eta_p > 0\} \in (0, +\infty],$$

where $\eta_p$ is defined as in (1.14) by writing $4\zeta_j$ instead of $\zeta_j$.

Concerning the multiplicative noise case, the time discretization of $(X_t(\xi, i), \Lambda^i(t))$, determined by (1.13) and (1.1), also inherits the exponentially ergodic property as the step size is sufficiently small, which is presented as below as another main result in this paper.

**Theorem 1.6.** *Assume* $(\mathbf{H_1})$, $(\mathbf{H_2})$, (1.21) *and* $\kappa^\star > 1$. *Then, for* $\nu_1, \nu_2 \in \mathcal{P}_{p,d}(\mathbf{E})$ *with* $p \in (0, 1]$, *there exist* $\delta_0 \in (0, 1)$ *and* $\alpha > 0$ *such that*

$$\mathbb{W}_{p,d}\big(\nu_1 P_{k\delta}^{(\delta)}, \nu_2 P_{k\delta}^{(\delta)}\big) \leq c\bigg\{1 + \int_{\mathscr{C}} \|\xi\|_\infty^p \nu_1(d\xi, \mathbf{S}) + \int_{\mathscr{C}} \|\xi\|_\infty^p \nu_2(d\xi, \mathbf{S})\bigg\} e^{-\alpha k\delta} \tag{1.23}$$

*for any* $k \geq 0$ *and* $\delta \in (0, \delta_0)$. *Furthermore,* (1.23) *implies that* $(Y_{k\delta}^{\xi,i}, \Lambda^i(k\delta))$, *determined by* (1.22) *and* (1.1), *admits a unique IPM* $\nu^{(\delta)} \in \mathcal{P}_{p,d}(\mathbf{E})$ *such that*

$$\mathbb{W}_{p,d}\big(\delta_{(\xi,i)} P_{k\delta}^{(\delta)}, \nu^{(\delta)}\big) \leq c\left\{1 + \|\xi\|_\infty^p + \int_{\mathscr{C}} \|\xi\|_\infty^p \nu^{(\delta)}(d\xi, \mathbf{S})\right\} e^{-\alpha k\delta}, \quad (\xi, i) \in \mathbf{E}.$$

Below, an example is provided to demonstrate the application of Theorem 1.6.

**Example 1.7.** Assume that $(\Lambda(t))_{t\geq 0}$ is a right-continuous Markov chain with the state space $\mathbb{S} = \{0, 1\}$ and the generator

$$Q = \begin{pmatrix} -\gamma & \gamma \\ 1 & -1 \end{pmatrix} \tag{1.24}$$

for some constant $\gamma > 0$. For $\theta > 0$, consider a scalar path-dependent random diffusion

$$\mathrm{d}X(t) = \big\{a_{\Lambda(t)} X(t) + b_{\Lambda(t)} X(t-1)\big\}\mathrm{d}t + \sigma_{\Lambda_t}(X(t) + X(t-1))\mathrm{d}W(t), \ t > 0 \tag{1.25}$$

where $a_0, b_0, b_1, \sigma_0, \sigma_1 > 0, a_1 < 0$ with $\lambda_1 := 2a_1 + b_1 + 2\sigma_1^2 < 2a_0 + b_0 + 2\sigma_0^2 =: \lambda_0$, and $(W(t))$ is a 1-dimensional Brownian motion. Set $a := \lambda_0 + 4\mathrm{e}^{-\lambda_1}\zeta_0$ and $b := \lambda_1 + 4\mathrm{e}^{-\lambda_1}\zeta_1$ with $\zeta_0 := b_0 + 2\sigma_0^2$ and $\zeta_1 := b_1 + 2\sigma_1^2$. If

$$(2 - b)\gamma + a(b - 1) > 0, \quad a + b < 1 + \gamma, \quad \lambda_0 + 4\mathrm{e}^{-\lambda_1}\zeta_0 + \big(\lambda_1 + 4\mathrm{e}^{-\lambda_1}\zeta_1\big)\gamma < 0, \tag{1.26}$$

then $(Y_{k\delta}^{\xi,i}, \Lambda^i(k\delta))$, determined by (1.24) and (1.25), has a unique IPM $\nu^{(\delta)} \in \mathcal{P}_{p,d}(\mathbf{E})$ for some $p > 1$ when the stepsize is sufficiently small and the corresponding numerical transition kernel converges exponentially under the Wasserstein distance to the IPM $\nu^{(\delta)}$.

Now, we explain this example. For $j \in \{0, 1\}$ and $\xi \in \mathscr{C}$, let

$$b(\xi, j) = a_j \xi(0) + b_j \xi(-1), \qquad \sigma(\xi, j) = \sigma_j(\xi(0) + \xi(-1)).$$

Then, direct calculations show that for $j \in \{0, 1\}$ and $\xi \in \mathscr{C}$,

$$2(\xi(0) - \eta(0))(b(\xi, j) - b(\eta, j)) \leq (2a_j + b_j)(\xi(0) - \eta(0))^2 + b_j(\xi(-1) - \eta(-1))^2$$

and

$$|\sigma(\xi, j) - \sigma(\eta, j)|^2 = 2\sigma_j^2\big((\xi(0) - \eta(0))^2 + (\xi(-1) - \eta(-1))^2\big).$$

Therefore, $(\mathbf{H_1})$ holds for $\alpha_j = 2a_j + b_j$ $\beta_j = b_j$ and $\mu(\mathrm{d}z) = \delta_{-1}(\mathrm{d}z)$ and $(\mathbf{H_2})$ is valid for $\theta_j = \gamma_j = 2\sigma_j^2$, which obviously imply

$$\lambda_j = 2a_j + b_j + 2\sigma_j^2, \qquad \zeta_j = b_j + 2\sigma_j^2.$$

Furthermore, (1.21) holds true with $L = 2((a_0^2 \vee b_0^2) + (a_1^2 \vee b_1^2))$.

Owing to $\lambda_1 < \lambda_0$, we readily obtain that for any $p \geq 0$,

$$Q_p = Q + p\,\mathrm{diag}\Big(\lambda_0 + 4\zeta_0 \int_{-\tau}^0 \mathrm{e}^{\widehat{\lambda}s}\delta_{-1}(\mathrm{d}s), \lambda_1 + 4\zeta_1 \int_{-\tau}^0 \mathrm{e}^{\widehat{\lambda}s}\delta_{-1}(\mathrm{d}s)\Big) = \begin{pmatrix} -\gamma + pa & \gamma \\ 1 & -1 + pb \end{pmatrix}$$

with $a := \lambda_0 + 4\zeta_0\mathrm{e}^{-\lambda_1}$ and $b := \lambda_1 + 4\zeta_1\mathrm{e}^{-\lambda_1}$. It is easy to see that the characteristic equation associated with $Q_p$ is given as below

$$\det\big(Q_p - \lambda I_{2\times 2}\big) = \lambda^2 - \big(p(a+b) - 1 - \gamma\big)\lambda + (2 - pb)\gamma + pa(pb - 1) = 0. \tag{1.27}$$

By taking (1.26) into consideration, there exists a constant $p > 1$ such that

$$(2 - pb)\gamma + pa(pb - 1) > 0, \qquad p(a + b) < 1 + \gamma$$

so that Eq. (1.27) has two negative roots. Therefore, we reach $\kappa^\star > 1$ by taking advantage of $p > 1$.

Due to $a_0, b_0, \sigma_0 > 0$, the following subsystem

$$\mathrm{d}X^{(0)}(t) = b(X_t^{(0)}, 0)\mathrm{d}t + \sigma(X_t^{(0)}, 0)\mathrm{d}W(t)$$

is fully non-dissipative so the functional solution $(X_t^{(0)})$ does not possess an IPM. Furthermore, a direct calculation shows that $(\Lambda(t))_{t\geq 0}$ has a unique IPM $\pi$, which is given explicitly by $\pi = (\pi_0, \pi_1) = (1/(1 + \gamma), \gamma/(1+\gamma))$. Hence, all assumptions in Theorem 1.6 are fulfilled so the assertion of Example 1.7 is followed.

With Theorems 1.3 and 1.6 at hand, by following exactly the proof of Corollary 1.5, we derive the following corollary.

**Corollary 1.8.** *Under the assumptions of Theorem 1.6,*

$$\lim_{\delta\to 0} \mathbb{W}_{p,d}\big(\nu, \nu^{(\delta)}\big) = 0,$$

*where $\nu \in \mathcal{P}_{p,d}(\mathbf{E})$ is the IPM of $(X_t^{\xi,i}, \Lambda^i(t))$ solving (1.13) and (1.1) and $\nu^{(\delta)} \in \mathcal{P}_{p,d}(\mathbf{E})$ is the IPM of $(Y_{k\delta}^{\xi,i}, \Lambda^i(k\delta))$, determined by (1.22) and (1.1).*

The remainder of this paper is arranged as follows. Section 2 is devoted to the proof of Theorem 1.1; Section 3 is concerned with the proofs of Theorem 1.3 and Example 1.2; In Section 4, we aim to investigate estimate on an exponential functional of the discrete observation for the Markov chain involved and meanwhile finish the proofs of Theorem 1.4 and Corollary 1.5; At length, we focus on the Markov property of time discretization version of $(X_t^{\xi,i}, \Lambda^i(t))$ and complete the proof of Theorem 1.6.

## 2. Proof of Theorem 1.1

The lemma below shows that, under suitable assumptions, the functional solutions starting from different initial points will be contractive to each other in the mean-square sense when the time parameter goes to infinity.

**Lemma 2.1.** *Under the assumptions of Theorem 1.1,*

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

*where $\eta_1 > 0$ is defined in (1.3) with $p = 1$.*

**Proof.** For each 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 have a 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 $\alpha_{\Lambda^{\omega_2}(t)}$ as follows

$$\alpha^\varepsilon_{\Lambda^{\omega_2}(t)} := \frac{1}{\varepsilon}\int_t^{t+\varepsilon} \alpha_{\Lambda^{\omega_2}(s)}\mathrm{d}s + \varepsilon t = \int_0^1 \alpha_{\Lambda^{\omega_2}(\varepsilon s+t)}\mathrm{d}s + \varepsilon t, \quad \varepsilon \in (0,1). \tag{2.2}$$

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

$$\Gamma^{\omega_2}(t) = X^{\omega_2,\xi,i}(t) - X^{\omega_2,\eta,i}(t), \quad t \ge 0 \tag{2.3}$$

and $(\Gamma_t^{\omega_2})$ be the corresponding segment process. Applying Itô's formula and taking $(\mathbf{A})$ into account ensures that

$$\begin{aligned} \mathrm{e}^{-\int_0^t \alpha^\varepsilon_{\Lambda^{\omega_2}(s)}\mathrm{d}s}|\Gamma^{\omega_2}(t)|^2 &= |\Gamma^{\omega_2}(0)|^2 + \int_0^t \mathrm{e}^{-\int_0^s \alpha^\varepsilon_{\Lambda^{\omega_2}(r)}\mathrm{d}r}\Big\{-\alpha^\varepsilon_{\Lambda^{\omega_2}(s)}|\Gamma^{\omega_2}(s)|^2 \\ &\quad + 2\big\langle \Gamma^{\omega_2}(s), b(X_s^{\omega_2,\xi,i}, \Lambda^{\omega_2}(s)) - b(X_s^{\omega_2,\eta,i}, \Lambda^{\omega_2}(s))\big\rangle\Big\}\,\mathrm{d}s \\ &\le |\Gamma^{\omega_2}(0)|^2 + \Psi^{\omega_2,\varepsilon}(t) + \int_0^t \beta_{\Lambda^{\omega_2}(s)}\mathrm{e}^{-\int_0^s \alpha^\varepsilon_{\Lambda^{\omega_2}(r)}\mathrm{d}r}\|\Gamma_s^{\omega_2}\|_\infty^2\mathrm{d}s, \end{aligned} \tag{2.4}$$

where

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

Due to the fact that

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

we therefore infer from (2.4) that

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

Since $\alpha^\varepsilon_{\Lambda^{\omega_2}(s)} \to \alpha_{\Lambda^{\omega_2}(s)}$ so that $\Psi^{\omega_2,\varepsilon}(t) \to 0$ as $\varepsilon \to 0$, one has

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

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

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

Due to $\kappa^* > 1$ and [4, Propositions 4.2], we have

$$\sum_{j\in\mathbf{S}}(\alpha_j + \mathrm{e}^{-\widehat{\alpha}\tau}\beta_j)\pi_j < 0. \tag{2.6}$$

Consequently, the desired assertion follows from [4, Propositions 4.1] and [4, Theorem 1.5]. □

The lemma below reveals that the functional solution is ultimately bounded in the mean-square sense.

**Lemma 2.2.** *Under the assumptions of Theorem 1.1,*

$$\sup_{t\ge 0}\mathbb{E}\|X_t^{\xi,i}\|_\infty^2\le c\,(1+\|\xi\|_\infty^2),\qquad (\xi,i)\in\mathbf{E}. \tag{2.7}$$

**Proof.** Let $\alpha^\varepsilon_{\Lambda^{\omega_2}(t)}$ (resp. $\beta^\varepsilon_{\Lambda^{\omega_2}(t)}$) be the smooth approximation of $\alpha_{\Lambda^{\omega_2}(t)}$ (resp. $\beta_{\Lambda^{\omega_2}(t)}$) defined as in (2.2), and write $(X_t)$ in lieu of $(X_t^{\xi,i})$ for notation brevity. By virtue of $(\mathbf{A})$, for all $\gamma>0$, there is a constant $c_\gamma>0$ such that

$$2\langle\xi(0),b(\xi,j)\rangle+\|\sigma(j)\|_{\rm HS}^2\le c_\gamma+(\gamma+\alpha_j)|\xi(0)|^2+\beta_j\|\xi\|_\infty^2,\quad (\xi,j)\in\mathbf{E}. \tag{2.8}$$

Employing Itô's formula and taking (2.8) into consideration provides

$$\begin{aligned}\mathrm{e}^{-\int_0^t(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(s)})\mathrm{d}s}|X^{\omega_2}(t)|^2&\le|\xi(0)|^2+c_\gamma\int_0^t\mathrm{e}^{-\int_0^s(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\mathrm{d}s+\Phi^{\omega_2,\varepsilon}(t)+M^{\omega_2,\varepsilon}(t)\\&\quad+\int_0^t\beta^\varepsilon_{\Lambda^{\omega_2}(s)}\mathrm{e}^{-\int_0^s(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\|X_s^{\omega_2}\|_\infty^2\mathrm{d}s,\end{aligned} \tag{2.9}$$

where

$$\begin{aligned}\Phi^{\omega_2,\varepsilon}(t):&=\int_0^t\mathrm{e}^{-\int_0^s(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\Big\{\,|\alpha_{\Lambda^{\omega_2}(s)}-\alpha^\varepsilon_{\Lambda^{\omega_2}(s)}|\cdot|X^{\omega_2}(s)|^2\\&\quad+|\beta_{\Lambda^{\omega_2}(s)}-\beta^\varepsilon_{\Lambda^{\omega_2}(s)}|\cdot\|X_s^{\omega_2}\|_\infty^2\,\Big\}\,\mathrm{d}s,\\M^{\omega_2,\varepsilon}(t):&=2\int_0^t\mathrm{e}^{-\int_0^s(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\big\langle X^{\omega_2}(s),\sigma(\Lambda^{\omega_2}(s))\mathrm{d}\omega_1(s)\big\rangle.\end{aligned} \tag{2.10}$$

Below, $\|\cdot\|$ means the operator norm. For $0\le s\le t$ with $t-s\in[0,\tau]$ and $\kappa\in(0,1)$, exploiting BDG's inequality yields

$$\begin{aligned}\mathbb{E}_{\mathbb{P}_1}\Big(\sup_{s\le r\le t}M^{\omega_2,\varepsilon}(r)\Big)&=\mathbb{E}_{\mathbb{P}_1}M^{\omega_2,\varepsilon}(s)+\mathbb{E}_{\mathbb{P}_1}\Big(\sup_{s\le r\le t}\big(M^{\omega_2,\varepsilon}(r)-M^{\omega_2,\varepsilon}(s)\big)\Big)\\&=2\,\mathbb{E}_{\mathbb{P}_1}\Big(\sup_{s\le r\le t}\Big|\int_s^r\mathrm{e}^{-\int_0^u(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\big\langle X^{\omega_2}(u),\sigma(\Lambda^{\omega_2}(u))\mathrm{d}\omega_1(u)\big\rangle\Big|\Big)\\&\le c\,\mathbb{E}_{\mathbb{P}_1}\Big(\int_s^t\mathrm{e}^{-2\int_0^u(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}|X^{\omega_2}(u)|^2\cdot\|\sigma(\Lambda^{\omega_2}(u))\|^2\mathrm{d}u\Big)^{1/2}\\&\le c\,\mathrm{e}^{-\int_0^t(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\mathbb{E}_{\mathbb{P}_1}\Big(\|X_t^{\omega_2}\|_\infty^2\int_s^t\mathrm{e}^{2\int_u^t(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\mathrm{d}u\Big)^{1/2}\\&\le c\,\mathrm{e}^{-\int_0^t(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\big(\mathbb{E}_{\mathbb{P}_1}\|X_t^{\omega_2}\|_\infty^2\big)^{1/2}\\&\le\kappa\,\mathrm{e}^{\widehat{\alpha}\tau}\mathrm{e}^{-\int_0^t(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\mathbb{E}_{\mathbb{P}_1}\|X_t^{\omega_2}\|_\infty^2+\widehat{c}_\kappa\,\mathrm{e}^{-\int_0^t(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\end{aligned} \tag{2.11}$$

for some constant $\widehat{c}_\kappa>0$, where the second identity holds true since $(M^{\omega_2,\varepsilon}(t))$ is a martingale. Then, (2.9) and (2.11) lead to

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

Then, applying Gronwall's inequality yields

$$\begin{aligned}
&\mathrm{e}^{-\int_0^t(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(s)})\mathrm{d}s}\mathbb{E}_{\mathbb{P}_1}\|X_t^{\omega_2}\|_\infty^2\\
&\le c\,\Big\{\,\|\xi\|_\infty^2+\int_0^t\mathrm{e}^{-\int_0^s(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\mathrm{d}s+\mathrm{e}^{-\int_0^t(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}+\mathbb{E}_{\mathbb{P}_1}\varPhi^{\omega_2,\varepsilon}(t)\\
&\quad+\|\xi\|_\infty^2\int_0^t\varTheta^{\omega_2,\varepsilon}(s)\mathrm{e}^{\int_s^t\varTheta^{\omega_2,\varepsilon}(r)\mathrm{d}r}\mathrm{d}s\\
&\quad+\int_0^t\int_0^s\mathrm{e}^{-\int_0^u(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\mathrm{d}u\,\varTheta^{\omega_2,\varepsilon}(s)\mathrm{e}^{\int_s^t\varTheta^{\omega_2,\varepsilon}(r)\mathrm{d}r}\mathrm{d}s\\
&\quad+\int_0^t\mathrm{e}^{-\int_0^s(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\varTheta^{\omega_2,\varepsilon}(s)\mathrm{e}^{\int_s^t\varTheta^{\omega_2,\varepsilon}(r)\mathrm{d}r}\mathrm{d}s\\
&\quad+\int_0^t\big(\mathbb{E}_{\mathbb{P}_1}\varPhi^{\omega_2,\varepsilon}(t)\big)\varTheta^{\omega_2,\varepsilon}(s)\mathrm{e}^{\int_s^t\varTheta^{\omega_2,\varepsilon}(r)\mathrm{d}r}\mathrm{d}s\,\Big\}\\
&=:c\Big(\sum_{j=1}^8\varXi^{j,\omega_2,\varepsilon}(t)\Big),
\end{aligned}\tag{2.12}$$

in which $\varTheta^{\omega_2,\varepsilon}(t):=\mathrm{e}^{-\widehat{\alpha}\tau}\beta^\varepsilon_{\Lambda^{\omega_2}(t)}/(1-\kappa)$. By $u$-substitution and Fubini's theorem, one has

$$\begin{aligned}
&\varXi^{5,\omega_2,\varepsilon}(t)+\varXi^{6,\omega_2,\varepsilon}(t)\\
&=\|\xi\|_\infty^2\Big(\mathrm{e}^{\int_0^t\varXi^{3,\omega_2,\varepsilon}(s)\mathrm{d}s}-1\Big)+\int_0^t\mathrm{e}^{-\int_0^s(\gamma+\alpha^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\Big(\mathrm{e}^{\int_s^t\varXi^{3,\omega_2,\varepsilon}(r)\mathrm{d}r}-1\Big)\mathrm{d}s.
\end{aligned}\tag{2.13}$$

Under (**A**), by using Hölder's inequality and BDG's inequality, it is more or less standard to show

$$\mathbb{E}_{\mathbb{P}_1}\Big(\sup_{0\le s\le t}\|X_s^{\omega_2}\|_\infty^2\Big)\le C_t,\quad t\ge0$$

for some nondecreasing function $t\mapsto C_t$. So, the dominated convergence theorem implies that

$$\varXi^{4,\omega_2,\varepsilon}(t)+\varXi^{8,\omega_2,\varepsilon}(t)\to0,\quad\varepsilon\downarrow0.\tag{2.14}$$

Whereafter, taking (2.13) and (2.14) into account and keeping $\alpha^\varepsilon_{\Lambda^{\omega_2}(t)}\to\alpha_{\Lambda^{\omega_2}(t)}$, $\beta^\varepsilon_{\Lambda^{\omega_2}(t)}\to\beta_{\Lambda^{\omega_2}(t)}$ as $\varepsilon\downarrow0$ in mind, we deduce from (2.12) that

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

Accordingly, by taking $\gamma,\kappa\in(0,1)$ sufficiently small, [4, Theorem 1.5] and [4, Propositions 4.1 and 4.2] imply (2.7) thanks to $\kappa^*>1$ and (2.6). □

We are now in a position to complete the

**Proof of Theorem 1.1.** Recall that $T\ge0$ is the coupling time of $(\Lambda^i(t),\Lambda^j(t))$ defined by

$$T=\inf\big\{t\ge0:\Lambda^i(t)=\Lambda^j(t)\big\}.$$

Below, we set $p \in (0,1]$. For $\beta \in (0,1)$ to be fixed, by Hölder's inequality, it follows that

$$\begin{aligned}\mathbb{W}_{p,d}\big(\delta_{(\xi,i)}P_t,\delta_{(\eta,j)}P_t\big) &\le \mathbb{E}\big\{\|X_t^{\xi,i}-X_t^{\eta,j}\|_\infty^p+\mathbf{1}_{\{\Lambda^i(t)\neq\Lambda^j(t)\}}\big\}\\ &=\mathbb{E}\big\{\big(\|X_t^{\xi,i}-X_t^{\eta,j}\|_\infty^p+\mathbf{1}_{\{\Lambda^i(t)\neq\Lambda^j(t)\}}\big)\mathbf{1}_{\{T\le\beta t\}}\big\}\\ &\quad+\mathbb{E}\big\{\big(\|X_t^{\xi,i}-X_t^{\eta,j}\|_\infty^p+\mathbf{1}_{\{\Lambda^i(t)\neq\Lambda^j(t)\}}\big)\mathbf{1}_{\{T>\beta t\}}\big\}\\ &\le \mathbb{E}\big\{\mathbf{1}_{\{T\le\beta t\}}\mathbb{E}\big(\|X_t^{\xi,i}-X_t^{\eta,j}\|_\infty^p|\mathscr{F}_T\big)\big\}\\ &\quad+\Big\{1+\sqrt{2\big(\mathbb{E}\|X_t^{\xi,i}\|_\infty^{2p}+\mathbb{E}\|X_t^{\eta,j}\|_\infty^{2p}\big)}\Big\}\sqrt{\mathbb{P}(T>\beta t)}\\ &\le c\,\mathbb{E}\big(\mathbf{1}_{\{T\le\beta t\}}\|X_T^{\xi,i}-X_T^{\eta,j}\|_\infty^p\mathrm{e}^{-\frac{p\eta_1}{2}(t-T)}\big)\\ &\quad+c\,(1+\|\xi\|_\infty^p+\|\eta\|_\infty^p)\mathrm{e}^{-\frac12\theta\beta t}\\ &\le c\,(1+\|\xi\|_\infty^p+\|\eta\|_\infty^p)(\mathrm{e}^{-\frac12\theta\beta t}+\mathrm{e}^{-\frac{p\eta_1}{2}(1-\beta)t}),\end{aligned}$$

where we used (1.4), (2.1) as well as (2.7) in the last two step, and utilized (2.7) once more in the last step. Optimizing over $\beta$ in order to guarantee $\theta\beta=p\eta_1(1-\beta)$ (i.e., $\beta=p\eta_1/(\theta+p\eta_1)$), leads to

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

Thus, substituting (2.16) into

$$\mathbb{W}_{p,d}(\nu_1P_t,\nu_2P_t)\le\int_{\mathbf{E}}W_{p,d}(\delta_{(\xi,i)}P_t,\delta_{(\eta,j)}P_t)\pi((\mathrm{d}\xi,\mathrm{d}\{i\}),(\mathrm{d}\eta,\mathrm{d}\{j\})),\quad \pi\in\mathcal{C}(\nu_1,\nu_2)$$

yields the desired assertion (1.5).

For any $\nu\in\mathcal{P}_{p,d}(\mathbf{E})$ and a fixed $t_0$, by the semigroup property of $(P_t)_{t\ge0}$, we derive from (1.5) that

$$\begin{aligned}\mathbb{W}_{p,d}\big(\nu P_{nt_0},\nu P_{(n+1)t_0}\big)&=\mathbb{W}_{p,d}\big(\nu P_{nt_0},(\nu P_{t_0})P_{nt_0}\big)\\ &\le c\Big\{1+\int_{\mathscr{C}}\|\xi\|_\infty^p\nu(\mathrm{d}\xi)+\int_{\mathscr{C}}\|\xi\|_\infty^p(\nu P_{t_0})(\mathrm{d}\xi)\Big\}\mathrm{e}^{-\frac{p\theta\eta_1 nt_0}{2(\theta+p\eta_1)}}.\end{aligned}$$

Whence, $(\nu P_{nt_0})_{n\ge1}$ is a Cauchy sequence in the Polish space $(\mathcal{P}_{p,d}(\mathbf{E}),\mathbb{W}_{p,d})$ so that there exists $\nu_\infty\in\mathcal{P}_{p,d}(\mathbf{E})$ such that $\nu P_{nt_0}$ converges weakly to $\nu_\infty$ as $n\to\infty$. Moreover, $(P_t)_{t\ge0}$ is a Feller process due to Lemma 2.1. Hence, we have $\nu_\infty P_{t_0}=\mu_\infty$. Now, let $\pi=\frac{1}{t_0}\int_0^{t_0}\nu_\infty P_s\mathrm{d}s$. Again, by the semigroup property of $(P_t)_{t\ge0}$, it is easy to see that $\pi\in\mathcal{P}_{p,d}(\mathbf{E})$ is an IPM of $(P_t)_{t\ge0}$. For any IPMs $\nu,\widehat{\nu}\in\mathcal{P}_{p,d}(\mathbf{E})$, we infer from (1.5) that

$$\mathbb{W}_{p,d}(\nu,\widehat{\nu})=\mathbb{W}_{p,d}(\nu P_t,\widehat{\nu}P_t)\le c\,\mathrm{e}^{-\alpha t}$$

for some constant $\alpha>0$. Whence, the uniqueness of IPM follows by approaching $t\uparrow\infty$. Finally, (1.6) follows by just taking $\nu=\delta_{(\xi,i)}$ in (1.5). □

## 3. Proof of Theorem 1.3

Recall $\lambda_i=\alpha_i+\theta_i,\quad \zeta_i=\beta_i+\gamma_i, i\in\mathbf{S}$, where $\alpha_i,\theta_i$ were given in $(\mathbf{H_1})$ and $\beta_i,\gamma_i$ were introduced in $(\mathbf{H_2})$. Below, $\lambda^\varepsilon_{\Lambda^{\omega_2}(t)}$ (resp. $\zeta^\varepsilon_{\Lambda^{\omega_2}(t)}$)be the smooth approximation of $\lambda_{\Lambda^{\omega_2}(t)}$ (resp. $\zeta_{\Lambda^{\omega_2}(t)}$), defined as in (2.2).

The following two lemmas play a crucial role in investigating the long time behavior of $(X_t^{\xi,i},\Lambda^i(t))$.

**Lemma 3.1.** *Under the assumptions of Theorem 1.3, it holds that*

$$\mathbb{E}\|X_t^{\xi,i}-X_t^{\eta,i}\|_\infty^2\le c\,\|\xi-\eta\|_\infty^2e^{-\eta_1 t},\quad(\xi,i),(\eta,i)\in\mathbf{E}, \tag{3.1}$$

*where $\eta_1>0$ is defined in as (1.14) with $p=1$.*

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

$$\mathrm{d}X^{\omega_2}(t) = b(X_t^{\omega_2}, \Lambda^{\omega_2}(t))\mathrm{d}t + \sigma(X_t^{\omega_2}, \Lambda^{\omega_2}(t))\mathrm{d}\omega_1(t), \quad t \geq 0$$

with the initial value $(X_0^{\omega_2}, \Lambda^{\omega_2}(0)) = (\xi, i) \in \mathbf{E}$. Let $\Gamma^{\omega_2}(t)$ be defined as in (2.3) and set

$$\begin{aligned} \Upsilon^{1,\omega_2,\varepsilon}(t) &:= \int_0^t \mathrm{e}^{-\int_0^s \lambda^\varepsilon_{\Lambda^{\omega_2}(r)}\mathrm{d}r} |\lambda_{\Lambda^{\omega_2}(s)} - \lambda^\varepsilon_{\Lambda^{\omega_2}(s)}| \cdot |\Gamma^{\omega_2}(s)|^2 \mathrm{d}s, \\ \Upsilon^{2,\omega_2,\varepsilon}(t) &:= \int_{-\tau}^0 \zeta_{\Lambda^{\omega_2}(t-s)} \mathrm{e}^{-\int_t^{t-s} \lambda^\varepsilon_{\Lambda^{\omega_2}(r)}\mathrm{d}r} \mu(\mathrm{d}s). \end{aligned}$$

By the Itô formula, we deduce from $(\mathbf{H_1})$ and $(\mathbf{H_2})$ that

$$\begin{aligned} &\mathrm{e}^{-\int_0^t \lambda^\varepsilon_{\Lambda^{\omega_2}(s)}\mathrm{d}s} \mathbb{E}_{\mathbb{P}_1}|\Gamma^{\omega_2}(t)|^2 \\ &= |\Gamma^{\omega_2}(0)|^2 + \int_0^t \mathrm{e}^{-\int_0^s \lambda^\varepsilon_{\Lambda^{\omega_2}(r)}\mathrm{d}r} \mathbb{E}_{\mathbb{P}_1} \Big\{ -\lambda^\varepsilon_{\Lambda^{\omega_2}(s)} |\Gamma^{\omega_2}(s)|^2 \\ &\quad + 2\langle \Gamma^{\omega_2}(s), b(X_s^{\omega_2,\xi,i}, \Lambda^{\omega_2}(s)) - b(X_s^{\omega_2,\eta,i}, \Lambda^{\omega_2}(s))\rangle \\ &\quad + \|\sigma(X_s^{\omega_2,\xi,i}, \Lambda^{\omega_2}(s)) - \sigma(X_s^{\omega_2,\eta,i}, \Lambda^{\omega_2}(s))\|^2_{\mathrm{HS}} \Big\} \mathrm{d}s \\ &\leq |\Gamma^{\omega_2}(0)|^2 + \mathbb{E}_{\mathbb{P}_1} \Upsilon^{1,\omega_2,\varepsilon}(t) + \int_0^t \zeta_{\Lambda^{\omega_2}(s)} \mathrm{e}^{-\int_0^s \lambda^\varepsilon_{\Lambda^{\omega_2}(r)}\mathrm{d}r} \int_{-\tau}^0 \mathbb{E}_{\mathbb{P}_1}|\Gamma^{\omega_2}(s+\theta)|^2 \mu(\mathrm{d}\theta)\mathrm{d}s \\ &\leq c\|\Gamma_0^{\omega_2}\|^2_\infty + \mathbb{E}_{\mathbb{P}_1}\Upsilon^{1,\omega_2,\varepsilon}(t) + \int_0^t \Upsilon^{2,\omega_2,\varepsilon}(s) \mathrm{e}^{-\int_0^s \lambda^\varepsilon_{\Lambda^{\omega_2}(r)}\mathrm{d}r} \mathbb{E}_{\mathbb{P}_1}|\Gamma^{\omega_2}(s)|^2 \mathrm{d}s, \end{aligned} \tag{3.2}$$

where in the last step we used the fact that

$$\begin{aligned} &\int_0^t \zeta_{\Lambda^{\omega_2}(s)} \mathrm{e}^{-\int_0^s \lambda^\varepsilon_{\Lambda^{\omega_2}(r)}\mathrm{d}r} \int_{-\tau}^0 \mathbb{E}_{\mathbb{P}_1}|\Gamma^{\omega_2}(s+\theta)|^2 \mu(\mathrm{d}\theta)\mathrm{d}s \\ &= \int_{-\tau}^0 \int_\theta^{t+\theta} \zeta_{\Lambda^{\omega_2}(s-\theta)} \mathrm{e}^{-\int_0^{s-\theta} \lambda^\varepsilon_{\Lambda^{\omega_2}(r)}\mathrm{d}r} \mathbb{E}_{\mathbb{P}_1}|\Gamma^{\omega_2}(s)|^2 \mathrm{d}s\mu(\mathrm{d}\theta) \\ &\leq c\|\Gamma_0^{\omega_2}\|^2_\infty + \int_0^t \Upsilon^{2,\omega_2,\varepsilon}(s) \mathrm{e}^{-\int_0^s \lambda^\varepsilon_{\Lambda^{\omega_2}(r)}\mathrm{d}r} \mathbb{E}_{\mathbb{P}_1}|\Gamma^{\omega_2}(s)|^2 \mathrm{d}s. \end{aligned}$$

Via a standard stopping time argument, we have $\mathbb{E}_{\mathbb{P}_1}|\Gamma^{\omega_2}(t)|^2 < \infty, t \geq 0$. Then, Gronwall's inequality is applicable so that

$$\mathbb{E}_{\mathbb{P}_1}|\Gamma^{\omega_2}(t)|^2 \leq \{c\|\Gamma_0^{\omega_2}\|^2_\infty + \mathbb{E}_{\mathbb{P}_1}\Upsilon^{1,\omega_2,\varepsilon}(t)\} \mathrm{e}^{\int_0^t (\lambda^\varepsilon_{\Lambda^{\omega_2}(s)} + \Upsilon^{2,\omega_2,\varepsilon}(s))\mathrm{d}s}, \tag{3.3}$$

where we also used $t \mapsto \mathbb{E}_{\mathbb{P}_1}\Upsilon^{1,\omega_2,\varepsilon}(t)$ is non-decreasing. Letting $\varepsilon \to 0$ followed by taking expectation w.r.t. $\mathbb{P}_2$ on both sides of (3.3), together with $\int_0^t \lambda^\varepsilon_{\Lambda^{\omega_2}(r)}\mathrm{d}r \to \int_0^t \lambda_{\Lambda^{\omega_2}(r)}\mathrm{d}r$ and $\mathbb{E}_{\mathbb{P}_1}\Upsilon^{1,\omega_2,\varepsilon}(t) \to 0$ as $\varepsilon \downarrow 0$, gives for $\Gamma(t) := X^{\xi,i}(t) - X^{\eta,i}(t)$,

$$\mathbb{E}|\Gamma(t)|^2 \leq c\|\Gamma_0\|^2_\infty \mathbb{E}\exp\bigg(\int_0^t \Big(\lambda_{\Lambda(s)} + \int_{-\tau}^0 \zeta_{\Lambda(s-\theta)} \mathrm{e}^{-\int_s^{s-\theta} \lambda_{\Lambda(r)}\mathrm{d}r} \mu(\mathrm{d}\theta)\Big)\mathrm{d}s\bigg). \tag{3.4}$$

It is easy to see that

$$\begin{aligned} \int_0^t \int_{-\tau}^0 \zeta_{\Lambda(s-\theta)} \mathrm{e}^{-\int_s^{s-\theta} \lambda_{\Lambda(r)}\mathrm{d}r} \mu(\mathrm{d}\theta)\mathrm{d}s &\leq \int_{-\tau}^0 \mathrm{e}^{\theta\widehat{\lambda}} \int_{-\theta}^{t-\theta} \zeta_{\Lambda(s)} \mathrm{d}s\mu(\mathrm{d}\theta) \\ &\leq c + \int_{-\tau}^0 \mathrm{e}^{\theta\widehat{\lambda}} \mu(\mathrm{d}\theta) \int_0^t \zeta_{\Lambda(s)}\mathrm{d}s. \end{aligned}$$

Inserting the estimate above into (3.4), one has

$$\mathbb{E}|\Gamma(t)|^2 \le c\,\|\Gamma_0\|_\infty^2 \mathbb{E}\exp\bigg(\int_0^t \Big(\lambda_{\Lambda(s)} + \int_{-\tau}^0 \mathrm{e}^{\widehat{\lambda}\theta}\mu(\mathrm{d}\theta)\zeta_{\Lambda(s)}\Big)\mathrm{d}s\bigg).$$

Moreover, the prerequisite $\kappa^{**} > 1$ and [4, Propositions 4.2] imply

$$\sum_{j\in\mathbf{S}}\Big(\lambda_j + \zeta_j\int_{-\tau}^0 \mathrm{e}^{\widehat{\lambda}s}\mu(\mathrm{d}s)\Big)\pi_j < 0. \tag{3.5}$$

Thus, [4, Theorem 1.5 & Propositions 4.1] yields

$$\mathbb{E}|\Gamma(t)|^2 \le c\,\mathrm{e}^{-\eta_1 t}\,\|\xi-\eta\|_\infty^2, \quad t\ge 0. \tag{3.6}$$

Next, for any $0\le s\le t$, applying Itô's formula and BDG's inequality and taking advantage of $(\mathbf{H_1})$ and $(\mathbf{H_2})$, we find that

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

which further implies

$$\mathbb{E}\Big(\sup_{s\le r\le t}|\Gamma(r)|^2\Big) \le c\Big\{\mathbb{E}|\Gamma(s)|^2 + \int_{s-\tau}^t \mathbb{E}|\Gamma(r)|^2\mathrm{d}r\Big\}, \quad 0\le s\le t \tag{3.7}$$

owing to $\mathbb{E}\Big(\sup_{s\le r\le t}|\Gamma(r)|^2\Big) < \infty$ by a more or less standard argument. (3.7), together with (3.6), leads to, for $t\ge 2\tau$,

$$\mathbb{E}\|\Gamma_t\|_\infty^2 = \mathbb{E}\Big(\sup_{t-\tau\le s\le t}|\Gamma(s)|^2\Big) \le c\Big\{\mathbb{E}|\Gamma(t-\tau)|^2 + \int_{t-2\tau}^t \mathbb{E}|\Gamma(r)|^2\mathrm{d}r\Big\} \le c\,\mathrm{e}^{-\eta_1 t}\,\|\xi-\eta\|_\infty^2. \tag{3.8}$$

On the other hand, we have for $t\in[0,2\tau]$,

$$\mathbb{E}\|\Gamma_t\|_\infty^2 \le \|\xi-\eta\|_\infty^2 + \mathbb{E}\Big(\sup_{0\le s\le t}|\Gamma(s)|^2\Big) \le c\|\xi-\eta\|_\infty^2 \le c\,\mathrm{e}^{2\eta_1\tau}\mathrm{e}^{-\eta_1 t}\|\xi-\eta\|_\infty^2. \tag{3.9}$$

As a result, (3.1) follows from (3.8) and (3.9). □

**Lemma 3.2.** *Under the assumptions of Theorem 1.3, one has*

$$\mathbb{E}\|X_t^{\xi,i}\|_\infty^2 \le c\,(1+\|\xi\|_\infty^2), \qquad (\xi,i)\in\mathbf{E}. \tag{3.10}$$

**Proof.** By virtue of $(\mathbf{H_1})$ and $(\mathbf{H_2})$, for any $\gamma>0$, there exists a constant $c_\gamma>0$ such that for all $(\xi,j)\in\mathbf{E}$,

$$\begin{aligned}2\langle\xi(0), b(\xi,j)\rangle &\le c_\gamma + (\gamma+\alpha_j)|\xi(0)|^2 + \beta_j\int_{-\tau}^0 |\xi(s)|^2\mu(\mathrm{d}s)\\ \|\sigma(\xi,j)\|_{\mathrm{HS}}^2 &\le c_\gamma + (\gamma+\theta_j)|\xi(0)|^2 + (\gamma+\gamma_j)\int_{-\tau}^0 |\xi(s)|^2\mu(\mathrm{d}s).\end{aligned} \tag{3.11}$$

Next, following the argument to derive (3.2) and making use of (3.11), we infer

$$\mathrm{e}^{-\int_0^t(2\gamma+\lambda^\varepsilon_{\Lambda^{\omega_2}(s)})\mathrm{d}s}\mathbb{E}_{\mathbb{P}_1}|X^{\omega_2}(t)|^2 \le c\,\|\xi\|_\infty^2 + \mathbb{E}_{\mathbb{P}_1}\Psi^{1,\omega_2,\varepsilon}(t) + c\int_0^t \mathrm{e}^{-\int_0^s(2\gamma+\lambda^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\mathrm{d}s + \int_0^t \Psi^{2,\omega_2,\varepsilon}(s)\mathrm{e}^{-\int_0^s(2\gamma+\lambda^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\mathbb{E}_{\mathbb{P}_1}|X^{\omega_2}(s)|^2\mathrm{d}s,$$

where

$$\Psi^{1,\omega_2,\varepsilon}(t) := \int_0^t \mathrm{e}^{-\int_0^s(2\gamma+\lambda^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\Big\{ |\lambda_{\Lambda^{\omega_2}(s)} - \lambda^\varepsilon_{\Lambda^{\omega_2}(s)}|\cdot|X^{\omega_2}(s)|^2 + |\zeta_{\Lambda^{\omega_2}(s)} - \zeta^\varepsilon_{\Lambda^{\omega_2}(s)}|\int_{-\tau}^0 |X^{\omega_2}(s+u)|^2\mu(\mathrm{d}u)\Big\}\mathrm{d}s,$$

$$\Psi^{2,\omega_2,\varepsilon}(t) := \int_{-\tau}^0 \big(\gamma+\zeta^\varepsilon_{\Lambda^{\omega_2}(t-\theta)}\big)\mathrm{e}^{-\int_t^{t-\theta}(2\gamma+\lambda^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\mu(\mathrm{d}\theta).$$

Subsequently, an application of Gronwall's inequality yields

$$\begin{aligned}\mathrm{e}^{-\int_0^t(2\gamma+\lambda^\varepsilon_{\Lambda^{\omega_2}(s)})\mathrm{d}s}\mathbb{E}_{\mathbb{P}_1}|X^{\omega_2}(t)|^2 &\le c\,\|\xi\|_\infty^2 + \mathbb{E}_{\mathbb{P}_1}\Psi^{1,\omega_2,\varepsilon}(t) + c\int_0^t \mathrm{e}^{-\int_0^s(2\gamma+\lambda^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\mathrm{d}s\\ &\quad + \int_0^t\Big(c\,\|\xi\|_\infty^2 + \mathbb{E}_{\mathbb{P}_1}\Psi^{1,\omega_2,\varepsilon}(s) + c\int_0^s \mathrm{e}^{-\int_0^u(2\gamma+\lambda^\varepsilon_{\Lambda^{\omega_2}(r)})\mathrm{d}r}\mathrm{d}u\Big)\\ &\quad \times \Psi^{2,\omega_2,\varepsilon}(s)\exp\Big(\int_s^t \Psi^{2,\omega_2,\varepsilon}(r)\mathrm{d}r\Big)\mathrm{d}s.\end{aligned}$$

Thus, by mimicking the line to derive (2.15), we arrive at

$$\mathbb{E}|X(t)|^2 \le c\,\|\xi\|_\infty^2\mathbb{E}\,\mathrm{e}^{\int_0^t(2\gamma+\lambda_{\Lambda(s)}+\Pi(s))\mathrm{d}s} + c\int_0^t \mathbb{E}\mathrm{e}^{\int_s^t(2\gamma+\lambda_{\Lambda(u)}+\Pi(u))\mathrm{d}u}\mathrm{d}s, \tag{3.12}$$

where

$$\Pi(t) := \int_{-\tau}^0(\gamma+\zeta_{\Lambda(t-\theta)})\mathrm{e}^{-\int_t^{t-\theta}(2\gamma+\lambda_{\Lambda(r)})\mathrm{d}r}\mu(\mathrm{d}\theta), \qquad t>0.$$

Plugging the fact that

$$\int_s^t \Pi(r)\mathrm{d}r \le c + \mu(\mathrm{e}^{\widehat{\lambda}\cdot})\int_s^t(\gamma+\zeta_{\Lambda(r)})\mathrm{d}r, \quad \mu(\mathrm{e}^{\widehat{\lambda}\cdot}) := \int_{-\tau}^0 \mathrm{e}^{\widehat{\lambda}\theta}\mu(\mathrm{d}\theta)$$

into (3.12) yields

$$\mathbb{E}|X(t)|^2 \le c\,\|\xi\|_\infty^2\mathbb{E}\exp\Big(\int_0^t(c\gamma+\lambda_{\Lambda(s)}+\mu(\mathrm{e}^{\widehat{\lambda}\cdot})\zeta_{\Lambda(s)})\mathrm{d}s\Big) + c\int_0^t\mathbb{E}\exp\Big(\int_s^t(c\gamma+\lambda_{\Lambda(r)}+\mu(\mathrm{e}^{\widehat{\lambda}\cdot})\zeta_{\Lambda(r)})\mathrm{d}r\Big)\mathrm{d}s$$

for some $c>0$. Thus, with the aid of [4, Theorem 1.5 & Propositions 4.1 and 4.2] and by choosing $\gamma>0$ such that $c\gamma=\eta_1/2$, we obtain from $\kappa^{**}>1$ as well as (3.5) that

$$\mathbb{E}|X(t)|^2 \le c\,\|\xi\|_\infty^2\mathrm{e}^{-\eta_1 t/2} + c\int_0^t \mathrm{e}^{-\frac{\eta_1}{2}s}\mathrm{d}s \le c(1+\|\xi\|_\infty^2). \tag{3.13}$$

Carrying out an analogous manner to derive (3.7), we have

$$\mathbb{E}\Big(\sup_{s\le r\le t}|X(r)|^2\Big) \le c\Big\{1+\|\xi\|_\infty^2+\mathbb{E}|X(s)|^2+\int_{s-\tau}^t\mathbb{E}|X(r)|^2\mathrm{d}r\Big\}, \quad 0\le s\le t. \tag{3.14}$$

Thereby, (3.10) is now available from (3.13) and (3.14). □

**Proof of Theorem 1.3.** With the aid of Lemmas 3.1 and 3.2, the argument of Theorem 1.3 can be done by repeating the proof of Theorem 1.1 so that we do not go into detail about the corresponding proof. □

## 4. Proof of Theorem 1.4

For $K:\mathbf{S}\to\mathbb{R}$ and $p\ge 0$, set

$$Q_p := Q + p\,\mathrm{diag}(K_1,\dots,K_N),\ \ \eta_p := \underset{\gamma\in\mathrm{spec}(Q_p)}{-\max}\ \mathrm{Re}(\gamma), \kappa^{\star\star} := \sup\{p\ge 0:\eta_p>0\}.$$

The lemma below, which is concerned with an estimate on the exponential functional of the discrete observation for the Markov chain involved and may be interesting by itself, plays a crucial role in the analyzing the long-time behavior of the discretization for $(X_t^{\xi,i},\Lambda^i(t))$.

**Lemma 4.1.** *Assume $\kappa^{\star\star}>1$. Then, there exist constants $\delta\in(0,1)$ and $\lambda>0$ such that*

$$\mathbb{E}\, e^{\int_0^t K_{\Lambda(s_\delta)}\,ds} \le c\, e^{-\lambda t},\quad t\ge 0. \tag{4.1}$$

**Proof.** By Hölder's inequality, it follows that

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

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

$$\begin{aligned}&\mathbb{P}(\Lambda(t+\triangle)=j|\Lambda(t)=j)=1+q_{jj}\triangle+o(\triangle),\\ &\mathbb{P}(\Lambda(t+\triangle)\ne j|\Lambda(t)=j)\le\kappa_0\triangle+o(\triangle),\end{aligned} \tag{4.3}$$

where $\kappa_0:=\max_{k\in\mathbf{S}}(-q_{kk})$. Utilizing Jensen's inequality and taking advantage of (4.3), we derive

$$\begin{aligned}
&\mathbb{E}\Big(\mathrm{e}^{\frac{1+\varepsilon}{\varepsilon}\int_{j\delta}^{(j+1)\delta\wedge t}(K_{\Lambda(j\delta)}-K_{\Lambda(s)})\mathrm{d}s}\Big|\Lambda(j\delta)\Big)\\
&=\frac{\sum_{k\in\mathbf{S}}\mathbf{1}_{\{\Lambda(j\delta)=k\}}}{(j+1)\delta\wedge t-j\delta}\int_{j\delta}^{(j+1)\delta\wedge t}\mathbb{E}\Big(\mathrm{e}^{\frac{1+\varepsilon}{\varepsilon}((j+1)\delta\wedge t-i\delta)(K_j-K_{\Lambda(s)})}\Big|\Lambda(j\delta)=k\Big)\mathrm{d}s\\
&=\frac{\sum_{k\in\mathbf{S}}\mathbf{1}_{\{\Lambda(j\delta)=k\}}}{(j+1)\delta\wedge t-j\delta}\int_{j\delta}^{(j+1)\delta\wedge t}\mathbb{E}\,(\mathbf{1}_{\{\Lambda(s)=k\}}|\Lambda(j\delta)=k)\mathrm{d}s\\
&\quad+\frac{\sum_{k\in\mathbf{S}}\mathbf{1}_{\{\Lambda(j\delta)=k\}}}{(j+1)\delta\wedge t-j\delta}\int_{i\delta}^{(j+1)\delta\wedge t}\mathbb{E}\Big(\mathrm{e}^{\frac{1+\varepsilon}{\varepsilon}((j+1)\delta\wedge t-j\delta)(K_j-K_{\Lambda(s)})}\mathbf{1}_{\{\Lambda(s)\ne j\}}\Big|\Lambda(j\delta)=j\Big)\mathrm{d}s\\
&\le\frac{\sum_{j\in\mathbf{S}}\mathbf{1}_{\{\Lambda(j\delta)=k\}}}{(j+1)\delta\wedge t-j\delta}\int_{j\delta}^{(j+1)\delta\wedge t}\mathbb{E}\,(\mathbf{1}_{\{\Lambda(s)=k\}}|\Lambda(j\delta)=k)\mathrm{d}s\\
&\quad+\mathrm{e}^{\frac{2(1+\varepsilon)\check K\delta}{\varepsilon}}\frac{\sum_{k\in\mathbf{S}}\mathbf{1}_{\{\Lambda(j\delta)=k\}}}{(j+1)\delta\wedge t-j\delta}\int_{j\delta}^{(j+1)\delta\wedge t}\mathbb{P}\,(\mathbf{1}_{\{\Lambda(s)\ne j\}}|\Lambda(j\delta)=k)\mathrm{d}s\\
&\le\frac{\sum_{k\in\mathbf{S}}\mathbf{1}_{\{\Lambda(j\delta)=k\}}}{(j+1)\delta\wedge t-j\delta}\int_{j\delta}^{(j+1)\delta\wedge t}(1+q_{kk}(s-j\delta)+o(s-j\delta))\mathrm{d}s\\
&\quad+\mathrm{e}^{\frac{2(1+\varepsilon)\check K\delta}{\varepsilon}}\frac{\sum_{k\in\mathbf{S}}\mathbf{1}_{\{\Lambda(j\delta)=k\}}}{(j+1)\delta\wedge t-j\delta}\int_{j\delta}^{(j+1)\delta\wedge t}(2\kappa_0(s-i\delta)+o(s-j\delta))\mathrm{d}s\\
&\le 1+\kappa_0\delta\,\mathrm{e}^{\frac{2(1+\varepsilon)\check\alpha\delta}{\varepsilon}}+o(\delta),\quad \delta\in(0,\delta_1).
\end{aligned} \tag{4.4}$$

By the property of conditional expectation, we deduce from (4.4) that

$$\begin{aligned}
&\mathbb{E}\,\mathrm{e}^{\frac{1+\varepsilon}{\varepsilon}\int_0^t (K_{\Lambda(s_\delta)}-K_{\Lambda(s)})\mathrm{d}s}\\
&=\mathbb{E}\Big(\mathrm{e}^{\frac{1+\varepsilon}{\varepsilon}\sum_{j=0}^{\lfloor t/\delta\rfloor-1}\int_{j\delta}^{(j+1)\delta}(K_{\Lambda(j\delta)}-K_{\Lambda(s)})\mathrm{d}s}\,\mathbb{E}\Big(\mathrm{e}^{\frac{1+\varepsilon}{\varepsilon}\int_{t_\delta}^{(t_\delta+\delta)\wedge t}(K_{\Lambda(t_\delta)}-K_{\Lambda(s)})\mathrm{d}s}\Big|\Lambda(t_\delta)\Big)\Big)\\
&\le\Big(1+\kappa_0\delta\,\mathrm{e}^{\frac{2(1+\varepsilon)\check{K}\delta}{\varepsilon}}+o(\delta)\Big)\mathbb{E}\Big(\mathrm{e}^{\frac{1+\varepsilon}{\varepsilon}\sum_{j=0}^{\lfloor t/\delta\rfloor-1}\int_{j\delta}^{(j+1)\delta}(K_{\Lambda(j\delta)}-K_{\Lambda(s)})\mathrm{d}s}\Big)\\
&\le\cdots\\
&\le\Big(1+\kappa_0\delta\mathrm{e}^{\frac{2(1+\varepsilon)\check{K}\delta}{\varepsilon}}+o(\delta)\Big)^{\lfloor t/\delta\rfloor+1},\qquad \delta\in(0,\delta_1).
\end{aligned}\tag{4.5}$$

For any $c_1,c_2>0$, by L'Hospital's rule,

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

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

$$\ln(1+c_1\delta\,\mathrm{e}^{c_2\delta})\le 2\,c_1\delta,\qquad \delta\in(0,\delta_2).\tag{4.6}$$

According to (4.6), for any $\delta\in(0,\delta_1\wedge\delta_2)$,

$$\begin{aligned}
\Big(1+\kappa_0\delta\,\mathrm{e}^{\frac{2(1+\varepsilon)\check{K}\delta}{\varepsilon}}+o(\delta)\Big)^{\frac{\varepsilon(\lfloor t/\delta\rfloor+1)}{1+\varepsilon}}&\le\exp\Big(\frac{\varepsilon(t+1)}{\delta(1+\varepsilon)}\ln\Big(1+\kappa_0\delta\,\mathrm{e}^{\frac{2(1+\varepsilon)\check{K}\delta}{\varepsilon}}+o(\delta)\Big)\Big)\\
&\le\mathrm{e}^{\kappa_\varepsilon(1+t)}
\end{aligned}\tag{4.7}$$

with $\kappa_\varepsilon:=2\kappa_0\varepsilon/(1+\varepsilon)$. Taking (4.5) and (4.7) into consideration, we deduce from (4.2) that

$$\mathbb{E}\,\mathrm{e}^{\int_0^t K_{\Lambda(s_\delta)}\mathrm{d}s}\le\mathrm{e}^{\kappa_\varepsilon(1+t)}\Big(\mathbb{E}\,\mathrm{e}^{(1+\varepsilon)\int_0^t K_{\Lambda(s)}\mathrm{d}s}\Big)^{\frac{1}{1+\varepsilon}}.\tag{4.8}$$

Using $\kappa^{\star\star}>1$ and [4, Propositions 4.2], we have

$$\sum_{j\in\mathbf{S}}K_j\pi_j<0.\tag{4.9}$$

This, together with [4, Theorem 1.5 & Propositions 4.1], implies that there exist $\varepsilon_0\in(0,1)$ sufficiently small and $\lambda>0$ such that

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

Inserting this into (4.8) yields that

$$\mathbb{E}\,\mathrm{e}^{\int_0^t K_{\Lambda(s_\delta)}\mathrm{d}s}\le\mathrm{e}^{\kappa_\varepsilon(1+t)-\frac{\lambda t}{1+\varepsilon}}.$$

Thus, the desired assertion follows by taking $0<\varepsilon<\varepsilon_0\wedge(2\kappa_0)^{-1}\lambda$. □

Next, we provide two crucial lemmas, where one of them is concerned with the distance in $L^2$-norm sense of the $\mathscr{C}$-valued stochastic process $Y_{t_\delta}$ starting from different points and another one is related to the uniform boundedness in $L^2$-norm sense.

**Lemma 4.2.** *Assume the assumptions of Theorem 1.1 and (1.19). Then, there exist $\delta_0\in(0,1)$ and $\alpha>0$ such that*

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

**Proof.** Hereinafter, we assume $t \ge \tau + 1$. 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 notation brevity, set

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

First of all, we verify

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

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

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

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

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

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

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

Thereby, (4.12) follows by applying (4.12) with $t = k\delta$. It is easy to see from (1.19) that

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

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

$$\begin{aligned} &\mathrm{e}^{-\int_0^t \alpha_{\Lambda^{\omega_2}(s_\delta)}\mathrm{d}s}|\Upsilon^{\omega_2}(t)|^2 \\ &\le |\Upsilon^{\omega_2}(0)|^2 + \int_0^t \mathrm{e}^{-\int_0^s \alpha_{\Lambda^{\omega_2}(r_\delta)}\mathrm{d}r}\Big\{ \alpha_{\Lambda^{\omega_2}(s_\delta)}(|\Upsilon^{\omega_2}(s_\delta)|^2 - |\Upsilon^{\omega_2}(s)|^2) + \beta_{\Lambda^{\omega_2}(s_\delta)}\|\Upsilon^{\omega_2}_{s_\delta}\|_\infty^2 \\ &\quad + 2\langle \Upsilon^{\omega_2}(s) - \Upsilon^{\omega_2}(s_\delta), b(Y^{\omega_2,\xi,i}_{s_\delta}, \Lambda^{\omega_2}(s_\delta)) - b(Y^{\omega_2,\eta,i}_{s_\delta}, \Lambda^{\omega_2}(s_\delta))\rangle \Big\}\mathrm{d}s \\ &\le |\Upsilon^{\omega_2}(0)|^2 + \int_0^t \mathrm{e}^{-\int_0^s \alpha_{\Lambda^{\omega_2}(r_\delta)}\mathrm{d}r}\Big\{ \frac{c}{\sqrt{\delta}}|\Upsilon^{\omega_2}(s) - \Upsilon^{\omega_2}(s_\delta)|^2 \\ &\quad + (c\sqrt{\delta} + \beta_{\Lambda_0^{\omega_2}(s_\delta)})\sup_{s-\tau-\delta\le r\le s}|\Upsilon^{\omega_2}(s)|^2 \Big\}\mathrm{d}s \\ &\le |\Upsilon^{\omega_2}(0)|^2 + \int_0^t (c\sqrt{\delta} + \beta_{\Lambda^{\omega_2}(s_\delta)})\mathrm{e}^{-\int_0^s \alpha_{\Lambda^{\omega_2}(r_\delta)}\mathrm{d}r}\sup_{s-\tau-\delta\le r\le s}|\Upsilon^{\omega_2}(s)|^2\mathrm{d}s. \end{aligned} \tag{4.14}$$

where in the penultimate display we used (4.13). Observe that

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

We therefore obtain from (4.14) that

$$\varPi^{\omega_2}(t)\le c\,\|\xi-\eta\|_\infty^2+\mathrm{e}^{-\widehat{a}(\tau+\delta)}\int_0^t\big(c\sqrt{\delta}+\beta_{\Lambda^{\omega_2}(s_\delta)}\big)\varPi^{\omega_2}(s)\mathrm{d}s.$$

This, together with Gronwall's inequality, implies that

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

Owing to (2.6), we can choose some $\delta_1\in(0,1)$ such that

$$\sum_{j\in\mathbf{S}}\big(\alpha_j+\mathrm{e}^{-\widehat{a}(\tau+\delta)}\beta_j\big)\pi_j<0,\qquad\delta\in(0,\delta_1).\tag{4.15}$$

Thus, according to Lemma 4.1, (4.15) as well as $\kappa^*>1$, there exists $\eta_1>0$ such that

$$\mathbb{E}\|Y_{t_\delta}^{\xi,i}-Y_{t_\delta}^{\eta,i}\|_\infty^2\le c\,\|\xi-\eta\|_\infty^2\mathrm{e}^{c\,\mathrm{e}^{-\widehat{a}(\tau+\delta)}\sqrt{\delta}t}\mathrm{e}^{-\eta_1t}.$$

Furthermore, take $\delta_2\in(0,\delta_1)$ such that $c\,\mathrm{e}^{-\widehat{a}(\tau+\delta_2)}\sqrt{\delta_2}<\eta_1$. As a consequence, (4.10) follows for any $\delta\in(0,\delta_2)$. □

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

$$\mathbb{E}\|Y_{t_\delta}^{\xi,i}\|_\infty^2\le c\,(1+\|\xi\|_\infty^2),\qquad t\ge\tau+1,\quad\delta\in(0,\delta_0),\quad(\xi,i)\in\mathbf{E}.\tag{4.16}$$

**Proof.** Below, we assume $t\ge\tau+1$ and, for notation brevity, write $Y^{\omega_2}$ instead of $Y^{\omega_2,\xi,i}$. Carrying out the procedure to gain (4.12), we obtain from (2.8) that

$$\begin{aligned}&\mathrm{e}^{-\int_0^t(\gamma+\alpha_{\Lambda^{\omega_2}(s_\delta)})\mathrm{d}s}|Y^{\omega_2}(t)|^2\\&=|\xi(0)|^2+\int_0^t\mathrm{e}^{-\int_0^s(\gamma+\alpha_{\Lambda^{\omega_2}(r_\delta)})\mathrm{d}r}\Big\{-(\gamma+\alpha_{\Lambda^{\omega_2}(s_\delta)})|Y^{\omega_2}(s)|^2\\&\quad+2\langle Y^{\omega_2}(s),b(Y_{s_\delta}^{\omega_2},\Lambda^{\omega_2}(s_\delta))\rangle+\|\sigma(\Lambda^{\omega_2}(s_\delta))\|_{\mathrm{HS}}^2\Big\}\mathrm{d}s+\varTheta^{\omega_2}(t),\end{aligned}\tag{4.17}$$

where

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

Thanks to (1.19), it follows that

$$|Y^{\omega_2}(t)-Y^{\omega_2}(t_\delta)|^2\le c\delta+4\delta L_0^2\|Y_{t_\delta}^{\omega_2}\|_\infty^2+c\,|\omega_1(t)-\omega_1(t_\delta)|^2.\tag{4.18}$$

Thus, by combining (2.8) with (4.17), it follows that

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

Following the argument to derive (2.11), 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

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

where

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

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

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

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

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

in which

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

By an inductive argument, we derive that for all $0\le s\le t$,

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

Subsequently, taking (4.21), (4.22) 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

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

Thus, with the help of $\kappa^* > 1$ and (4.16) follows from Lemma 4.1 and by taking $\gamma, \delta, \kappa \in (0, 1)$ sufficiently small. □

**Proof of Theorem 1.4.** With Lemmas 4.2 and 4.3 and Lemma 5.1 below at hand, we can complete the argument of Theorem 1.4 by mimicking the proof of Theorem 1.3. □

**Proof of Corollary 1.5.** For any $k \geq 1$, by the triangle inequality, we have

$$\mathbb{W}_{p,d}\big(\nu, \nu^{(\delta)}\big) \leq \mathbb{W}_{p,d}\big(\nu, \delta_{(\xi,i)}P_{k\delta}\big) + \mathbb{W}_{p,d}\big(\nu^{(\delta)}, \delta_{(\xi,i)}P^{(\delta)}_{k\delta}\big) + \mathbb{W}_{p,d}\big(\delta_{(\xi,i)}P_{k\delta}, \delta_{(\xi,i)}P^{(\delta)}_{k\delta}\big).$$

In terms of Theorems 1.1 and 1.4, there exist constants $\alpha > 0$ and $c_1$, which is independent of $k$ and $\delta$, such that

$$\mathbb{W}_{p,d}\big(\nu, \delta_{(\xi,i)}P_{k\delta}\big) + \mathbb{W}_{p,d}\big(\nu^{(\delta)}, \delta_{(\xi,i)}P^{(\delta)}_{k\delta}\big) \leq c\,\big(1 + \|\xi\|^p_\infty\big)\mathrm{e}^{-\alpha k\delta}. \tag{4.23}$$

Moreover, by a more or less standard argument, there exists a constant $c_2$, independent of $k$ and $\delta$ such that

$$\mathbb{W}_{p,d}\big(\delta_{(\xi,i)}P_{k\delta}, \delta_{(\xi,i)}P^{(\delta)}_{k\delta}\big) \leq c_2\big(1 + \|\xi\|^p_\infty\big)\mathrm{e}^{c_2(k\delta)^2}\delta^{p/2}. \tag{4.24}$$

In particular, in (4.23) and (4.24), taking

$$k = \Big\lfloor \frac{1}{\sqrt{c_2}\delta}\big(\ln \delta^{-\frac{p}{4}}\big)^{\frac{1}{2}}\Big\rfloor$$

and approaching $\delta \to 0$ yields the desired assertion. □

## 5. Proof of Theorem 1.6

Before we complete the proof of Theorem 1.6, let us make some preparations. For any $t \geq 0$, let $\mathcal{F}_t = \sigma((W(u), \Lambda(u)), 0 \leq u \leq t) \vee \mathcal{N}$, 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., for any $A \in \mathscr{B}(\mathscr{C})$ and $(\xi, i) \in \mathbf{E}$,*

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

*and*

$$\begin{aligned}&\mathbb{P}\big((Y_{(k+1)\delta}, \Lambda((k+1)\delta)) \in A \times \{i\}|\mathscr{F}_{k\delta}\big)\\ &= \mathbb{P}\big((Y_{(k+1)\delta}, \Lambda((k+1)\delta)) \in A \times \{i\}|(Y_{k\delta}, \Lambda(k\delta))\big).\end{aligned} \tag{5.2}$$

**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 (1.18) that

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

Observe from (1.22) and (5.3) that

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

and that

$$Y_{(k+1)\delta}(\theta)=\begin{cases}Y_{k\delta}(0)+\frac{\theta+\delta}{\delta}\big\{Y((k+1)\delta)-Y_{k\delta}(0)\big\}, & \theta\in[-\delta,0]\\ Y_{k\delta}((1+i)\delta)+\frac{\theta-i\delta}{\delta}\big\{Y_{k\delta}((2+i)\delta)-Y_{k\delta}((1+i)\delta)\big\}, & \theta\in[i\delta,(i+1)\delta],i\neq-1\end{cases}$$
$$=\begin{cases}Y_{k\delta}(0)+\frac{\theta+\delta}{\delta}\Big\{\,b(Y_{k\delta},\Lambda(k\delta))\delta\\ \qquad+\sigma(Y_{k\delta},\Lambda(k\delta))(W((k+1)\delta)-W(k\delta))\,\Big\}, & \theta\in[-\delta,0]\\ Y_{k\delta}((1+i)\delta)+\frac{\theta-i\delta}{\delta}\big\{Y_{k\delta}((2+i)\delta)-Y_{k\delta}((1+i)\delta)\big\}, & \theta\in[i\delta,(i+1)\delta],i\neq-1.\end{cases} \tag{5.5}$$

Thus, comparing (5.4) 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 $(Y_{k\delta},\Lambda(k\delta))=(Y_0,\Lambda(0))=(\xi,i)$. Therefore, (5.1) holds immediately.

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

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

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 $(Y_{k\delta},\Lambda(k\delta))$. Applying [16, Theorem 2.24, p.46] and taking (5.6) into consideration yields that

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

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

**Lemma 5.2.** *Assume* $(\mathbf{H_1})$, $(\mathbf{H_2})$, (1.21) *and suppose further* $\kappa^\star>1$. *Then, there exist* $\delta_0\in(0,1)$ *and* $\alpha>0$ *such that*

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

**Proof.** For fixed $\omega_2$, let $Y^{\omega_2}$ solve the following SDE

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

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 $(\mathbf{H_2})$ and (1.19), it is easy to see that

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

Following the procedure to derive (4.12), we obtain from $(\mathbf{H_1})$, $(\mathbf{H_2})$, (1.21) and (5.8) that

$$\begin{aligned} &\mathrm{e}^{-\int_0^t \lambda_{\Lambda^{\omega_2}(s_\delta)}\mathrm{d}s}\mathbb{E}_{\mathbb{P}_1}|\Upsilon^{\omega_2}(t)|^2 \\ &\le |\Upsilon^{\omega_2}(0)|^2 + \int_0^t \mathrm{e}^{-\int_0^s \lambda_{\Lambda^{\omega_2}(r_\delta)}\mathrm{d}r}\Big\{ c\sqrt{\delta}\mathbb{E}_{\mathbb{P}_1}|\Upsilon^{\omega_2}(s_\delta)|^2 \\ &\quad + \big(c+\zeta_{\Lambda^{\omega_2}(s_\delta)}\big)\int_{-\tau}^0 \mathbb{E}_{\mathbb{P}_1}|\Upsilon^{\omega_2}_{s_\delta}(\theta)|^2\mu(\mathrm{d}\theta) + \frac{c}{\sqrt{\delta}}\mathbb{E}_{\mathbb{P}_1}|\Upsilon^{\omega_2}(s)-\Upsilon^{\omega_2}(s_\delta)|^2\Big\}\,\mathrm{d}s \\ &\le |\Upsilon^{\omega_2}(0)|^2 + c\sqrt{\delta}\int_0^t \mathrm{e}^{-\int_0^s \lambda_{\Lambda^{\omega_2}(r_\delta)}\mathrm{d}r}\mathbb{E}_{\mathbb{P}_1}|\Upsilon^{\omega_2}(s_\delta)|^2\mathrm{d}s + \Psi^{\omega_2}(t), \end{aligned} \tag{5.9}$$

where

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

By virtue of (5.4), we deduce that

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

where

$$\Theta^{\omega_2}(t) := 2\mathrm{e}^{-\widehat{a}\tau}\sum_{j=-N}^{-1}\int_{j\delta}^{(j+1)\delta}\big\{c\sqrt{\delta}+\zeta_{\Lambda^{\omega_2}(t_\delta-i\delta)}+\zeta_{\Lambda^{\omega_2}(t_\delta-(j+1)\delta)}\big\}\mu(\mathrm{d}\theta). \tag{5.11}$$

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

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

This, together with the fact that

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

implies that

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

Thus, an application of Gronwall's inequality leads to

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

Furthermore, observe that

$$\begin{aligned} \int_0^t \Theta^{\omega_2}(s)\mathrm{d}s &= 2\,\mathrm{e}^{-\widehat{\lambda}\tau}\sum_{j=-N}^{-1}\int_{j\delta}^{(j+1)\delta}\int_{-j\delta}^{t-j\delta}\big\{c\sqrt{\delta}+\zeta_{\Lambda^{\omega_2}(s_\delta)}\big\}\mathrm{d}s\mu(\mathrm{d}\theta) \\ &\quad + 2\,\mathrm{e}^{-\widehat{\lambda}\tau}\sum_{j=-N}^{-1}\int_{i\delta}^{(j+1)\delta}\int_{-(j+1)\delta}^{t-(j+1)\delta}\zeta_{\Lambda^{\omega_2}(s_\delta)}\mathrm{d}s\mu(\mathrm{d}\theta) \\ &\le c + 4\,\mathrm{e}^{-\widehat{\lambda}\tau}\int_0^t \big\{c\sqrt{\delta}+\zeta_{\Lambda^{\omega_2}(s_\delta)}\big\}\mathrm{d}s. \end{aligned} \tag{5.14}$$

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

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

By applying Lemma 4.1 and $\kappa^\star > 1$, there exist $\delta_0 \in (0,1)$ and $\alpha > 0$ such that

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

With $(\mathbf{H_2})$ and (5.15) in hand, (5.7) can be obtained via a standard procedure. □

**Lemma 5.3.** *Under the assumptions of Lemma 5.2, there exists some $\delta_0 \in (0,1)$ such that*

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

**Proof.** Mimicking the procedure to derive (4.17), we have

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

where $\gamma > 0$ is introduced in (3.11). By $(\mathbf{H_2})$ and (1.21), it follows that

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

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

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

where

$$\Psi_0^{\omega_2}(t) := \int_0^t \big(c\sqrt{\delta}+\zeta_{\Lambda^{\omega_2}(s_\delta)}\big)\mathrm{e}^{-\int_0^s (2\gamma+\lambda_{\Lambda^{\omega_2}(r_\delta)})\mathrm{d}r}\int_{-\tau}^0 \mathbb{E}_{\mathbb{P}_1}|Y_{s_\delta}^{\omega_2}(\theta)|^2\mu(\mathrm{d}\theta)\mathrm{d}s.$$

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

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

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

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

This, applying the Gronwall inequality and utilizing (5.12) with $\Upsilon^{\omega_2}$ being replaced by $Y^{\omega_2}$ enables us to obtain

$$\Pi^{\omega_2}(t) \le c\,\|\xi\|_\infty^2 + c\int_0^t \mathrm{e}^{-\int_0^s (2\gamma+\lambda_{\Lambda^{\omega_2}(r_\delta)})\mathrm{d}r}\mathrm{d}s + c\,\mathrm{e}^{-\widehat{\lambda}\delta}\int_0^t\Big\{\|\xi\|_\infty^2 + \int_0^s \mathrm{e}^{-\int_0^u (2\gamma+\lambda_{\Lambda^{\omega_2}(r_\delta)})\mathrm{d}r}\mathrm{d}u\Big\}$$
$$\times\big(c\sqrt{\delta}+\Theta^{\omega_2}(s)\big)\mathrm{e}^{\int_s^t \mathrm{e}^{-\widehat{a}\delta}(c\sqrt{\delta}+\Theta^{\omega_2}(r))\mathrm{d}r}\mathrm{d}s.$$

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

**Proof of Theorem 1.6.** With the help of Lemmas 5.1–5.3, we can finish the proof by following the argument of Theorem 1.3. □


## Acknowledgments

We would like to express our great gratitude to the anonymous referee for the insightful comments, which improve considerably our work.


## References


[1] W.J. Anderson, Continuous-Time Markov Chains, Springer-Verlag, New York, 1991.
[2] Y. Bakhtin, T. Hurth, Invariant densities for dynamical systems with random switching, Nonlinearity 25 (2012) 2937–2953.
[3] J. Bao, J. Shao, C. Yuan, Approximation of invariant measures for regime-switching diffusions, Potential Anal. 44 (2016) 707–727.
[4] J.B. Bardet, H. Guérin, F. Malrieu, Long time behavior of diffusions with Markov switching, ALEA Lat. Am. J. Probab. Math. Stat. 7 (2010) 151–170.
[5] M. Benaim, S. Le Borgne, F. Malrieu, P.A. Zitt, Quantitative ergodicity for some switched dynamical systems, Electron. Commun. Probab. 17 (2012) 14.
[6] B. Cloez, M. Hairer, Exponential ergodicity for Markov processes with random switching, Bernoulli 21 (2015) 505–536.
[7] B. de Saporta, J. Yao, Tail of a linear diffusion with Markov switching, Ann. Appl. Probab. 15 (2005) 992–1018.
[8] R. Douc, A. Fort, A. Guillin, Subgeometric rates of convergence of $f$-ergodic strong Markov processes, Stochastic Process. Appl. 119 (2009) 897–923.
[9] E. Eberle, Reflection couplings and wasserstein contractivity with convexity, Probab. Theory Related Feilds 166 (2016) 851–886.
[10] A. Eberle, A. Guillin, R. Zimmer, Quantitative Harris-type theorem for diffusions and McKean-Vlasov processes, Trans. Amer. Math. Soc. 371 (2019) 7135–7173.
[11] A.A. Gushchin, U. Küchler, On stationary solutions of delay differential equations driven by a Lévy process, Stochastic Process. Appl. 88 (2000) 195–211.
[12] M. Hairer, J.C. Mattingly, M. Scheutzow, Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations, Probab. Theory Related Fields 149 (2011) 223–259.
[13] M. Hutzenthaler, A. Jentzen, P.E. Kloeden, Strong and weak divergence in finite time of Euler's method for stochastic differential equations with non-globally Lipschitz continuous coefficients, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 467 (2011) 1563–1576.
[14] N. Ikeda, S. Watanabe, Stochastic Differential Equations and Diffusion Processes, North-Holland, 1989.
[15] K. Itô, M. Nisio, On stationary solutions of a stochastic differential equation, J. Math. Kyoto Univ. 4 (1964) 1–75.
[16] Fima C. Klebaner, Introduction to Stochastic Calculus with Applications, third ed., Imperial College Press, London, 2012.
[17] U. Küchler, B. Mensch, Langevins stochastic differential equation extended by a time-delayed term, Stochastics 40 (1992) 23–42.
[18] X. Li, Q. Ma, H. Yang, C. Yuan, The numerical invariant measure of stochastic differential equations with Markovian switching, SIAM J. Numer. Anal. 56 (2018) 1435–1455.
[19] Z.-W. Liao, J. Shao, Long time behavior of Lévy-driven Ornstein–Uhlenbeck process with regime switching, arXiv:1906.08426.
[20] D. Luo, J. Wang, Exponential convergence in $L^p$-Wasserstein distance for diffusion processes with uniformly dissipative drift, Math. Nachr. 289 (2016) 1909–1926.
[21] A.J. Majda, X. Tong, Geometric ergodicity for piecewise contracting processes with applications for tropical stochastic lattice models, Comm. Pure Appl. Math. 69 (2016) 1110–1153.
[22] X. Mao, C. Yuan, Stochastic Differential Equations with Markovian Switching, Imperial College Press, London, 2006.

[23] S-E.A. Mohammed, Stochastic Functional Differential Equations, Pitman, Boston, 1984.
[24] M. Pinsky, R. Pinsky, Transience recurrence and central limit theorem behavior for diffusions in random temporal enrivoments, Ann. Probab. 21 (1993) 433–452.
[25] M. Pinsky, M. Scheutzow, Some remarks and examples concerning the transience and recurrence of random diffusions, Ann. Inst. H. Poincaré 28 (1992) 519–536.
[26] M. Reiß, M. Riedle, O. van Gaans, Delay differential equations driven by Lévy processes: stationarity and feller properties, Stochastic Process. Appl. 116 (2006) 1409–1432.
[27] J. Shao, Ergodicity of regime-switching diffusions in Wasserstein distances, Stochastic Process. Appl. 125 (2015) 739–758.
[28] J. Shao, F. Xi, Strong ergodicity of the regime-switching diffusion processes, Stochastic Process. Appl. 123 (2013) 3903–3918.
[29] Q. Song, G. Yin, C. Zhu, Optimal switching with constraints and utility maximization of an indivisible market, SIAM J. Control Optim. 50 (2012) 629–651.
[30] X. Tong, A.J. Majda, Moment bounds and geometric ergodicity of diffusions with random switching and unbounded transition rates, Res. Math. Sci. 3 (41) (2016) 33.
[31] A.Y. Veretennikov, Bounds for the mixing rate in the theory of stochastic equations, Theory Probab. Appl. 32 (1988) 273–281.
[32] Max-K. von Renesse, M. Scheutzow, Existence and uniqueness of solutions of stochastic functional differential equations, Random Oper. Stoch. Equ. 18 (2010) 267–284.
[33] F. Xi, C. Zhu, On Feller and strong Feller properties and exponential ergodicity of regime-switching jump diffusion processes with countable regimes, SIAM J. Control Optim. 55 (2017) 1789–1818.
[34] C. Yuan, X. Mao, Asymptotic stability in distribution of stochastic differential equations with Markovian switching, Stochastic Process. Appl. 103 (2003) 277–291.
[35] C. Yuan, X. Mao, Stationary distributions of Euler–Maruyama-type stochastic difference equations with Markovian switching and their convergence, J. Difference Equ. Appl. 11 (2005) 29–48.
[36] C. Yuan, J. Zou, X. Mao, Stability in distribution of stochastic differential delay equations with Markovian switching, Systems Control Lett. 50 (2003) 195–207.
[37] C. Zhu, G. Yin, On competitive Lotka–Volterra model in random environments, J. Math. Anal. Appl. 357 (2009) 154–170.