Ergodicity for neutral type SDEs with infinite length of memory [*]

Jianhai Bao[b),c)], **Feng-Yu Wang**[a),c)], **Chenggui Yuan**[c)]

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

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

[c)]Department of Mathematics, Swansea University, Bay Campus, SA1 8EN, UK

jianhaibao@csu.edu.cn, wangfy@bnu.edu.cn, C.Yuan@swansea.ac.uk

September 10, 2019

Abstract

In this paper, the weak Harris theorem developed in [18] is illustrated by using a straightforward Wasserstein coupling, which implies the exponential ergodicity of the functional solutions to a range of neutral type SDEs with infinite length of memory. A concrete example is presented to illustrate the main result.

AMS Subject Classification: 34K50, 37A30, 60J05
Keywords: ergodicity, neutral type stochastic differential equation, infinite memory, weak Harris' theorem, Wasserstein coupling

1 Introduction

The ergodicity theory is a rich and active area in the study of Markov processes and related topics. Existing results include both qualitative characterizations (for instance, existence and uniqueness of invariant probability measures, strong Feller property, irreducibility) and quantitative estimates (convergence rate of Markov transition semigroups, gradient and heat kernel estimates, etc.). Among many other references, we would like to mention [10, 12, 26, 31] for the study of non-degenerate stochastic differential equations (SDEs) and stochastic partial differential equations (SPDEs), and [13, 15, 16, 17, 25, 30, 33] for degenerate SDEs/SPDEs. In these references, several different probability distances (for

---

example, total variational distance, L^2 distance, and Wasserstein distance) have been adopted to measure the convergence rate of Markov transition semigroups. Efficient tools developed in the literature include functional inequalities (for instance, weak Poincaré, Poincaré, and log-Sobolev inequalities), Lyapunov type criteria, Harris' theorem, and coupling method, etc.

For path-dependent SDEs (i.e., the coefficients depend on the history), which are also called functional SDEs or SDEs with memory, the solutions are no longer Markovian. In this case, one investigates the functional solutions (i.e., the segment process, also called window process, of the solutions), which are Markov processes on the path space determined by the length of memory. However, the above tools mentioned are very hard to apply to such kind of infinite-dimensional Markov processes:

- Due to the lack of characterization on Dirichlet forms, functional inequalities are not yet established;

- The Lyapunov condition on the path space is less explicit since the formulation of infinitesimal generator is not yet available; on account of the same reason, the classical coupling argument via coupling operator is invalid;

- The classical Harris' theorem does not apply since the functional solutions are highly degenerate (infinite-dimensional Markov processes with finite-dimensional noises).

To see that the above mentioned tools do not work for path-dependent SDEs, we consider the following instructive example taken from [8, 18]

$$(1.1) \qquad \mathrm{d}X(t) = b(X(t-1))\mathrm{d}t + \sigma(X(t-1))\mathrm{d}W(t), \quad t \geq 0,$$

with the initial datum $X(\theta) = \xi(\theta), \theta \in [-1, 0]$, where $b : \mathbb{R} \to \mathbb{R}$ is a Lipschitz function, $\sigma : \mathbb{R} \to \mathbb{R}$ is a positive strictly increasing bounded Lipschitz function, $\xi : [-1, 0] \to \mathbb{R}$ is a continuous function, and W is a 1-dimensional Brownian motion. As shown in [27] that (1.1) has a reconstruction property, i.e., given the trajectory $(X(t, \omega))_{t \in [N, N+1]}$ for some $N > 0$, the initial path $(\xi(t))_{t \in [-1, 0]}$ can be reconstructed with probability one. Whence, the functional solution to (1.1) does not admit the strong Feller property and nor the mixing property. Moreover, a "small set" has to be a singleton, and the classical Lyapunov condition (based on infinitesimal generator) does not hold true any more since both the drift and the diffusion are lack of the present information.

In recent years, some new approaches have been developed to investigate the ergodicity and related properties for path-dependent SDEs. When the noise term is path-independent and the drift depends only on a finite segment of path, the ergodicity under the total variational distance was investigated in [8], while gradient estimates and Harnack type inequalities (which in particular imply the strong Feller and irreducibility) have been established in [3, 34, 32], to name a few, by using coupling by change of measures. When the noise part is also path-dependent, but both drift and noise parts depend only on a fixed length of past path, a weak Harris' theorem has been established in [18] to derive the exponential ergodicity under the Wasserstein distance. In case the noise is

2

path-dependent, we would like to emphasize that the ergodicity under the total variational distance and the strong Feller property are not available since the laws of functional solutions with different initial data are mutually singular. The weak Harris' theorem has been applied in, e.g., [7, 11, 23, 29] to establish the ergodicity for highly degenerate stochastic dynamical systems including Markov processes with random switching.

In this paper, we aim to investigate by using a straightforward Wasserstein coupling the exponential ergodicity for a class of neutral type SDEs with infinite length of memory. Such kind of model fits more real world systems whose time evolution depends on the whole history (cf. [24, Chapter 6]). Intuitively, the farer the history, the weaker the influence to the evolution of the system. So, in the following we will take a reference norm on the path space which indicates that the influence of history decays exponentially when the time goes to $-\infty$.

For an integer $d \geq 1$, let $(\mathbb{R}^d, \langle \cdot, \cdot \rangle, |\cdot|)$ be the standard d-dimensional Euclidean space, and $\mathbb{R}^d \otimes \mathbb{R}^d$ the family of all $d \times d$-matrices equipped with the Hilbert-Schmidt norm $\|\cdot\|_{\mathrm{HS}}$. $\mathscr{C} = C((-\infty, 0]; \mathbb{R}^d)$ stands for the space of all continuous maps $f : (-\infty, 0] \to \mathbb{R}^d$. For a map $f(\cdot) : (-\infty, \infty) \to \mathbb{R}^d$, define its segment map $f_\cdot : [0, \infty) \to \mathscr{C}$ by setting

$$f_t(\theta) := f(t + \theta), \quad t \geq 0, \theta \in (-\infty, 0].$$

For a fixed number $r \in (0, \infty)$, let

$$\mathscr{C}_r = \left\{ \phi \in \mathscr{C} : \|\phi\|_r := \sup_{-\infty < \theta \leq 0} (e^{r\theta} |\phi(\theta)|) < \infty \right\}.$$

Then, $(\mathscr{C}_r, \|\cdot\|_r)$ is a Polish space. The norm $\|\cdot\|_r$ fits the intuition of exponential decay with regard to the influence of history; that is, the contribution to the norm from the history at time $\theta < 0$ has a minus exponential discount $e^{r\theta}$. For any $\theta \in (-\infty, 0]$, let $\xi_0(\theta) \equiv \mathbf{0}$, a d-dimensional zero vector.

Consider the following path-dependent SDE of neutral type

$$(1.2) \qquad d\{X(t) - G(X_t)\} = b(X_t)dt + \sigma(X_t)dW(t), \quad t > 0, \quad X_0 = \xi \in \mathscr{C}_r,$$

where $G, b : \mathscr{C}_r \to \mathbb{R}^d$ and $\sigma : \mathscr{C}_r \to \mathbb{R}^d \otimes \mathbb{R}^d$ are measurable with $G(\xi_0) = \mathbf{0}$, $(X_t)_{t \geq 0}$ is the segment process associated with $(X(t))_{t \geq 0}$, $(W(t))_{t \geq 0}$ is the d-dimensional Brownian motion on a complete filtration probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t \geq 0}, \mathbb{P})$. For more motivating examples of (1.2), please refer to [24, p.201-202].

A continuous adapted process $(X(t))_{t \geq 0}$ is called a solution to (1.2) with the initial value X_0, if $\mathbb{P}$-a.s.

$$X(t) = X(0) + G(X_t) - G(X_0) + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dW(s), \quad t \geq 0.$$

We call $(X_t^\xi)_{t \geq 0}$ a functional solution to (1.2) with the initial value $X_0^\xi = \xi \in \mathscr{C}_r$.

Throughout this paper, we assume that

3

(**A0**) b and σ are continuous and bounded on bounded subsets of $\mathscr{C}_r$, and there exists $\alpha \in (0,1)$ such that

$$|G(\xi) - G(\eta)| \le \alpha \|\xi - \eta\|_r, \quad \xi, \eta \in \mathscr{C}_r.$$

(**A1**) There exists a constant $L_0 > 0$ such that

$$\langle \xi(0) - \eta(0) + G(\eta) - G(\xi), b(\xi) - b(\eta) \rangle^+ + \|\sigma(\xi) - \sigma(\eta)\|_{\mathrm{HS}}^2 \le L_0 \|\xi - \eta\|_r^2, \quad \xi, \eta \in \mathscr{C}_r.$$

(**A2**) For each $\xi \in \mathscr{C}_r$, $\sigma(\xi)$ is invertible, and $\sup_{\xi \in \mathscr{C}_r} \{\|\sigma(\xi)\| + \|\sigma(\xi)^{-1}\|\} < \infty$.

(**A3**) There exists a continuous function $V : \mathscr{C}_r \to \mathbb{R}_+$ with $\lim_{\|\xi\|_r \to \infty} V(\xi) = \infty$ such that

$$P_t V(\xi) \le K \mathrm{e}^{-\gamma t} V(\xi) + K$$

holds for some constants $K, \gamma > 0$.

Under (**A0**) and (**A1**), (1.2) has a unique functional solution $(X_t^\xi)_{t \ge 0}$; see [2, 4] for more details. Before we proceed to investigate ergodicity of the functional solution to (1.2), we introduce some additional notation. For the functional solution X_t^ξ, which is a Markov process, we define the Markov semigroup P_t by

$$P_t f(\xi) = \mathbb{E} f(X_t^\xi) = \int_{\mathscr{C}_r} f(\eta) P_t(\xi, \mathrm{d}\eta), \quad t \ge 0, \ f \in \mathscr{B}_b(\mathscr{C}_r), \ \xi \in \mathscr{C}_r,$$

where $P_t(\xi, A) := \mathbb{E} \mathbf{1}_A(X_t^\xi), A \in \mathscr{B}(\mathscr{C}_r)$, which is the transition kernel of X_t^ξ. As explained above, $P_t(\xi, \cdot)$ does not converges in the total variational distance in particular when the noise term is path-dependent. Next, we take the Wasserstein distance induced by the distance

(1.3) $$\rho_r(\xi, \eta) := 1 \wedge \|\xi - \eta\|_r, \quad \xi, \eta \in \mathscr{C}_r.$$

For any $\mu, \nu \in \mathscr{P}(\mathscr{C}_r)$, the collection of all probability measures on $\mathscr{C}_r$, the L^1-Wasserstein distance between μ and ν induced by ρ_r is defined by

(1.4) $$\mathbb{W}_{\rho_r}(\mu, \nu) = \inf_{\pi \in \mathscr{C}(\mu, \nu)} \int_{\mathscr{C}_r \times \mathscr{C}_r} \rho_r(\xi, \eta) \pi(\mathrm{d}\xi, \mathrm{d}\eta),$$

where $\mathscr{C}(\mu, \nu)$ is the set of all couplings of μ and ν; that is, $\pi \in \mathscr{C}(\mu, \nu)$ if and only if π is a probability measure on $\mathscr{C}_r \times \mathscr{C}_r$ such that $\pi(\cdot \times \mathscr{C}_r) = \mu$ and $\pi(\mathscr{C}_r \times \cdot) = \nu$. For the Lyapunov function V in (**A3**), let

$$\rho_{r,V}(\xi, \eta) = \sqrt{\rho_r(\xi, \eta)(1 + V(\xi) + V(\eta))}, \quad \xi, \eta \in \mathscr{C}_r.$$

4

Theorem 1.1. *Assume* (**A0**)-(**A3**). *Then P_t has a unique invariant probability measure π, and there exist constants $c, \lambda > 0$ such that*

$$(1.5) \qquad \mathbb{W}_{\rho_{r,V}}(\mu P_t, \nu P_t) \le c\, \mathrm{e}^{-\lambda t} \mathbb{W}_{\rho_{r,V}}(\mu, \nu), \quad \mu, \nu \in \mathscr{P}(\mathscr{C}_r),\, t \ge 0,$$

where $\mathbb{W}_{\rho_{r,V}}$ is defined as in (1.4) *with ρ_r therein replaced by the pseuso-metric $\rho_{r,V}$. Consequently, there exists a constant $C > 0$ such that*

$$(1.6) \qquad \mathbb{W}_{\rho_{r,V}}(P_t(\xi, \cdot), \pi) \le C\, \mathrm{e}^{-\lambda t} \sqrt{1 + V(\xi)}, \quad t \ge 0.$$

Theorem 1.1 only considers the exponential ergodicity of the solution to (1.2) in the Wasserstein distance. However, the method used in its proof also works in the study of the ergodicity with slower (i.e., subexponential, polynomial) convergence rates by replacing (**A3**) with

$$P_t V(\xi) \le V(\xi) - \int_0^t P_s(\varphi \circ V)(\xi)\mathrm{d}s + Kt, \quad \xi \in \mathscr{C}, \; t \ge 0,$$

where function $V : \mathscr{C} \to \mathbb{R}$ is measurable with $\lim_{\|\xi\|_r \to \infty} V(\xi) = \infty$, $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$ is concave with $\varphi(0) = 0$ and $\varphi(s) \uparrow \infty$ as $s \uparrow \infty$, and $K \ge 0$ is a constant. See, for instance, [8, 21] for the study of path-dependent SDEs with finite length of memory. We also remark that, in [14], sufficient conditions are provided for subgeometric rates of convergence in Wasserstein distance concerning general state-space Markov chains which need not to be irreducible.

The proof of Theorem 1.1 is based on the weak Harris' theorem developed [18] (see Lemma 2.1 below for more details) and is completed by constructing a straightforward Wasserstein coupling. It is worth pointing out that straightforward Wasserstein coupling, also called generalized coupling, has been applied to study the exponential ergodicity in Wasserstein type metrics for path-dependent SDEs, SPDEs, and some other infinite-dimensional processes in the Wasserstein metric, see for instance [9, 22] and references within.

To meet the conditions of Lemma 2.1 below, one has to overcome the difficulties caused by the infinite length of memory. Unlike conditions (**A0**) and (**A1**), which are explicitly imposed on the coefficients, the Lyapunov condition (**A3**) is set by means of the semigroup P_t which is less explicit. In many cases, one may verify (**A3**) by using the Lyapunov condition

$$(1.7) \qquad \mathscr{L}V(\xi) \le -\lambda V(\xi) + c, \quad \xi \in \mathscr{C}_r$$

for some constants $c, \lambda > 0$, where $\mathscr{L}$ is the extended generator corresponding to the semigroup $(P_t)_{t \ge 0}$. However, as already explained before, $\mathscr{L}$ is not yet available for the present model. In this spirit, we present below explicit conditions for (**A3**).

Proposition 1.2. *Let $\mu_0 \in \mathscr{P}((-\infty, 0])$ such that*

$$(1.8) \qquad \delta_r(\mu_0) := \int_{-\infty}^0 \mathrm{e}^{-2r\theta} \mu_0(\mathrm{d}\theta) < \infty,$$

5

and set

$$(1.9) \qquad \beta := \left(1 + \sqrt{\alpha_1 + \alpha_2 \delta_r(\mu_0)}\right)^2.$$

Then (**A3**) *holds for* $V(\xi) := \|\xi\|_r^2$ *provided that the following two conditions hold:*

(i) *For any* $\xi \in \mathscr{C}_r$, *there exist constants* $\alpha_1, \alpha_2 > 0$ *with* $\alpha_1 + \alpha_2 \delta_r(\mu_0) < 1$ *such that*

$$(1.10) \qquad |G(\xi)|^2 \leq \alpha_1 |\xi(0)|^2 + \alpha_2 \int_{-\infty}^0 |\xi(\theta)|^2 \mu_0(\mathrm{d}\theta).$$

(ii) *There exist constants* $c_0, \lambda_1, \lambda_2 > 0$ *with* $\gamma := \lambda_1 - 2r\beta - \lambda_2 \delta_r(\mu_0) > 0$ *such that*

$$(1.11) \quad 2\langle \xi(0) - G(\xi), b(\xi) \rangle + \|\sigma(\xi)\|_{\mathrm{HS}}^2 \leq c_0 - \lambda_1 |\xi(0)|^2 + \lambda_2 \int_{-\infty}^0 |\xi(\theta)|^2 \mu_0(\mathrm{d}\theta),$$

$$(1.12) \qquad \|\sigma(\xi)\|_{\mathrm{HS}}^2 \leq c_0 \left(1 + |\xi(0)|^2 + \int_{-\infty}^0 |\xi(\theta)|^2 \mu_0(\mathrm{d}\theta)\right).$$

The assumption on the Lyapunov function (i.e., $\lim_{\|\xi\|_r \to \infty} V(\xi) = \infty$) seems to be a little bit restrictive. Whereas, provided that (i) the diffusion coefficient is uniformly bounded, and the drift coefficient can be decomposed into two terms

$$b(\xi) = b_1(\xi) + b_2(\xi(0)), \quad \xi \in \mathscr{C}_r,$$

where $b_1 : \mathscr{C} \to \mathbb{R}^d$ is uniformly bounded; (ii) the drift b satisfies the generalized Veretennikov-Khasminskii condition, i.e., there exist constants $\kappa, M > 0$ such that

$$(1.13) \qquad \langle \xi(0) - G(\xi), b(\xi) \rangle \leq -\kappa |\xi(0)|, \quad \xi \in \mathscr{C}_r, \quad |\xi(0)| \geq M,$$

the Lyapunov function can be designed via the present state but not the whole path. See e.g. [7, Theorem 3.2 (ii) & Theorem 3.3 (i)] for further details. For further examples which satisfy the generalized Veretennikov-Khasminskii condition, please refer to [8, Section 3] for more details. If the drift term is pure delay, the condition (1.11) no loner holds true. In this setup, we can replace (1.11) by the generalized Veretennikov-Khasminskii condition (1.13) concerning the drift term b. Moreover, for the following semi-linear SDE of neutral type

$$\mathrm{d}\left(X(t) - \int_{-\tau}^0 X(t+\theta)\rho(\mathrm{d}\theta)\right) = \left(\int_{-\tau}^0 X(t+\theta)\mu(\mathrm{d}\theta)\right)\mathrm{d}t + \sigma(X_t)\mathrm{d}W(t),$$

where ρ, μ are signed measure on $[-\tau, 0]$, and σ is bounded Lipschitz and non-degenerate, we can examine that (**A3**) still is satisfied by taking $V(\xi) = |\xi(0)|^2$ and employing a variation-of-constants formula although (1.11) is invalid; see [5, Section 4] for more details.

To conclude this section, we present below a concrete example to illustrate Theorem 1.1.

Example 1.3. *Let* $\mu_0(\mathrm{d}\theta) = \frac{1}{r_0} \mathrm{e}^{r_0\theta} \mathrm{d}\theta \in \mathscr{P}((-\infty, 0])$ *for some* $r_0 > 2r$ *and let*

$$G(\xi) = \gamma_1 \int_{-\infty}^{0} \xi(\theta)\mu_0(\mathrm{d}\theta), \quad \sigma(\xi) = 1 + \gamma_2 \int_{-\infty}^{0} (1 \wedge |\xi(\theta)|)\mu_0(\mathrm{d}\theta),$$

$$b(\xi) = -\gamma_3 \xi(0) - \gamma_4 \left(\xi(0) - \gamma_1 \int_{-\infty}^{0} \xi(\theta)\mu_0(\mathrm{d}\theta)\right)^{\frac{1}{3}} + \gamma_5 \int_{-\infty}^{0} \xi(\theta)\mu_0(\mathrm{d}\theta)$$

for some constants $\gamma_i > 0, i = 1, \cdots, 5$. *If*
(1.14)

$$\gamma_1^2 < r_0(r_0 - 2r) \quad \text{and} \quad 2\gamma_3 > 2r\left(1 + \frac{\gamma_1}{\sqrt{r_0(r_0 - 2r)}}\right)^2 + \frac{\gamma_2^2}{r_0(r_0 - 2r)} + \frac{2(\gamma_5 + \gamma_1\gamma_3)}{\sqrt{r_0(r_0 - 2r)}},$$

assertions in Theorem 1.1 hold.

Remark 1.1. Path-dependent SDEs of neutral type have been utilized to model some evolution phenomena arising in e.g. physics, biology and engineering; see, for instance, [1, 19, 20]. Here, whenever μ_0 is a signed measure on the finite time interval $[-\tau, 0]$ and $\gamma_4 = 0$, this example has been investigated in [1]. Moreover, the example can demonstrate Proposition 1.2 in the case $\gamma_4 = 0$ (the natural case). Whereas, we herein allow $\gamma_4 \neq 0$ just to emphasize that the drift term can be singular.

The remainder of this paper is organized as follows. In Section 2, Theorem 1.1 is proved by using weak Harris' theorem and Wasserstein coupling. Section 4 is devoted to the proofs on Proposition 1.2 and Example 1.3.

2 Proof of Theorem 1.1

Now, we shall start to complete the proof of Theorem 1.1 by the aid of weak Harris' theorem introduced in [18]. For readers' convenience, we state it below in details. To begin, we recall some notions.

Definition 2.1. Let $\mathbb{X}$ be a Polish space, and $(P_t)_{t\geq 0}$ a Markov semigroup with transition kernel $P_t(\xi, \cdot)$ on $\mathbb{X}$.

(1) A continuous function $V : \mathbb{X} \to \mathbb{R}_+$ is called a Lyapunov function for $(P_t)_{t\geq 0}$, if there exist constants $\gamma, K > 0$ such that

$$(2.1) \qquad P_t V(\xi) := \int_{\mathbb{X}} V(\eta) P_t(\xi, \mathrm{d}\eta) \leq K \mathrm{e}^{-\gamma t} V(\xi) + K, \qquad \xi \in \mathbb{X}, \quad t \geq 0.$$

(2) A function $\rho : \mathbb{X} \times \mathbb{X} \to [0, 1]$ is said to be distance-like if it is symmetric, lower semi-continuous, and $\rho(\xi, \eta) = 0$ if and only if $\xi = \eta$.

(3) A set $A \subset \mathbb{X}$ is said to be ρ-small for P_t, if there exists $\varepsilon \in (0, 1)$ such that

$$\mathbb{W}_\rho(P_t(\xi, \cdot), P_t(\eta, \cdot)) \leq 1 - \varepsilon, \quad \xi, \eta \in A,$$

where $\mathbb{W}_\rho$ is defined as in (1.4) for $(\mathbb{X}, \rho)$ replacing $(\mathscr{C}_r, \rho_r)$.

(4) ρ is said to be contractive for P_t, if there exists $\varepsilon \in (0,1)$ such that

$$\mathbb{W}_\rho(P_t(\xi, \cdot), P_t(\eta, \cdot)) \le \varepsilon \rho(\xi, \eta), \quad \xi, \eta \in \mathbb{X} \text{ with } \rho(\xi, \eta) < 1.$$

The following result is due to [18, Theorem 4.8].

Lemma 2.1. *Let ρ be a distance-like function on $\mathbb{X} \times \mathbb{X}$, and V a Lyapunov function such that (2.1) holds for some constants $\gamma, K > 0$. If there exists a constant $t^* > 0$ such that $\{V \le 4K\}$ is ρ-small and ρ is contractive for P_{t^*}, then there exists a constant $t > 0$ such that*

$$\mathbb{W}_{\rho_V}(\mu P_t, \nu P_t) \le \frac{1}{2} \mathbb{W}_{\rho_V}(\mu, \nu), \quad \forall \mu, \nu \in \mathscr{P}(\mathbb{X}),$$

where $\rho_V(\xi, \eta) := \sqrt{\rho(\xi, \eta)(1 + V(\xi) + V(\eta))}, \xi, \eta \in \mathbb{X}$.

To apply this result to the present model, for any $\delta > 0$ and $R > 0$, let

$$\rho_{r,\delta} = 1 \wedge (\delta^{-1}\rho_r), \quad B_R = \{\xi \in \mathscr{C}_r : \|\xi\|_r \le R\},$$

(2.2) $$t_{R,\delta} = 1 + \frac{1}{2r} \log \left(\frac{3}{2\delta^2} \left(\frac{2e^{2r}}{(1-\alpha)^2} \left(2R + \frac{\delta}{3} \right)^2 + \frac{R^2}{1-\alpha} \right) \right),$$

where ρ_r was given in (1.3). Obviously, the metric $\rho_{r,\delta}$ is equivalent to ρ_r. To check the conditions in Lemma 2.1 for the present setup, we need to prepare the following four lemmas concerned, respectively, with the (local) irreducibility, the continuity with respect to the initial variable, $\rho_{r,\delta}$-small property, and $\rho_{r,\delta}$-contractive property for the Markov transition kernel.

Lemma 2.2. *Under the conditions of Theorem 1.1, for any $R, \delta > 0$,*

(2.3) $$\inf_{\xi \in B_R} \mathbb{P}(X_t^\xi \in B_\delta) > 0, \quad t \ge t_{R,\delta}.$$

Proof. The idea of proof is essentially borrowed from that of [28, Lemma 2.2], where the crucial point is to apply a standard result (e.g., [6, Lemma I.8.3]) that a uniform elliptic diffusion process is irreducible. So, below we will compare the radial process $|X^\xi|(s)$ with an elliptic diffusion process. For any $\xi \in B_R$ and $\delta > 0$, let $h \in C_b^\infty(\mathbb{R}_+; \mathbb{R}^d)$ such that

(2.4) $h(0) = \xi(0) - G(\xi) - \dfrac{\delta(1-\alpha)}{3}(1, 0, \cdots, 0), \quad |h| \le |h(0)|, \quad \text{and } h(s) = \mathbf{0} \text{ for } s \ge 1,$

where $\alpha \in (0,1)$ was introduced in (**A0**). Let

(2.5) $$\Lambda^{X^\xi}(t) = X^\xi(t) - G(X_t^\xi), \quad \Lambda^{X^\xi, X^\eta}(t) = \Lambda^{X^\xi}(t) - \Lambda^{X^\eta}(t), \quad t \ge 0.$$

By taking (**A0**) into consideration, we have

(2.6) $(1-\alpha)^2 \sup_{0 \le u \le s} (e^{2ru}|X^\xi(u)|^2) \le (1-\alpha)\|\xi\|_r^2 + \sup_{0 \le u \le s} (e^{2ru}|\Lambda^{X^\xi}(u)|^2), \quad s \ge 0.$

8

Consider the following radial process

(2.7)
$$D(s) := |\Lambda^{X^\xi}(s) - h(s)|^2 - \frac{\delta^2(1-\alpha)^2}{9}, \quad s \geq 0.$$

By Itô's formula, it follows that

(2.8)
$$\begin{aligned}
\mathrm{d}(\mathrm{e}^{2rs}D(s)) &= 2\,r\mathrm{e}^{2rs}D(s)\mathrm{d}s + \mathrm{e}^{2rs}\mathrm{d}D(s) \\
&= \mathrm{e}^{2rs}\big\{2\,rD(s) + 2\langle\Lambda^{X^\xi}(s) - h(s), b(X_s^\xi) - h'(s)\rangle + \|\sigma(X_s^\xi)\|_{\mathrm{HS}}^2\big\}\mathrm{d}s \\
&\quad + 2\,\mathrm{e}^{2rs}\langle\Lambda^{X^\xi}(s) - h(s), \sigma(X_s^\xi)\mathrm{d}W(s)\rangle, \quad s \geq 0.
\end{aligned}$$

Define the stopping time

(2.9)
$$\tau = \inf\left\{s \geq 0 : \mathrm{e}^{2rs}|D(s)| \geq \frac{\delta^2(1-\alpha)^2}{18}\right\}.$$

Since $D(0) = 0$ and $D(s)$ is continuous with respect to s, we have $\mathbb{P}(\tau > 0) = 1$. In terms of (2.7) and (2.9), we therefore have

$$\frac{\delta^2(1-\alpha)^2}{18} \geq \mathrm{e}^{2rs}|D(s)| \geq |D(s)| \geq \frac{\delta^2(1-\alpha)^2}{9} - |\Lambda^{X^\xi}(s) - h(s)|^2, \quad s \in [0, \tau].$$

As a consequence, we arrive at

(2.10)
$$|\Lambda^{X^\xi}(s) - h(s)|^2 \geq \frac{\delta^2(1-\alpha)^2}{18}, \quad s \in [0, \tau].$$

Combining (2.10) with (**A2**), we obtain from (2.8) that

(2.11)
$$\frac{\mathrm{d}}{\mathrm{d}s}\langle\mathrm{e}^{2rs}D(s)\rangle = 4\,\mathrm{e}^{4rs}|\sigma^*(X_s^\xi)(\Lambda^{X^\xi}(s) - h(s))|^2 \in [c_1, c_2], \quad s \in [0, \tau \wedge t]$$

for some constants $c_2 > c_1 > 0$. Herein, $\langle\cdot\rangle$ means the quadratic variation of a continuous semi-martingale and $t \geq t_{R,\delta}$, which is to be fixed in what follows. Next, we are going to claim that (2.11) implies that

(2.12)
$$\mathbb{P}\left(\sup_{0 \leq u \leq s}(\mathrm{e}^{2ru}|D(u)|) < \frac{\delta^2(1-\alpha)^2}{18}\right) > 0, \quad s \geq 0.$$

To achieve (2.12), we extend $(\mathrm{e}^{2rs}D(s))_{s\in[0,\tau]}$ into $(\mathrm{e}^{2rs}D(s))_{s\geq 0}$ in the following manner

(2.13)
$$Y(s) := \mathrm{e}^{2r(s\wedge\tau)}D(s \wedge \tau) + \mathbf{1}_{\{s>\tau\}}(W^1(s) - W^1(\tau)), \quad s \geq 0,$$

where $(W^1(s))_{s\geq 0}$ stands for the first component of $(W(s))_{s\geq 0}$. Consequently, (2.11) gives that

$$\frac{\mathrm{d}}{\mathrm{d}s}\langle Y(s)\rangle \in [c_1 \wedge 1, c_2 \vee 1].$$

9

By applying [6, Lemma I.8.3] and using $Y(0) = \mathbf{0}$, this yields

(2.14)
$$\mathbb{P}\Big(\sup_{0 \leq u \leq s} |Y(u)| < c \Big) > 0, \quad s, c > 0.$$

Combining this with (2.9) and (2.13), we obtain

$$\mathbb{P}\Big(\sup_{0 \leq u \leq s} (\mathrm{e}^{2ru}|D(u)|) < \frac{\delta^2(1-\alpha)^2}{18} \Big) = \mathbb{P}\Big(\sup_{0 \leq u \leq s \wedge \tau} (\mathrm{e}^{2ru}|D(u)|) < \frac{\delta^2(1-\alpha)^2}{18}, s < \tau \Big)$$

$$= \mathbb{P}\Big(\sup_{0 \leq u \leq s} |Y(u)| < \frac{\delta^2(1-\alpha)^2}{18} \Big) > 0.$$

So, (2.12) holds true.

By using the fundamental inequality: $\frac{1}{2}|u|^2 - |v|^2 \leq |u-v|^2, u, v \in \mathbb{R}^d$, and recalling that $h(s) = \mathbf{0}$ for $s \geq 1$, we deduce from (2.6) and $\xi \in B_R$ that

$$\mathbb{P}\Big(\sup_{0 \leq s \leq t} (\mathrm{e}^{2rs}|D(s)|) \leq \frac{\delta^2(1-\alpha)^2}{18} \Big)$$

$$\leq \mathbb{P}\Big(\sup_{0 \leq s \leq t} (\mathrm{e}^{2rs}|\Lambda^{X^\xi}(s) - h(s)|^2) \leq \frac{\delta^2(1-\alpha)^2}{6} \mathrm{e}^{2rt} \Big)$$

(2.15)
$$\leq \mathbb{P}\Big(\sup_{0 \leq s \leq t} \Big(\frac{1}{2} \mathrm{e}^{2rs}|\Lambda^{X^\xi}(s)|^2 - \mathrm{e}^{2rs}|h(s)|^2 \Big) \leq \frac{\delta^2(1-\alpha)^2}{6} \mathrm{e}^{2rt} \Big)$$

$$\leq \mathbb{P}\Big(\sup_{0 \leq s \leq t} (\mathrm{e}^{2rs}|\Lambda^{X^\xi}(s)|^2) \leq 2 \sup_{0 \leq s \leq 1} (\mathrm{e}^{2rs}|h(s)|^2) + \frac{\delta^2(1-\alpha)^2}{3} \mathrm{e}^{2rt} \Big)$$

$$\leq \mathbb{P}\Big(\sup_{0 \leq s \leq t} (\mathrm{e}^{2rs}|X^\xi(s)|^2) \leq \frac{1}{1-\alpha} R^2 + \frac{2}{(1-\alpha)^2} \sup_{0 \leq s \leq 1} (\mathrm{e}^{2rs}|h(s)|^2) + \frac{\delta^2}{3} \mathrm{e}^{2rt} \Big).$$

On the other hand, we observe that

$$\mathbb{P}(X_t^\xi \in B_\delta) = \mathbb{P}\Big(\mathrm{e}^{-2rt} \sup_{-\infty < s \leq t} (\mathrm{e}^{2rs}|X^\xi(s)|^2) \leq \delta^2 \Big)$$

$$= \mathbb{P}\Big(\|\xi\|_r^2 \vee \sup_{0 \leq s \leq t} (\mathrm{e}^{2rs}|X^\xi(s)|^2) \leq \mathrm{e}^{2rt}\delta^2 \Big)$$

$$\geq \mathbb{P}\Big(\sup_{0 \leq s \leq t} (\mathrm{e}^{2rs}|X^\xi(s)|^2) \leq \mathrm{e}^{2rt}\delta^2 \Big).$$

Combining this with (2.12) and (2.15), it follows that we need to show

(2.16)
$$\frac{1}{1-\alpha} R^2 + \frac{2}{(1-\alpha)^2} \sup_{0 \leq s \leq 1} (\mathrm{e}^{2rs}|h(s)|^2) + \frac{\delta^2}{3} \mathrm{e}^{2rt} \leq \mathrm{e}^{2rt}\delta^2, \quad t \geq t_{R,\delta}.$$

By (2.4), (**A0**), $G(\xi_0) = \mathbf{0}$ and $\xi \in B_R$, we infer that

$$|h|^2 \leq |h(0)|^2 \leq \Big(2R + \frac{\delta}{3} \Big)^2,$$

10

which incurs

$$\sup_{0 \leq s \leq 1} (e^{2rs}|h(s)|^2) \leq e^{2r}\Big(2R + \frac{\delta}{3}\Big)^2.$$

Then (2.16) holds definitely provided that

$$\frac{2\delta^2}{3}e^{2rt} \geq \frac{2e^{2r}}{(1-\alpha)^2}\Big(2R + \frac{\delta}{3}\Big)^2 + \frac{R^2}{1-\alpha},$$

which indeed is true for $t \geq t_{R,\delta}$. $\qquad\square$

Lemma 2.3. *Under conditions of Theorem 1.1, there exists a constant $K > 0$ such that*

$$(2.17) \qquad \mathbb{E}\|X_t^\xi - X_t^\eta\|_r^2 \leq Ke^{Kt}\,\|\xi - \eta\|_r^2, \quad t \geq 0, \ \xi, \eta \in \mathscr{C}_r.$$

Proof. The proof of Lemma 2.3 is more or less standard (cf. [2, 4]). Whereas we herein provide an outline of the argument just to make the content self-contained and emphasize some corresponding differences due to the infinite memory and the neutral term. Let $\Lambda^{\xi,\eta}(t) = \Lambda^{X^\xi, X^\eta}(t)$ for simplicity. Using (**A0**) we have

$$(2.18) \qquad e^{2rt}\|X_t^\xi - X_t^\eta\|_r^2 \leq \frac{1}{1-\alpha}\|\xi - \eta\|_r^2 + \frac{1}{(1-\alpha)^2}\sup_{0 \leq s \leq t}(e^{2rs}|\Lambda^{\xi,\eta}(s)|^2).$$

Thus, to obtain the desired assertion (2.17), it is sufficient to show that

$$(2.19) \qquad \Gamma(t) := \mathbb{E}\Big(\sup_{0 \leq s \leq t}(e^{2rs}|\Lambda^{\xi,\eta}(s)|^2)\Big) \leq Je^{Jt}\|\xi - \eta\|_r^2, \quad t \geq 0, \ \xi, \eta \in \mathscr{C}_r$$

for some constant $J > 0$. Applying Itô's formula and using (**A1**) and (**A2**), we obtain

$$e^{2rt}|\Lambda^{\xi,\eta}(t)|^2 \leq 4\|\xi - \eta\|_r^2 + \int_0^t e^{2rs}\{2r|\Lambda^{\xi,\eta}(s)|^2 + L_0\|X_s^\xi - X_s^\eta\|_r^2\}ds$$

$$+ 2\int_0^t e^{2rs}\langle\Lambda^{\xi,\eta}(s), (\sigma(X_s^\xi) - \sigma(X_s^\eta))dW(s)\rangle.$$

Combining this with (2.18), (**A2**) and BDG's inequality, we find out constants $c_1, c_2, c_3 > 0$ such that

$$\Gamma(t) \leq 4\|\xi - \eta\|_r^2 + c_1\int_0^t \{\|\xi - \eta\|_r^2 + \Gamma(s)\}ds$$

$$+ \mathbb{E}\Big(\sup_{0 \leq s \leq t}(e^{2rs}|\Lambda^{\xi,\eta}(s)|^2)\int_0^t e^{2rs}\|\sigma(X_s^\xi) - \sigma(X_s^\eta)\|_{HS}^2 ds\Big)^{1/2}$$

$$\leq 4\|\xi - \eta\|_r^2 + c_1\int_0^t \{\|\xi - \eta\|_r^2 + \Gamma(s)\}ds + \frac{1}{2}\Gamma(t) + c_2\int_0^t \mathbb{E}(e^{2rs}\|X_s^\xi - X_s^\eta\|_r^2)ds$$

$$\leq \frac{1}{2}\Gamma(t) + c_3(1+t)\|\xi - \eta\|_r^2 + c_3\int_0^t \Gamma(s)ds.$$

11

Consequently,

$$\Gamma(t) \le 2c_3(1+t)\|\xi - \eta\|_r^2 + 2c_3 \int_0^t \Gamma(s)\mathrm{d}s.$$

By Gronwall's inequality, we obtain

$$\Gamma(t) \le 2c_3(1+t)\mathrm{e}^{2c_3 t}\|\xi - \eta\|_r^2.$$

Therefore, (2.19) follows from (2.18) immediately. $\qquad\square$

Lemma 2.4. *Under conditions of Theorem* 1.1, *for any* $R, \delta > 0$,

$$(2.20) \qquad \mathbb{W}_{\rho_{r,\delta}}(P_t(\xi, \cdot), P_t(\eta, \cdot)) \le 1 - \frac{\alpha_t^2}{2} < 1, \quad t \ge t_{R,\delta/4}, \ \xi, \eta \in B_R,$$

holds for $t_{R,\delta/4}$ *in* (2.2) *and* $\alpha_t := \inf_{\xi \in B_R} \mathbb{P}(X_t^\xi \in B_{\delta/4})$.

Proof. For any $\xi, \eta \in B_R$, let $(X_t^\xi)_{t \ge 0}$ be the functional solution to (1.2) with the initial value $X_0^\xi = \xi \in \mathscr{C}_r$, and $(\widetilde{X}_t^\eta)_{t \ge 0}$ the functional solution to (1.2) with the initial datum $\widetilde{X}_0^\eta = \eta$ but for an independent Brownian motion $(\widetilde{W}(t))_{t \ge 0}$ replacing $(W(t))_{t \ge 0}$. We call $(X_t^\xi, \widetilde{X}_t^\eta)$ an independent coupling of the functional solutions to (1.2). In view of the independence of $(X_t^\xi)_{t \ge 0}$ and $(\widetilde{X}_t^\eta)_{t \ge 0}$, we deduce that

$$
\begin{aligned}
\mathbb{W}_{\rho_{r,\delta}}(P_t(\xi, \cdot), P_t(\eta, \cdot)) &\le \mathbb{E}(1 \wedge (\delta^{-1}\|X_t^\xi - X_t^\eta\|_r)) \\
&\le \frac{1}{2}\mathbb{P}(X_t^\xi \in B_{\delta/4}, X_t^\eta \in B_{\delta/4}) + \mathbb{P}(\{X_t^\xi \notin B_{\delta/4}\} \cup \{X_t^\eta \notin B_{\delta/4}\}) \\
&= 1 - \frac{1}{2}\mathbb{P}(X_t^\xi \in B_{\delta/4})\mathbb{P}(X_t^\eta \in B_{\delta/4}) \\
&\le 1 - \frac{\alpha_t^2}{2}.
\end{aligned}
$$

Hence, (2.20) holds true due to Lemma 2.2. $\qquad\square$

Lemma 2.5. *Under the conditions of Theorem* 1.1, *for any* $\beta \in (0,1)$ *there exist constants* $\delta_\beta, t_\beta > 0$ *such that*

$$(2.21) \qquad \mathbb{W}_{\rho_{r,\delta}}(P_t(\xi, \cdot), P_t(\eta, \cdot)) \le \beta \, \rho_{r,\delta}(\xi, \eta), \qquad t \ge t_\beta, \ \delta \in (0, \delta_\beta]$$

for any $\xi, \eta \in \mathscr{C}_r$ *with* $\rho_{r,\delta}(\xi, \eta) < 1$.

Proof. Our proof is based on the Girsanov transform and Warsserstein coupling, which is more straightforward than the "binding construction" argument adopted in [18, p.254-257]. For $\xi, \eta \in \mathscr{C}_r$, let $(X_s^\xi)_{s \ge 0}$ be the functional solution to (1.2), and let $(Y^\eta(s))_{s \ge 0}$ solve the following SDE

$$(2.22) \quad \mathrm{d}\{Y^\eta(s) - G(Y_s^\eta)\} = \{b(Y_s^\eta) + \lambda\Lambda^{\xi,\eta}(s)\}\mathrm{d}s + \sigma(Y_s^\eta)\mathrm{d}W(s), \qquad s \ge 0, \qquad Y_0^\eta = \eta,$$

12

where $\lambda > 0$ is a constant, $\Lambda^{\xi,\eta}(s) := \Lambda^{X^\xi, X^\eta}(s)$ is defined in (2.5). For $\lambda > 0$ sufficiently large, the additional drift $\lambda \Lambda^{\xi,\eta}(s)$ strongly pushes Y_s^η moving toward to X_s^ξ whenever $s \uparrow \infty$. Indeed, when $\lambda > 0$ is sufficiently large, for any $r_0 \in (0, r)$, there exists a constant $c > 0$ such that

$$(2.23) \qquad \mathbb{E}\|X_s^\xi - Y_s^\eta\|_r^2 \leq c\, \mathrm{e}^{-r_0 s}\|\xi - \eta\|_r^2, \quad s \geq 0,\ \xi, \eta \in \mathscr{C}_r,$$

and, for any stopping time τ,

$$(2.24) \qquad \mathbb{E}\|X_{s\wedge\tau}^\xi - Y_{s\wedge\tau}^\eta\|_r^2 \leq c\,\|\xi - \eta\|_r^2, \quad s \geq 0,\ \xi, \eta \in \mathscr{C}_r$$

as shown in the proof of [4, (3.11)],

To compare Y_s^η with X_s^η via the Girsanov theorem, let $h(s) = \lambda\,\sigma^{-1}(Y_s^\eta)\Lambda^{\xi,\eta}(s)$ and set

$$R_t := \exp\left(- \int_0^t \langle h(s), \mathrm{d}W(s)\rangle - \frac{1}{2}\int_0^t |h(s)|^2 \mathrm{d}s \right).$$

Generally, R_t may not be a well defined probability density, so we shall restrict it by the following stopping time

$$\tau_\varepsilon = \inf\left\{ s \geq 0 : \int_0^s |h(s)|^2 \mathrm{d}s \geq \varepsilon^{-1}\|\xi - \eta\|_r^2 \right\}$$

for some constant $\varepsilon \in (0, 1)$ sufficiently small to be determined later. By Girsanov theorem, $\mathrm{d}\mathbb{Q}_\varepsilon := R_{t\wedge\tau_\varepsilon}\mathrm{d}\mathbb{P}$ is a probability measure on $(\Omega, \mathscr{F})$ under which

$$\widetilde{W}(s) := W(s) + \int_0^{s\wedge\tau_\varepsilon} h(u)\mathrm{d}u, \quad s \geq 0$$

is a d-dimensional Brownian motion. Let $\widetilde{Y}^\eta(s)$ solve the SDE

$$\mathrm{d}\{\widetilde{Y}^\eta(s) - G(\widetilde{Y}_s^\eta)\} = \{b(\widetilde{Y}_s^\eta) + \mathbf{1}_{\{\tau_\varepsilon \geq s\}}\lambda \widetilde{\Lambda}^{\xi,\eta}(s)\}\mathrm{d}s + \sigma(\widetilde{Y}_s^\eta)\mathrm{d}\widetilde{W}(s), \quad s \geq 0, \quad \widetilde{Y}_0^\eta = \eta,$$

where $\widetilde{\Lambda}^{\xi,\eta}(s) := \Lambda^{X^\xi, \widetilde{Y}^\eta}(s)$. By the weak uniqueness of solutions to (2.22) up to time $t\wedge\tau_\varepsilon$, we have

$$(2.25) \qquad \mathbb{P}(X_t^\eta \in \cdot) = \mathbb{Q}_\varepsilon(\widetilde{Y}_t^\eta \in \cdot), \quad \widetilde{Y}_{t\wedge\tau_\varepsilon}^\eta = Y_{t\wedge\tau_\varepsilon}^\eta, \quad t \geq 0.$$

To estimate $\mathbb{W}_{\rho_{r,\delta}}(P_t(\xi, \cdot), P_t(\eta, \cdot))$, we take the following Wasserstein coupling of $\mathbb{P}$ and $\mathbb{Q}_\varepsilon$:

$$(2.26) \qquad \begin{aligned} \Pi(\mathrm{d}\omega, \mathrm{d}\widetilde{\omega}) =\ &(1 \wedge R_{t\wedge\tau_\varepsilon})(\omega)\mathbb{P}(\mathrm{d}\omega)\delta_\omega(\mathrm{d}\widetilde{\omega}) \\ &+ \frac{(1 - R_{t\wedge\tau_\varepsilon})^+(\omega)(R_{t\wedge\tau_\varepsilon} - 1)^+(\widetilde{\omega})}{\mathbb{E}[(1 - R_{t\wedge\tau_\varepsilon})^+]}\mathbb{P}(\mathrm{d}\omega)\mathbb{P}(\mathrm{d}\widetilde{\omega}), \end{aligned}$$

where δ_ω is the Dirac measure at point ω, and the last term vanishes if $\mathbb{E}[(1 - R_{t \wedge \tau_\varepsilon})^+] = 0$ which is only possible when $\xi = \eta$. Combining this coupling with (2.25), and noting that $\rho_{r,\delta} \leq 1$, we obtain that

(2.27)
$$
\begin{aligned}
&\mathbb{W}_{\rho_{r,\delta}}(P_t(\xi, \cdot), P_t(\eta, \cdot)) \\
&\leq \int_{\Omega \times \Omega} \rho_{r,\delta}(X_t^\xi(\omega), \widetilde{Y}_t^\eta(\widetilde{\omega})) \Pi(\mathrm{d}\omega, \mathrm{d}\widetilde{\omega}) \\
&\leq \mathbb{E}\big[\rho_{r,\delta}(X_t^\xi, \widetilde{Y}_t^\eta)(1 \wedge R_{t \wedge \tau_\varepsilon})\big] + \mathbb{E}\big[(R_{t \wedge \tau_\varepsilon} - 1)^+\big] \\
&\leq \mathbb{E}\big[\mathbf{1}_{\{t \leq \tau_\varepsilon\}} \rho_{r,\delta}(X_t^\xi, Y_t^\eta)\big] + \mathbb{E}\big[\mathbf{1}_{\{t > \tau_\varepsilon\}} \rho_{r,\delta}(X_t^\xi, \widetilde{Y}_t^\eta)\big] + \mathbb{E}\big[(R_{t \wedge \tau_\varepsilon} - 1)^+\big] \\
&=: I_1(t) + I_2(t) + I_3(t).
\end{aligned}
$$

Next we are going to estimate three terms above, one-by-one.

Firstly, by (2.23), there exists a constant $c > 0$ such that

(2.28)
$$
I_1(t) \leq c\, \mathrm{e}^{-r_0 t/2} \delta^{-1} \|\xi - \eta\|_r = c\, \mathrm{e}^{-r_0 t/2} \rho_{r,\delta}(\xi, \eta)
$$

for arbitrary $\xi, \eta \in \mathscr{C}_r$ with $\rho_{r,\delta}(\xi, \eta) < 1$. Next, by Hölder's inequality, the strong Markov property, (2.17), (2.23), (2.24), $\widetilde{Y}_{t \wedge \tau_\varepsilon}^\eta = Y_{t \wedge \tau_\varepsilon}^\eta$ due to (2.25), and noting that the SDE for $\widetilde{Y}_s^\eta$ coincides with (1.2) when $s \geq \tau_\varepsilon$, we obtain that

(2.29)
$$
\begin{aligned}
I_2(t) &\leq \delta^{-1} \mathbb{E}\big[\|X_t^\xi - \widetilde{Y}_t^\eta\|_r \mathbf{1}_{\{\tau_\varepsilon < t\}}\big] \\
&= \delta^{-1} \mathbb{E}\Big[\mathbf{1}_{\{\tau_\varepsilon < t\}} \big\{\mathbb{E}\|X_{t-\tau_\varepsilon}^{\xi'} - \widetilde{Y}_{t-\tau_\varepsilon}^{\eta'}\|_r \big\}\big|_{(\xi', \eta') = (X_{t \wedge \tau_\varepsilon}^\xi, Y_{t \wedge \tau_\varepsilon}^\eta)}\Big] \\
&\leq \delta^{-1} \sqrt{\mathbb{P}(\tau_\varepsilon < t) K \mathrm{e}^{Kt} \mathbb{E}\|X_{t \wedge \tau_\varepsilon}^\xi - Y_{t \wedge \tau_\varepsilon}^\eta\|_r^2} \\
&\leq \sqrt{cK \mathrm{e}^{Kt} \mathbb{P}(\tau_\varepsilon < t)}\, \rho_{r,\delta}(\xi, \eta)
\end{aligned}
$$

for any $\xi, \eta \in \mathscr{C}_r$ with $\rho_{r,\delta}(\xi, \eta) < 1$. On the other hand, Chebyshev's inequality, (**A0**)-(**A2**) and (2.23) imply

(2.30)
$$
\begin{aligned}
\mathbb{P}(\tau_\varepsilon < t) &\leq \mathbb{P}\bigg(\int_0^t |h(s)|^2 \mathrm{d}s \geq \varepsilon^{-1} \|\xi - \eta\|_r^2\bigg) \\
&\leq c_1 \varepsilon \|\xi - \eta\|_r^{-2} \int_0^t \mathbb{E}\|X_s^\xi - Y_s^\eta\|_r^2 \mathrm{d}s \\
&\leq c_2 \varepsilon \int_0^t \mathrm{e}^{-r_0 s} \mathrm{d}s \leq \frac{c_2 \varepsilon}{r_0}
\end{aligned}
$$

for some constants $c_1, c_2 > 0$. Combining (2.30) with (2.29), we may find out a constant $c_3 > 0$ such that

(2.31)
$$
I_2(t) \leq c_3 \sqrt{\varepsilon} \mathrm{e}^{c_3 t} \rho_{r,\delta}(\xi, \eta), \quad \xi, \eta \in \mathscr{C}_r, \quad \rho_{r,\delta}(\xi, \eta) < 1.
$$

14

By Hölder's inequality and the definition of the stopping time τ_ε, we obtain

$$I_3^2(t) \leq \mathbb{E}R_{t\wedge\tau_\varepsilon}^2 - 1$$

$$\leq \mathbb{E}\exp\left(-2\int_0^{t\wedge\tau_\varepsilon}\langle h(s), \mathrm{d}W(s)\rangle - \int_0^{t\wedge\tau_\varepsilon}|h(s)|^2\mathrm{d}s\right) - 1$$

$$\leq \left(\mathbb{E}\exp\left(6\int_0^{t\wedge\tau_\varepsilon}|h(s)|^2\mathrm{d}s\right)\right)^{1/2} - 1$$

$$\leq \mathrm{e}^{3\varepsilon^{-1}\|\xi-\eta\|_r^2} - 1 \leq 3\varepsilon^{-1}\|\xi-\eta\|_r^2\mathrm{e}^{3\varepsilon^{-1}\|\xi-\eta\|_r^2},$$

where the last step is due to the inequality: $\mathrm{e}^x - 1 \leq x\,\mathrm{e}^x$, $x \geq 0$. Hence, for any $\xi, \eta \in \mathscr{C}_r$ with $\rho_{r,\delta}(\xi, \eta) < 1$, we infer that

$$I_3(t) \leq \sqrt{3}\varepsilon^{-1/2}\mathrm{e}^{\frac{3}{2}\varepsilon^{-1}\delta^2}\|\xi-\eta\|_r = \sqrt{3}\varepsilon^{-1/2}\delta\mathrm{e}^{\frac{3}{2}\varepsilon^{-1}\delta^2}\rho_{r,\delta}(\xi, \eta).$$

Combining this with (2.27), (2.28) and (2.31), we arrive at

$$\mathbb{W}_{\rho_{r,\delta}}(P_t(\xi, \cdot), P_t(\eta, \cdot)) \leq c_3\left\{\mathrm{e}^{-r_0 t/2} + \mathrm{e}^{Kt/2}\varepsilon + \varepsilon^{-1/2}\delta\mathrm{e}^{\frac{3}{2}\varepsilon^{-1}\delta^2}\right\}\rho_{r,\delta}(\xi, \eta)$$

for any $\xi, \eta \in \mathscr{C}_r$ with $\rho_{r,\delta}(\xi, \eta) < 1$. Thus, (2.21) holds by taking $t > 0$ sufficiently large and $\varepsilon = \delta \in (0, 1)$ sufficiently small. $\qquad\square$

Proof of Theorem 1.1. Since $\lim_{\|\xi\|_r\to\infty} V(\xi) = \infty$, there is a constant $R > 0$ such that $\{V \leq 4K\} \subset B_R$. By Lemmas 2.3 and 2.4, there exists $t_0 \geq t_{R,\delta/4}$ such that $\{V \leq 4K\}$ is $\rho_{r,\delta}$-small and $\rho_{r,\delta}$ is contractive for P_t for any $t \geq t_0$ and $\delta > 0$. So, in terms of Lemma 2.1, P_t has a unique probability measure π, and there exists a constant $t_1 > 0$ such that

$$(2.32) \qquad \mathbb{W}_{\rho_{r,\delta,V}}(\mu P_{t_1}, \nu P_{t_1}) \leq \frac{1}{2}\mathbb{W}_{\rho_{r,\delta,V}}(\mu, \nu), \qquad \mu, \nu \in \mathscr{P}(\mathscr{C}_r).$$

Combining this with the semigroup property, to prove (1.5) it suffices to find out a constant $C > 0$ such that

$$(2.33) \qquad \mathbb{W}_{\rho_{r,\delta,V}}(\delta_\xi P_t, \delta_\eta P_t) \leq C\rho_{r,\delta,V}(\xi, \eta), \quad t \in [0, t_1], \xi, \eta \in \mathscr{C}_r.$$

By (2.1) and (2.17), besides Hölder's inequality, there exists a constant $C > 0$ such that

$$\mathbb{W}_{\rho_{r,\delta,V}}(\delta_\xi P_t, \delta_\eta P_t) \leq \mathbb{E}\sqrt{\rho_{r,\delta}(X_t^\xi, X_t^\eta)(1 + V(X_t^\xi) + V(X_t^\eta))}$$

$$\leq \sqrt{\mathbb{E}\rho_{r,\delta}(X_t^\xi, X_t^\eta)\mathbb{E}(1 + V(X_t^\xi) + V(X_t^\eta))}$$

$$\leq C\sqrt{\rho_r(\xi, \eta)(1 + V(\xi) + V(\eta))}$$

$$= C\rho_{r,\delta,V}(\xi, \eta)$$

for any $t \in [0, t_1], \xi, \eta \in \mathscr{C}_r$. Therefore, (2.33) holds true so that (1.5) is available by in addition taking the equivalence of ρ_r and $\rho_{r,\delta}$.

Next, by (2.1), we have $\pi(V) := \int_{\mathscr{C}_r} V \mathrm{d}\pi < \infty$, so that (1.5) implies

$$\mathbb{W}_{\rho_{r,V}}(P_t(\xi, \cdot), \pi) = \mathbb{W}_{\rho_{r,V}}(\delta_\xi P_t, \pi P_t) \leq c\,\mathrm{e}^{-\lambda t}\int_{\mathscr{C}_r} \rho_{r,V}(\xi, \eta)\pi(\mathrm{d}\eta) \leq C\,\mathrm{e}^{-\lambda t}\sqrt{1 + V(\xi)}$$

for some constant $C > 0$. $\qquad\square$

3 Proofs of Proposition 1.2 and Example 1.3

Proof of Proposition 1.2 . For simplicity, we write $X(t) = X^\xi(t)$ and $X_t = X_t^\xi$. By (2.6), it is sufficient to find out a constant $c > 0$ such that

$$(3.1) \qquad \mathbb{E}\Big(\sup_{0 \le s \le t} (\mathrm{e}^{2rs}|\Lambda^X(s)|^2) \Big) \le c \Big((1+t)\|\xi\|_r^2 + \mathrm{e}^{2rt} \Big).$$

By Fubini's theorem and integration by substitution, we deduce from (1.8) that

$$
\int_0^t \int_{-\infty}^0 \mathrm{e}^{2rs} |X(s+\theta)|^2 \mu_0(\mathrm{d}\theta)\mathrm{d}s
$$

$$(3.2) \quad = \int_0^t \int_{-\infty}^{-s} \mathrm{e}^{-2r\theta} \mathrm{e}^{2r(s+\theta)} |X(s+\theta)|^2 \mu_0(\mathrm{d}\theta)\mathrm{d}s + \int_{-t}^0 \mathrm{e}^{-2\theta} \int_0^{t+\theta} \mathrm{e}^{2rs}|X(s)|^2 \mathrm{d}s \mu_0(\mathrm{d}\theta)$$

$$
\le \delta_r(\mu_0)\|\xi\|_r^2 t + \delta_r(\mu_0) \int_0^t \mathrm{e}^{2rs}|X(s)|^2 \mathrm{d}s,
$$

which, together with (1.10), leads to: for any $\varepsilon > 0$, there exists a constant $c_\varepsilon > 0$ such that

$$
\int_0^t \mathrm{e}^{2rs}|\Lambda^X(s)|^2 \mathrm{d}s \le \Big(1 + \varepsilon + (1 + 1/\varepsilon)\alpha_1 \Big) \int_0^t \mathrm{e}^{2rs}|X(s)|^2 \mathrm{d}s
$$

$$
+ (1 + 1/\varepsilon)\alpha_2 \int_0^t \int_{-\infty}^0 \mathrm{e}^{2rs}|X(s+\theta)|^2 \mu_0(\mathrm{d}\theta)\mathrm{d}s
$$

$$
\le c_\varepsilon \|\xi\|_r^2 t + \Big(1 + \alpha_1 + \alpha_2 \delta_r(\mu_0) + \varepsilon + (\alpha_1 + \alpha_2 \delta_r(\mu_0))/\varepsilon \Big) \int_0^t \mathrm{e}^{2rs}|X(s)|^2 \mathrm{d}s.
$$

Taking $\varepsilon = (\alpha_1 + \alpha_2 \delta_r(\mu_0))^{1/2}$, we find out a constant $c_1 > 0$ such that

$$(3.3) \qquad \int_0^t \mathrm{e}^{2rs}|\Lambda^X(s)|^2 \mathrm{d}s \le c_1 \|\xi\|_r^2 t + \beta_1 \int_0^t \mathrm{e}^{2rs}|X(s)|^2 \mathrm{d}s$$

where $\beta_1 > 0$ is in (1.9). Now, by Itô's formula, it follows from (1.11) that

$$
\mathrm{e}^{2rt}|\Lambda^X(t)|^2 \le |\Lambda^\xi(0)|^2 + \int_0^t \mathrm{e}^{2rs}\Big\{ c_0 + 2r|\Lambda^X(s)|^2 - \lambda_1|X(s)|^2
$$

$$(3.4)
$$

$$
+ \lambda_2 \int_{-\infty}^0 |X(s+\theta)|^2 \mu_0(\mathrm{d}\theta) \Big\}\mathrm{d}s + 2 \int_0^t \mathrm{e}^{2rs}\langle \Lambda^X(s), \sigma(X_s)\mathrm{d}W(s)\rangle.
$$

Plugging (3.2) and (3.3) into (3.4) and utilizing (**A0**) gives

$$
\mathrm{e}^{2rt}\mathbb{E}|\Lambda^X(t)|^2 \le c_2\Big((1+t)\|\xi\|_r^2 + \mathrm{e}^{2rt} \Big) - \Big(\lambda_1 - 2r\beta_1 - \lambda_2\delta_r(\mu_0) \Big) \int_0^t \mathrm{e}^{2rs}\mathbb{E}|X(s)|^2 \mathrm{d}s
$$

for some constant $c_2 > 0$. Since $\lambda_1 - 2r\beta_1 - \lambda_2\delta_r(\mu_0) > 0$, this implies

$$(3.5) \qquad \int_0^t \mathrm{e}^{2rs}\mathbb{E}|X(s)|^2 \mathrm{d}s \le c_3(1+t)\|\xi\|_r^2 + c_3\mathrm{e}^{2rt}$$

for some constant $c_3 > 0$. On the other hand, by BDG's inequality, we deduce from (1.12) and (3.2) that

$$
2 \sup_{0 \le s \le t} \left| \int_0^s e^{2ru} \langle \Lambda^X(u), \sigma(X_u) dW(u) \rangle \right|
$$

(3.6)
$$
\le 8\sqrt{2}\, \mathbb{E} \left(\sup_{0 \le s \le t} (e^{2rs}|\Lambda^X(s)|^2) \int_0^t e^{2rs} \|\sigma(X_s)\|_{\mathrm{HS}}^2 ds \right)^{1/2}
$$

$$
\le \frac{1}{2} \mathbb{E} \left(\sup_{0 \le s \le t} (e^{2rs}|\Lambda^X(s)|^2) \right) + c_4 \left(e^{2rt} + t\|\xi\|_r^2 + \int_0^t e^{2rs} \mathbb{E}|X(s)|^2 ds \right)
$$

for some $c_4 > 0$. Thus, by taking (3.3) and (3.4) into account and making use of (3.6) and (**A0**), there exists a constant $c_5 > 0$ such that

$$
\mathbb{E} \left(\sup_{0 \le s \le t} (e^{2rs}|\Lambda^X(s)|^2) \right) \le c_5 \left((1+t)\|\xi\|_r^2 + e^{2rt} + \int_0^t e^{2rs} \mathbb{E}|X(s)|^2 ds \right).
$$

Henceforth, (3.1) follows directly from (3.5). $\qquad\square$

Proof of Example 1.3. It suffices to verify (**A0**)-(**A2**) and conditions in Proposition 1.2.
Thanks to Hölder's inequality, one has

(3.7) $\quad |G(\xi) - G(\eta)|^2 \le \gamma_1^2 \int_{-\infty}^0 |\xi(\theta) - \eta(\theta)|^2 \mu_0(d\theta) \le \gamma_1^2 \delta_r(\mu_0) \|\xi - \eta\|_r^2, \quad \xi, \eta \in \mathscr{C}_r.$

Therefore, (**A0**) holds for $\alpha = \gamma_1 \sqrt{\delta_r(\mu_0)} < 1$ owing to (1.14). By Hölder's inequality and (3.7), we can find some constants $c_1, c_2 > 0$ such that

$$
\langle \xi(0) - \eta(0) - (G(\xi) - G(\eta)), b(\xi) - b(\eta) \rangle^+ + |\sigma(\xi) - \sigma(\eta)|^2
$$

$$
\le \left((\xi(0) - \eta(0) - (G(\xi) - G(\eta))) \left\{ -\gamma_3(\xi(0) - \eta(0)) + \gamma_5 \int_{-\infty}^0 (\xi(\theta) - \eta(\theta)) \mu_0(d\theta) \right\} \right)^+
$$

$$
+ \gamma_2^2 \left(\int_{-\infty}^0 |\xi(\theta) - \eta(\theta)| \mu_0(d\theta) \right)^2
$$

$$
\le c_1 \left\{ |\xi(0) - \eta(0)|^2 + |G(\xi) - G(\eta)|^2 \right\}
$$

$$
\le c_2 \|\xi - \eta\|_r^2, \quad \xi, \eta \in \mathscr{C}_r,
$$

where the first inequality is due to

$$
-\gamma_4 \langle \xi(0) - G(\xi) - (\eta(0) - G(\eta)), (\xi(0) - G(\xi))^{1/3} - (\eta(0) - G(\eta))^{1/3} \rangle \le 0.
$$

Consequently, (**A1**) holds true. According to the formula of $\sigma(\xi)$, (**A1**) holds trivially.
Finally, we verify conditions in Proposition 1.2. Obviously, (1.8) holds with $\delta_r(\mu_0) = \frac{1}{r_0(r_0 - 2r)} < \infty$ due to $r_0 > 2r$. Thanks to $G(\xi_0) = \mathbf{0}$ and (3.7), (1.10) holds for $\alpha_1 = 0$ and $\alpha_2 = \gamma_1^2$. By Hölder's inequality, it follows that

(3.8) $\qquad |\sigma(\xi)|^2 \le 1 + \frac{1}{\alpha} + (1+\alpha)\gamma_2^2 \int_{-\infty}^0 |\xi(\theta)|^2 \mu_0(d\theta), \quad \alpha > 0, \xi \in \mathscr{C}_r.$

Next, by using (3.8), for any $\varepsilon > 0$ and $\xi \in \mathscr{C}_r$,

$$2(\xi(0) - G(\xi))b(\xi) + |\sigma(\xi)|^2$$

$$\leq 1 + \frac{1}{\alpha} - 2\gamma_3\xi^2(0) + 2(\gamma_5 + \gamma_1\gamma_3)\xi(0)\int_{-\infty}^{0}\xi(\theta)\mu_0(\mathrm{d}\theta) + (1+\alpha)\gamma_2^2\int_{-\infty}^{0}|\xi(\theta)|^2\mu_0(\mathrm{d}\theta)$$

$$\leq 1 + \frac{1}{\alpha} - \left(2\gamma_3 - (\gamma_5 + \gamma_1\gamma_3)\varepsilon\right)\xi^2(0) + \left(\frac{1}{\varepsilon}(\gamma_5 + \gamma_1\gamma_3) + (1+\alpha)\gamma_2^2\right)\int_{-\infty}^{0}|\xi(\theta)|^2\mu_0(\mathrm{d}\theta).$$

Taking $\varepsilon = \sqrt{\delta_r(\mu_0)}$ and $\alpha \in (0,1)$ sufficiently small and combining (1.14), we conclude that (1.11) holds.

$\square$

References

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

[2] Asker, H., Well-posedness and exponential estimates for the solutions to neutral stochastic functional differential equations with infinite delay, arXiv:1805.10674v3.

[3] Bao, J., Wang, F.-Y., Yuan, C., Derivative formula and Harnack inequality for degenerate functional SDEs, *Stoch. Dyn.*, **13** (2013), 1250013, 22pp.

[4] Bao, J., Wang, F.-Y., Yuan, C., Asymptotic Log-Harnack Inequality and Applications for Stochastic Systems of Infinite Memory, to appear on Stochastic Process. Appl., https://doi.org/10.1016/j.spa.2018.12.010,

[5] Bao, J., Yin, G., Yuan, C., Stationary distributions for retarded stochastic differential equations without dissipativity, *Stochastics*, **89** (2017), 530–549.

[6] Bass, R., Diffusions and elliptic operators, Springer, New York, 1998.

[7] Butkovsky, O., Subgeometric rates of convergence of Markov processes in the Wasserstein metric, *Ann. Appl. Probab.*, **24** (2014), 526–552.

[8] Butkovsky, O., Scheutzow, M., Invariant measures for stochastic functional differential equations, *Electron. J. Probab.*, **22** (2017), 1–23

[9] Butkovsky, O., Kulik, K., Scheutzow, M., Generalized coupling and ergodic rates for SPDEs and other Markov models, arXiv:1806.00395.

[10] Chen, M.-F., Li, S., Coupling methods for multidimensional diffusion processes, *Ann. Probab.*, **17** (1989), 151–177.

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

[12] Da Prato, G., Zabczyk, J., Ergodicity for Infinite-Dimensional Systems, in: London Mathematical Society, Lecture Note Series, vol. 229, Cambridge University Press, Cambridge, 1996.

[13] Dolbeault, J., Mouhot, C., Schmeiser, C., Hypocoercivity for linear kinetic equations conserving mass, *Trans. Amer. Math. Soc.*, **367** (2015), 3807–3828.

[14] Durmus, A., Fort, G., Moulines, É., Subgeometric rates of convergence in Wasserstein distance for Markov chains, *Ann. Inst. Heri Poincaré Probab. Sta.*, **52** (2016), 1799-1822.

[15] M. Grothaus, P. Stilgenbauer, Hypocoercivity for kolmogorov backward evolution equations and applications, *J. Funct. Anal.*, **267** (2014), 3515–3556.

[16] M. Grothaus, F.-Y. Wang, Weak Poincaré inequalities for convergence rate of degenerate diffusion processes, arXiv:1703.04821.

[17] Hairer, M., Mattingly, Jonathan C., Ergodicity of the 2D Navier-Stokes equations with degenerate stochastic forcing, *Ann. of Math.*, **164** (2006), 993–1032.

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

[19] Hale, J., Forward and backward continuation of neutral functional differential equations, *J. Differential Equations*, **9** (1971), 168-181.

[20] Hale, J., Verduyn Lunel, S. M., *Introduction to functional differential equations*, New York, Springer-Verlag, 1993.

[21] Kulik, A., *Ergodic behavior of Markov processes with applications to limit theorems*, De Gruyter Studies in Mathematics, 67, De Gruyter, Berlin, 2018.

[22] Kulik, A., Scheutzow, M., Well-posedness, stability, and sensitivities for stochastic delay equations: a generalized coupling approach, arXiv:1808.06050.

[23] Majda, A. J., Tong, X. T., Geometric ergodicity for piecewise contracting processes with applications for tropical stoshastic lattice models, *Commu. Pure Appl. Math.*, **LXIX** (2016), 1110–1153.

[24] Mao, X., *Stochastic Differential Equations and Applications*, Second Ed., Horwood Publishing Limited, Chichester, 2008.

[25] Mattingly, J. C., Stuart, A. M., Higham, D. J., Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise, *Stochastic Process. Appl.*, **101** (2002), 185–232.

[26] Meyn, S. P., Tweedie, R. L., *Markov Chains and Stochastic Stability*, Springer-Verlag, Berlin, 1993.

[27] Scheutzow, M., Exponential growth rates for stochastic delay differential equations, *Stoch. Dyn.*, **5** (2005), 163–174.

[28] Scheutzow, M., Steinsaltz, D., Chasing balls through martingale fields, *Ann. Probab.*, **30** (2002), 2046–2080.

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

[30] Villani, C., Hypocoercivity, *Mem. Amer. Math. Soc.*, **202** (2009), iv+141.

[31] Wang, F.-Y., *Functional Inequalities, Markov Semigroups and Spectral Theory*, Science Press, 2005, Beijing.

[32] Wang, F.-Y., *Harnack inequalities and Applications for Stochastic Partial Differential Equations*, Springer, 2013, Berlin.

[33] Wang, F.-Y., Hypercontractivity and applications for stochastic Hamiltonian systems, *J. Funct. Anal.*, **272** (2017), 5360–5383.

[34] Wang, F.-Y., Yuan, C., Harnack inequalities for functional SDEs with multiplicative noise and applications, *Stochastic Process. Appl.*, **121** (2011), 2692–2710.