Invariant measures for stochastic functional differential equations

Oleg Butkovsky[*‡§]   Michael Scheutzow[‡¶]

Abstract

We establish new general sufficient conditions for the existence of an invariant measure for stochastic functional differential equations and exponential or subexponential convergence to the equilibrium. The obtained conditions extend the Veretennikov–Khasminskii conditions for SDEs and are optimal in a certain sense.

1. Introduction

While ergodic properties of stochastic differential equations (SDEs) are more or less understood by now, less is known about ergodic properties of stochastic functional (or delay) differential equations (SFDEs). In this article we establish new general sufficient conditions for existence of an invariant measure for SFDEs and obtain estimates for the rate of convergence to the equilibrium.

SFDEs in general have quite a peculiar ergodic behavior that can be very different from the ergodic behavior of SDEs. Let us briefly describe the main features. First of all, as was shown in [24], an SFDE might have a *reconstruction property*. Namely, consider the equation

$$dX^{(x)}(t) = f(X^{(x)}(t-1))dt + g(X^{(x)}(t-1))dW(t), \quad t \geq 0, \tag{1.1}$$

$$X^{(x)}(t) = x(t), \quad t \in [-1, 0].$$

[*]Technion — Israel Institute of Technology, Faculty of Industrial Engineering and Management Haifa, 3200003, Israel.
[‡]Technische Universität Berlin, Institut für Mathematik, MA 7-5, Fakultät II, Straße des 17. Juni 136, 10623 Berlin, Germany.
[§]E-mail: oleg.butkovskiy@gmail.com
[¶]E-mail: ms@math.tu-berlin.de

where $f\colon \mathbb{R} \to \mathbb{R}$ is a Lipschitz function, $g\colon \mathbb{R} \to \mathbb{R}$ is a positive strictly increasing bounded Lipschitz function, $x\colon [-1, 0] \to \mathbb{R}$ is a continuous function and W is a 1-dimensional Brownian motion. It turns out that if for some $N > 0$ (that might be arbitrarily large) one observes a single piece of trajectory $\{X(t, \omega), t \in [N, N + 1]\}$, then with probability 1 one can reconstruct the initial condition $\{X(t), t \in [-1, 0]\}$ of the SFDE. Clearly, this is not the case for SDEs.

As a result of this, the solution to (1.1) is not a strong Feller process and it does not have a *mixing property*. Indeed, if $x \neq y$, then the measures $\mathrm{Law}\{X^{(x)}(t), t \in [N, N + 1]\}$ and $\mathrm{Law}\{X^{(y)}(t), t \in [N, N + 1]\}$ are mutually singular for any $N > 0$. Therefore, one cannot hope to construct a classical coupling between these measures to show asymptotic stability.

SFDEs might also have a *resonance property*. If one considers a delay version of a classical Ornstein–Uhlenbeck process,

$$dX(t) = -\lambda X(t - 1)\, dt + dW(t), \quad t \geq 0, \tag{1.2}$$

where $\lambda > 0$, then (contrary to the non-delay case) for large enough λ (more precisely, $\lambda \geq \pi/2$) this equation does not have an invariant measure [14]. Moreover, for large λ the equation oscillates to infinity with rapidly increasing diameter of oscillations.

Due to the above mentioned challenges the question of existence of an invariant measure and rate of convergence to the equilibrium remained open even for a relatively simple SFDE (1.1) if f is not affine. In the current paper we present an answer to this question.

Let us recall that there are two quite general approaches that are used to study the ergodic properties of Markov processes. The first approach is based on functional inequalities, see, e.g., [1]. The second approach is based on the concept of small sets and utilizes the coupling method, see, e.g., [17]. Using these techniques, it was shown that if the drift vector field of an SDE points towards the origin (the so-called Veretennikov–Khasminskii condition), then, under some further non-degeneracy assumptions, the SDE has a unique invariant measure and converges to it in total variation, see [4], [13], [26]. More general SDEs are treated in [11].

Unfortunately, these methods are not applicable for SFDEs due to their lack of mixing properties. Note though that ergodic properties of affine SFDEs can be treated by comparison with the deterministic case and by studying the fundamental solutions, see [7], [14], [18], [20]. However this technique also does not work for non-affine SFDEs.

Some sufficient conditions for the existence of an invariant measure for SFDEs are obtained in [12, Theorem 3]. Let us note though that it might be quite hard to verify these conditions in practice.

To overcome these difficulties and to derive verifiable sufficient conditions M. Hairer, J. Mattingly and M. Scheutzow suggested a new approach targeted specifically at Markov processes with bad mixing properties [10]. They introduced a new concept of a d-small set, and showed that under certain conditions (much weaker than mixing) a Markov process has a unique invariant measure and converges to it. The price to pay is that this convergence occurs in the Wasserstein metric rather than in total variation. This approach was further developed in [3].

In this paper we apply this general approach to SFDEs. The main obstacle here is to construct a proper Lyapunov function. Due to the memory property it is much more challenging than in the SDE case. Indeed, a solution to an SFDE is an infinite dimensional Markov process with non-locally compact state space and rather involved generator. We develop a new technique inspired by some ideas from [23].

Another obstacle was to obtain a condition that is general enough to cover drifts in (1.1) of the form $f(x) = -|x(-1)|^{\beta}\,\mathrm{sign}(x(-1))$, $\beta \in [0, 1)$ (in this case an invariant

measure exists), but not "too general" since (1.1) with the drift $f(x) = -\lambda x(-1)$ does not have an invariant measure for $\lambda \geq \pi/2$.

The obtained result can be formulated as follows: one should check that the drift vector field $f(x)$ points towards the origin only for "typical" x. This extends and generalizes the corresponding theorems for SFDEs in [23], [10], [3]. The obtained conditions and rates are optimal in a certain sense. We explain our result in more details below in Section 2.

Note also that there is an alternative fruitful approach, which is also suitable for SFDEs, that was suggested and developed in [10], [16]. It is based on the generalized coupling method. Using this approach it is possible to establish uniqueness of an invariant measure and asymptotic stability under some natural conditions. However, this approach does not allow directly to obtain the results on existence of an invariant measure and on the convergence rate. Therefore we do not use it here.

The paper is organized as follows. We formulate and discuss our main results in Section 2. Section 3 contains specific applications of our results to different SFDEs as well as some counterexamples. All proofs are placed in Section 4.

2. Main results

We assume that all random objects are defined on a common probability space $(\Omega, \mathcal{F}, \mathrm{P})$. Fix $r > 0$, positive integers d, m and let $\mathcal{C} := \mathcal{C}([-r, 0], \mathbb{R}^d)$ be the space of continuous functions endowed with the supremum norm $\|\cdot\|$. We study a stochastic functional differential equation

$$dX(t) = f(X_t)\, dt + g(X_t)\, dW(t), \quad t \geq 0 \tag{2.1}$$
$$X_0 = x$$

where $f : \mathcal{C} \to \mathbb{R}^d$ and $g : \mathcal{C} \to \mathbb{R}^{d \times m}$ are measurable functions, W is an m-dimensional Brownian motion, the initial condition $x \in \mathcal{C}$, and we used the standard notation $X_t(s) := X(t + s)$, $s \in [-r, 0]$.

For a matrix $M \in \mathbb{R}^{d \times m}$ we denote by $\|M\|$ its Frobenius norm, that is, $\|M\| := \sqrt{\sum M_{ij}^2}$. For a real a we put $a_+ := \max(a, 0)$. We suppose that the drift and diffusion of (2.1) satisfy the following condition:

Assumption A1. The drift f is continuous and bounded on bounded subsets of $\mathcal{C}$. The diffusion g is non-degenerate, that is, for any $x \in \mathcal{C}$ the matrix $g(x)$ admits a right inverse $g^{-1}(x)$ and

$$\sup_{x \in \mathcal{C}} \|g^{-1}(x)\| < \infty.$$

Furthermore, f satisfies the one-sided Lipschitz condition and g is Lipschitz. Namely, there exists $C > 0$ such that for any $x, y \in \mathcal{C}$ we have

$$\langle f(x) - f(y), x(0) - y(0) \rangle_+ + \|g(x) - g(y)\|^2 \leq C\|x - y\|^2.$$

It follows from [21] that under Assumption **A1** SFDE (2.1) has a unique strong solution. Moreover, this solution $X = (X_t)_{t \geq 0}$ is a strong Markov process with the state space $(\mathcal{C}, \mathcal{B}(\mathcal{C}))$, see Proposition 4.1 below. We denote the transition probabilities of X by $P_t(x, \cdot)$, where $t \geq 0$, $x \in \mathcal{C}$.

In this article we study the invariant probability measures of X. Further, we will drop the word "probability" and refer to these measures just as invariant measures.

It was shown in [10, Theorems 3.1 and 3.7] (see also [16, Section 6.1]) that under **A1** X has at most one invariant measure and if it has one, then the transition probabilities weakly converge to this measure. Note however that **A1** does not guarantee the *existence* of the invariant measure of X. Indeed, the equation

$$dX(t) = dW(t), \quad t \geq 0$$

satisfies **A1** but does not have an invariant measure. Also assumption **A1** alone does not imply any bound on convergence rate, see [10, Remark 3.4].

We will provide two different sets of conditions for the existence of an invariant measure for SFDE (2.1) and present upper bounds for the rate of convergence to the equilibrium. To formulate our results we need to introduce some notation.

Let $(E, \mathcal{B}(E))$ be a Polish space. Recall that the *Wasserstein* (or *Kantorovich*) *distance* between two probability measures μ, ν on $(E, \mathcal{B}(E))$ is defined as follows:

$$W_d(\mu, \nu) := \inf \mathsf{E}d(X, Y),$$

where d is a lower semicontinuous metric on E and the infimum is taken over all random variables X, Y that are distributed as μ and ν, correspondingly. If the metric d is the discrete metric, that is $d(x, y) = \mathbb{1}(x \neq y)$, then the Wasserstein distance is equivalent to the *total variation distance* which is defined by

$$d_{TV}(\mu, \nu) := \inf \mathsf{P}(X \neq Y) = \sup_{A \in \mathcal{B}(E)} |\mu(A) - \nu(A)|,$$

where again the infimum is taken over all random variables X, Y that are distributed as μ and ν, correspondingly. In the paper we will consider only bounded distances d. In this case, convergence in total variation implies convergence in the Wasserstein metric; the latter is also equivalent to the weak convergence (see, e.g., [2]).

Throughout the paper, we will take the space $\mathcal{C}$ as the state space E. For $x \in \mathcal{C}$ we denote the diameter of the range of x by

$$D(x) := \sup_{t_1, t_2 \in [-r, 0]} |x(t_1) - x(t_2)|.$$

As in [10, Section 5], we consider the following family of distances on $\mathcal{C}$:

$$d_\rho(x, y) := \frac{\|x - y\|}{\rho} \wedge 1, \quad x, y \in \mathcal{C},$$

where $\rho > 0$.

Now we are in position to present our main results. We consider two different groups of conditions which are sufficient for the existence of invariant measure and exponential or subexponential convergence to the equilibrium.

Assumption A2 (Exponential convergence)**.** The diffusion g is globally bounded and the drift f is sublinear. The latter means that there exist constants $\beta \in [0, 1)$, $C > 0$ such that

$$|f(x)| \leq C(1 + \|x\|^\beta), \quad x \in \mathcal{C}. \tag{2.2}$$

Furthermore, there exist constants $\sigma, M > 0$ and a function $\kappa : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\lim_{z \to \infty} \left(\kappa(z) z^{-\beta} \right) = \infty$ and

$$\langle f(x), x(0) \rangle \leq -\sigma |x(0)|, \quad \text{for any } x \in \mathcal{C} \text{ with } D(x) \leq \kappa(|x(0)|) \text{ and } |x(0)| \geq M. \quad (2.3)$$

Assumption A3 (Subexponential convergence). The diffusion g and drift f are globally bounded. Furthermore, there exist $\alpha \in (0, 1)$, $\sigma > 0$, $M > 0$ and a function $\kappa : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\lim_{t \to \infty} \left(\kappa(z)/\sqrt{\log z} \right) = \infty$ and

$$\langle f(x), x(0) \rangle \leq -\sigma |x(0)|^\alpha, \quad \text{for any } x \in \mathcal{C} \text{ with } D(x) \leq \kappa(|x(0)|) \text{ and } |x(0)| \geq M. \quad (2.4)$$

We will also present results concerning convergence in the total variation distance. To state these results we need an additional assumption on the structure of the drift and the diffusion.

Assumption A4 (Convergence in total variation). The drift f is globally Lipschitz and the diffusion g depends on x only through $x(0)$.

Theorem 2.1. *Suppose that Assumptions **A1** and **A2** hold. Then SFDE (2.1) has a unique invariant measure π and the transition probabilities $P_t(x, \cdot)$ converge to it exponentially in the Wasserstein metric. That is, for any $\rho > 0$ there exist $C > 0$, $\lambda_1 > 0$, $\lambda_2 > 0$ such that for all $x \in \mathcal{C}$ we have*

$$W_{d_\rho}(P_t(x, \cdot), \pi) \leq C e^{\lambda_1 |x(0)| + D(x)} e^{-\lambda_2 t}, \quad t \geq 0. \quad (2.5)$$

*Moreover, if additionally Assumption **A4** holds, then the convergence in the Wasserstein metric in (2.5) can be replaced by convergence in total variation metric.*

It is interesting to compare the obtained theorem with the corresponding result for SDEs. Recall that in the non-delay case the following condition is sufficient [25] for existence and uniqueness of the invariant measure and exponential convergence of transition probabilities in total variation:

$$\langle f(y), y \rangle \leq -\sigma |y|, \quad |y| \geq M, y \in \mathbb{R}^d, \quad (2.6)$$

where $M > 0$, $\sigma > 0$. In other words, for large enough $y \in \mathbb{R}^d$ the drift f should point towards the origin. Therefore, condition (2.3) is a direct equivalent of (2.6) for SFDEs. We can call it the *extended Veretennikov–Khasminskii condition*.

Note that it is sufficient to check (2.3) only for trajectories x with not too large diameters. This is quite important as it makes verifying (2.3) in practice much easier, see Section 3. The intuition here is the following. As one can see from the results of Section 4 below, for large enough n with high probability $D(X_n)$ is approximately of the size $O(|X(n)|^\beta)$ regardless of the initial conditions. Thus, it is very unlikely that the trajectory will have a much bigger diameter. Even if it happens, one can just wait till the trajectory has a smaller diameter and then the drift would point towards the origin. Thus, one has to check the extended Veretennikov–Khasminskii condition only for "typical" trajectories. Note that this additional assumption $\lim_{z \to \infty} \left(\kappa(z) z^{-\beta} \right) = \infty$ is optimal, see Section 3 for counterexamples.

The convergence in the Wasserstein metric in (2.5) cannot be replaced by the convergence in total variation without additional Assumption **A4**. This is due to the reconstruction property discussed above. If the diffusion does not depend on the past, then SFDE does not have the reconstruction property and the convergence occurs in total variation.

Let us also mention that one cannot hope to replace (2.3) by something like

$$\langle f(x), x(-1) \rangle \leq -\sigma |x(-1)|, \quad |x(-1)| \geq M.$$

Indeed, the delayed Ornstein–Uhlenbeck equation (1.2) satisfies this assumption, but it does not have an invariant measure.

Let us move on to our second main result that concerns subgeometrical convergence.

Theorem 2.2. *Suppose that Assumptions **A1** and **A3** hold. Then SFDE* (2.1) *has a unique invariant measure π and the transition probabilities $P_t(x, \cdot)$ converge to it subexponentially in the Wasserstein metric. That is, for any $\rho > 0$ there exist $C > 0$, $\lambda_1 > 0$, $\lambda_2 > 0$ such that for all $x \in \mathcal{C}$ we have*

$$W_{d_\rho}(P_t(x, \cdot), \pi) \leq C e^{\lambda_1 |x(0)|^\alpha + \lambda_1 D(x)^2} e^{-\lambda_2 t^{\alpha/(2-\alpha)}}, \quad t \geq 0. \tag{2.7}$$

*Moreover, if additionally Assumption **A4** holds, then the convergence in the Wasserstein metric in* (2.7) *can be replaced by convergence in total variation metric.*

We see that in the subgeometrical case it is also enough to check the extended Veretennikov–Khasminskii condition only for trajectories with not too big diameter. The explanation is the same. It is worth mentioning that since the drift f "pushes" to the origin weaker than in the exponential case, one has to check (2.4) for a slightly bigger set of x than just "typical trajectories".

We also would like to mention that the obtained rate of convergence to infinity in the right-hand side of (2.7) matches the corresponding rate for the SDE case. The latter cannot be improved, see [8, Section 7.1].

The proofs of Theorems 2.1 and 2.2 are postponed till Section 4.

Convention on constants. Throughout the paper, we denote by C a positive constant whose value may change from line to line.

3. Examples and counterexamples

In this section we present a number of examples showing how the theoretical results from Section 2 can be used for studying convergence of SFDEs. In addition to it, we provide some counterexamples that show the optimality (in a certain sense) of Assumptions **A2** and **A3**.

We begin with the following example.

Example 3.1. Let $d = m = 1$. Consider an equation

$$dX(t) = h(X(t-r))dt + g(X_t)dW(t), \quad t \geq 0, \tag{3.1}$$

where the memory $r \geq 0$, $h\colon \mathbb{R} \to \mathbb{R}$ is a smooth function such that $h(z) = -|z|^\gamma \operatorname{sign} z$ for $|z| \geq 1$, $\gamma \in (-1, 1)$ and the diffusion g is bounded Lipschitz and non-degenerate.

Clearly, Assumption **A1** holds and hence equation (3.1) has a unique strong solution. Let us check **A2** and **A3**. Put $\kappa(z) := z^{(1+\gamma)/2}$. Note that there exists large enough M_0 such that for any $x \in \mathcal{C}$ with $D(x) \leq \kappa(|x(0)|)$ and $|x(0)| \geq M_0$ we have

$$|x(-r)| \geq |x(0)| - D(x) \geq |x(0)| - |x(0)|^{(1+\gamma)/2} \geq 1 \tag{3.2}$$

and hence $h(x(-r)) = -\operatorname{sign}(x(-r))|x(-r)|^\gamma$.

Take now any $x \in \mathcal{C}$ with $D(x) \leq \kappa(|x(0)|)$ and $|x(0)| \geq M_0$. Using (3.2), we derive

$$
\begin{aligned}
h(x(-r))x(0) &= -\operatorname{sign}(x(-r))|x(-r)|^\gamma x(0) \\
&= -\operatorname{sign}(x(-r))|x(-r)|^\gamma (x(-r) + x(0) - x(-r)) \\
&\leq -|x(-r)|^{\gamma+1} + |x(-r)|^\gamma D(x) \\
&\leq -\big|x(0) - |x(0)|^{(1+\gamma)/2}\big|^{\gamma+1} + (|x(0)|^\gamma + D(x)^\gamma + 1)D(x) \\
&\leq -\big|x(0) - |x(0)|^{(1+\gamma)/2}\big|^{\gamma+1} + 3|x(0)|^{(1+\gamma)/2 + (\gamma \vee 0)}.
\end{aligned}
$$

This implies that there exists large enough $M > M_0$, such that if $|x(0)| \geq M$ and $D(x) \leq \kappa(|x(0)|)$, then

$$h(x(-r))x(0) \leq -\frac{1}{2}|x(0)|^{\gamma+1}.$$

Now if $\gamma \in [0,1]$, then **A2** holds. Therefore, by Theorem 2.1 SFDE (3.1) has a unique invariant measure and converges to it exponentially in the Wasserstein metric.

If $\gamma \in (-1, 0)$, then **A3** holds. In this case we apply Theorem 2.2. We obtain that (3.1) still has a unique invariant measure but converges to it subexponentially with the rate given in (2.7). $\qquad\square$

Remark 3.2. In Example 3.1 it was crucial that it was sufficient to check condition (2.3) or (2.4) only for $x \in \mathcal{C}$ with not "too large" diameter. Evidently, these conditions are not satisfied for **all** $x \in \mathcal{C}$. Thus, the exponential/subexponential ergodicity of (3.1) cannot be obtained by [10, Remark 5.2] or [3, Theorem 3.3].

Example 3.3. Using the same method we can study more general equations. Let $d, m \in \mathbb{N}$, $r \geq 0$. We are interested in ergodic properties of the SFDE

$$dX(t) = h\Big(\int_{-r}^{0} X(t+s)\mu(ds)\Big)dt + g(X_t)dW(t), \quad t \geq 0, \tag{3.3}$$

where $h : \mathbb{R}^d \to \mathbb{R}^d$ is a smooth function with $h(z) := -z|z|^{\gamma-1}$ for $|z| > 1$, $\gamma \in (-1, 1)$; μ is a finite signed measure with $\mu([-r, 0]) > 0$; g is as in Example 3.1. We see that the drift and diffusion of equation (3.3) satisfy Assumption **A1** and thus this equation has a unique strong solution.

In order to verify **A2** and **A3**, we choose again $\kappa(z) := z^{(1+\gamma)/2}$. We consider the Jordan decomposition of the measure μ:

$$\mu = \mu^+ - \mu^-,$$

where μ^+ and μ^- are two finite nonnegative measures, and note that for any $x \in \mathcal{C}$ with $D(x) \leq \kappa(|x(0)|)$ we have

$$\Big|\int_{-r}^{0} x(s)\mu(ds)\Big| \geq \Big|\int_{-r}^{0} x(s)\mu^+(ds)\Big| - \Big|\int_{-r}^{0} x(s)\mu^-(ds)\Big|$$

$$\geq \mu^+([-r, 0])|x(0)| - \int_{-r}^{0} |x(s) - x(0)|\mu^+(ds)$$

$$- \mu^-([-r, 0])|x(0)| - \int_{-r}^{0} |x(s) - x(0)|\mu^-(ds)$$

$$\geq |x(0)|\mu([-r, 0]) - D(x)(\mu^+([-r, 0]) + \mu^-([-r, 0]))$$

$$\geq c_1|x(0)| - c_2|x(0)|^{(1+\gamma)/2},$$

where $c_1 = \mu([-r, 0])$ and $c_2 = (\mu^+([-r, 0]) + \mu^-([-r, 0]))$. By our assumptions, we have $c_1 > 0$. The verification of **A2** and **A3** is completed exactly as in Example 3.1. Thus, applying Theorems 2.1 and 2.2, we obtain that X has a unique invariant measure and converges to it exponentially if $\gamma \in [0, 1)$ or subexponentially if $\gamma \in [-1, 0)$. $\qquad\square$

Now we move on and present some counterexamples to demonstrate a certain optimality of the conditions in Theorems 2.1 and 2.2.

First, we consider the case $\beta = 0$. The next example shows that in this case it may happen that no invariant measure exists if the drift and diffusion satisfy all the conditions of Theorem 2.1 with the only exception that the condition $\lim_{z\to\infty} \kappa(z) = \infty$ in Assumption **A2** is replaced by $\liminf_{z\to\infty} \kappa(z) \geq N$, where $N > 0$ is an arbitrarily large constant.

Example 3.4. Let $d = m = 1$, $r = 2$, $N > 0$. Put $\kappa(z) := N$, $z \geq 0$. Consider an equation

$$dX(t) = f(X_t)dt + dW(t), \quad t \geq 0$$

where f is a Lipschitz continuous function which takes values in $[-1, A]$, and satisfies $f(x) = A$ whenever $D(x) \geq N + 1$ and $\langle f(x), x(0) \rangle \leq -|x(0)|$ whenever $D(x) \leq N$ and $|x(0)| \geq 1$. We claim that $A > 0$ can be chosen in such a way that for every fixed initial condition we have

$$\lim_{t \to +\infty} X(t) = +\infty \quad \text{a.s.} \tag{3.4}$$

This would imply in particular that X does not have an invariant measure.

To verify the claim we fix the initial condition $x \in \mathcal{C}$ and introduce an auxiliary sequence

$$Y(n) := x(0) - n + W(n) + A \sum_{i=1}^{n-1} \mathbb{1}\big(W(i) - W(i-1) \geq N + 2\big), \quad n \in \mathbb{Z}_+.$$

Since the drift f is bounded from below by -1, we derive for $n \in \mathbb{Z}_+$, $n \geq 1$

$$\begin{aligned}
X(n+1) - X(n) &= W(n+1) - W(n) + \int_n^{n+1} f(X_t)dt \\
&\geq W(n+1) - W(n) - 1 + A\,\mathbb{1}\big(\inf_{t \in [n,n+1]} D(X_t) \geq N + 1\big) \\
&\geq W(n+1) - W(n) - 1 + A\,\mathbb{1}\big(X(n) - X(n-1) \geq N + 1\big) \\
&\geq W(n+1) - W(n) - 1 + A\,\mathbb{1}\big(W(n) - W(n-1) \geq N + 2\big) \\
&= Y(n+1) - Y(n), \tag{3.5}
\end{aligned}$$

where we also used the fact that the memory $r = 2$ and hence $X(n) - X(n-1) \geq N + 1$ implies that $D(X_t) \geq N + 1$ for all $t \in [n, n+1]$. Recall that by definition $X(0) = x(0) = Y(0)$ and $X(1) \geq x(0) - 1 + W(1) = Y(1)$. Therefore (3.5) implies $X(n) \geq Y(n)$ for any $n \in \mathbb{Z}_+$.

By the strong law of large numbers,

$$Y(n)/n \to -1 + A\mathsf{P}(\xi \geq N + 2), \quad \text{a.s. whenever } n \to \infty,$$

where ξ denotes the standard Gaussian random variable. Hence by taking large enough A we get $Y(n)/n \to 1$, a.s. as $n \to \infty$. Since $X(n) \geq Y(n)$, this yields (3.4); thus the claim is proved and the process X does not have an invariant measure. $\qquad\square$

Next we consider the case $\beta \in (0, 1)$. We show that the condition $\lim_{z \to \infty} \kappa(z)z^{-\beta} = \infty$ in **A2** cannot be replaced with the condition $\liminf_{z \to \infty} \kappa(z)z^{-\beta} \geq N$. Since we need to construct an example with unbounded κ, the proof here will be different from the proof in Example 3.4.

Example 3.5. Let $\beta \in (0, 1)$, $N > 1$. Put $\kappa(z) := (N-1)z^{\beta}$, $z \geq 0$. Consider the following stochastic delay equation for $d = m = 1$, $r = 2$:

$$dX(t) = f(X_t)dt + dW(t), \tag{3.6}$$

where f is a Lipschitz continuous function such that $f(x) = 5Nx(0)^{\beta}$ if $x(0) \geq 1$ and $D(x) \geq Nx(0)^{\beta}$. Similar to Example 3.4, one can easily extend f in such a way that Assumptions **A1** and **A2** hold with the only exception that condition (2.3) is satisfied for all $x \in \mathcal{C}$ with $D(x) \leq (N-1)|x(0)|^{\beta}$ and $|x(0)| \geq 1$. Let us prove that SFDE (3.6) does not have an invariant measure.

Put $z_0 := (2N)^{1/(1-\beta)}$. We need the following technical statement.

Lemma 3.6. Let $x \in \mathcal{C}$ be such that for some $t_1, t_2 \in [-2, 0]$ we have $x(t_2) \geq z_0$ and

$$x(t_1) \geq x(t_2) + 2Nx(t_2)^{\beta}.$$

Then $D(x) \geq N|x(0)|^{\beta}$.

Proof. We consider two different cases. If $x(0) < 0$, then

$$D(x) \geq x(t_2) - x(0) \geq N^{1/(1-\beta)} + |x(0)| = N|x(0)|^\beta \Big(\frac{N^{\beta/(1-\beta)}}{|x(0)|^\beta} + \frac{|x(0)|^{1-\beta}}{N} \Big).$$

If now $|x(0)|^{1-\beta} > N$, then by above $D(x) \geq N|x(0)|^\beta$. If $|x(0)|^{1-\beta} \leq N$, then $N^{\beta/(1-\beta)} \geq |x(0)|^\beta$ and, again, by above $D(x) \geq N|x(0)|^\beta$.

If $x(0) \geq 0$ we derive

$$
\begin{aligned}
Nx(0)^\beta &\leq N(x(0) \vee x(t_1))^\beta \\
&\leq N((x(0) \vee x(t_1)) - x(t_2))^\beta + Nx(t_2)^\beta \\
&\leq ((x(0) \vee x(t_1)) - x(t_2)) \Big(\frac{N}{(x(t_1) - x(t_2))^{1-\beta}} + \frac{1}{2} \Big) \\
&\leq D(x) \Big(\frac{N}{(2N)^{1-\beta} z_0^{\beta(1-\beta)}} + \frac{1}{2} \Big) \\
&= D(x). \hspace{4cm} \square
\end{aligned}
$$

Now we go back to our equation (3.6). Define the "bad" set

$$G := \{ x \in \mathcal{C} : x(-1) \geq z_0 \ \text{and} \ x(0) \geq x(-1) + 2Nx(-1)^\beta \}.$$

Let us prove that if the process X starts with any initial condition from G, then it tends to infinity with positive probability.

Put $\tau := \inf\{t \geq 0 : X(t) = 1\}$ and $W_* := \inf_{s \in [0,1]} W(s)$. Note that if $X_0 \in G$, then, thanks to Lemma 3.6, we have $D(X_s) \geq N|X(s)|^\beta$ for any $s \in [0,1]$. Hence, it follows from the definition of f that

$$\mathsf{P}_x\big(f(X_{s \wedge \tau}) = 5NX(s \wedge \tau)^\beta \ \text{for every} \ s \in [0,1]\big) = 1,$$

for any $x \in G$. This and (3.6) imply that if $X_0 \in G$, then for any $s \in [0,1]$ we have

$$X(s \wedge \tau) \geq X(0) + W(s \wedge \tau) \geq X(0) + W_*. \tag{3.7}$$

Therefore on the set $\{W_* \geq -X(0) + 1\}$ we have $\tau \geq 1$. We employ this observation together with (3.7) to deduce for any $x \in G$

$$
\begin{aligned}
\mathsf{P}_x(X_1 \in G) &= \mathsf{P}_x\big(X(1) \geq x(0) + 2Nx(0)^\beta\big) \\
&\geq \mathsf{P}_x\big(X(1) \geq x(0) + 2Nx(0)^\beta, \ W_* \geq -x(0)^\beta/2\big) \\
&\geq \mathsf{P}_x\big(5N(x(0) + W_*)^\beta + W_* \geq 2Nx(0)^\beta, \ W_* \geq -x(0)^\beta/2\big) \\
&= \mathsf{P}(W_* \geq -x(0)^\beta/2) \\
&\geq 1 - 2\exp\{-x(0)^{2\beta}/8\},
\end{aligned}
$$

where in the fourth transition we used the fact that $x(0) \geq 1$ and hence

$$5N(x(0) + W_*)^\beta + W_* \geq 5N(x(0) - x(0)^\beta/2)^\beta - x(0)^\beta/2 \geq x(0)^\beta(5N-1)/2 \geq 2Nx(0)^\beta,$$

whenever $W_* \geq -x(0)^\beta/2$.

We apply the Markov property of X to get for any $x \in G$

$$\mathsf{P}_x(X_n \in G \ \text{for all} \ n \in \mathbb{Z}_+) \geq 1 - 2\sum_{n=0}^{\infty} \exp\{-y_n^{2\beta}/8\}, \tag{3.8}$$

where we defined recursively $y_0 := x(0)$ and $y_n := y_{n-1} + 2Ny_{n-1}^\beta$. Since $\beta > 0$ and $y_n \geq x(0) + n$, we see that there exists large enough $Z_0 \geq 0$ such that the right–hand side of (3.8) is positive whenever $x(0) \geq Z_0$. Thus for any $x \in G' := G \cap \{x(0) \geq Z_0\}$ we have $\mathsf{P}_x(\lim_{n \to \infty} X(n) = +\infty) > 0$. This implies by [23, Theorem 3a and 3c] that X does not have an invariant measure. $\hspace{2cm} \square$

Example 3.7. Finally, let us mention that the condition that the diffusion g depends on x only through $x(0)$ in Assumption **A4** also cannot be dropped. Indeed, consider again SFDE (3.1) with $\gamma = 0$, $g(x) = \widetilde{g}(x(-r))$, $r > 0$, $x \in \mathcal{C}$ and $\widetilde{g} \colon \mathbb{R} \to \mathbb{R}_+$ is a bounded increasing and strictly positive function. This equation satisfies Assumptions **A1** and **A2** and its drift is Lipschitz. Nevertheless, as shown in [24] this equation converges to its invariant measure only weakly and not in total variation. Hence without this additional assumption, one cannot replace convergence in the Wasserstein metric in (2.5) and (2.7) by the convergence in total variation.

4. Proofs of the Theorems 2.1 and 2.2

Till the end of this section without loss of generality and to simplify the notation we assume that the memory $r = 1$. In Section 4.1 we establish general lemmas that are useful for the proofs of our main results. In Sections 4.2 and 4.3 we prove Theorems 2.1 and 2.2.

4.1. General tools

First let us verify that the strong solution to SFDE (2.1) has indeed a Markov property. Whilst this statement is well–known for the case of Lipschitz drift and diffusion, we were not able to find in the literature the proof of the Markov property of SFDE in the case of the one–sided Lipschitz drift. Thus we provide it here for the sake of completeness.

Proposition 4.1. *Suppose that Assumption **A1** holds. Then the unique strong solution to (2.1) $X = (X_t)_{t \geq 0}$ is a strong Markov process with the state space $(\mathcal{C}, \mathcal{B}(\mathcal{C}))$.*

Proof. We establish the Markov property using the standard technique (see, e.g., [20, proof of Proposition 3.4]). The authors are grateful to Alexei Kulik for communicating the main idea of the proof.

Fix $t \geq s \geq 0$ and a bounded measurable function $f \colon \mathcal{C} \to \mathbb{R}$. Introduce the filtration $\mathcal{F}_r := \sigma(W(u), 0 \leq u \leq r) \vee \mathcal{N}$, where $r \geq 0$ and $\mathcal{N}$ denotes the collection of null-sets in $\mathcal{F}$. Similarly, put $\mathcal{G}_{r,s} := \sigma(W(u) - W(s), s \leq u \leq r) \vee \mathcal{N}$, $r \geq s$. Our goal is to show that

$$\mathsf{E}(f(X_t)|\mathcal{F}_s) = \mathsf{E}(f(X_t)|X_s). \tag{4.1}$$

Since the function f is arbitrary, (4.1) would imply the Markov property for X.

To establish (4.1) consider the equation

$$X^{(s,x)}(r) = x(0) + \int_s^r b(X_u^{(s,x)})du + \int_s^r \sigma(X_u^{(s,x)})dW(u), \quad r \geq s,\, x \in \mathcal{C}. \tag{4.2}$$

It follows from [21, Theorem 2.3] and a simple shift argument that for each fixed $x \in \mathcal{C}$, equation (4.2) has a unique strong solution and $X_t^{(s,x)}$ is a $\mathcal{G}_{t,s}|\mathcal{B}(\mathcal{C})$ measurable function.

Introduce now a function $\Phi \colon \mathcal{C} \times \Omega \to \mathcal{C}$, $(x, \omega) \mapsto X_t^{(s,x)}(\omega)$. By above, for any fixed $x \in \mathcal{C}$ the function $\Phi(x, \cdot)$ is $\mathcal{G}_{t,s}|\mathcal{B}(\mathcal{C})$–measurable. By [10, Proposition 5.4], there exists $C > 0$ such that

$$\mathsf{E}\|X_t^{(s,x)} - X_t^{(s,y)}\|^4 \leq e^{C(1+t)^2}\|x - y\|^4, \quad x, y \in \mathcal{C}. \tag{4.3}$$

Therefore $\Phi(x, \cdot)$ is continuous in probability with respect to x. Since the space $\mathcal{C}$ is Polish, [6, Theorem 3.1] implies that Φ has a modification $\widetilde{\Phi}$ that is $(\mathcal{B}(\mathcal{C}) \otimes \mathcal{G}_{t,s})|\mathcal{B}(\mathcal{C})$–measurable.

Strong uniqueness of solutions to (4.2) ([21, Theorems 2.2 and 2.3]) yields that

$$X_t(\omega) = \widetilde{\Phi}(X_s, \omega) \quad \text{a.s.,}$$

where we also used the fact that X_s is $\mathcal{F}_s$ measurable and the σ-algebras $\mathcal{F}_s$ and $\mathcal{G}_{t,s}$ are independent.

Now let us prove (4.1). It follows from the measurability properties of $\widetilde{\Phi}$ established above, that $f(\widetilde{\Phi}(X_s, \cdot))$ is $\sigma(X_s, \mathcal{G}_{t,s})$–measurable. Using again the independence of $\mathcal{F}_s$ and $\mathcal{G}_{t,s}$ and a standard approximation argument (see, e.g., [19, Theorem 7.1.2]), we derive

$$\mathsf{E}(f(X_t)|\mathcal{F}_s) = \mathsf{E}(f(\widetilde{\Phi}(X_s, \cdot))|\mathcal{F}_s) = \mathsf{E}f(\widetilde{\Phi}(x, \cdot))|_{x=X_s}.$$

Similarly,

$$\mathsf{E}(f(X_t)|X_s) = \mathsf{E}(f(\widetilde{\Phi}(X_s, \cdot))|X_s) = \mathsf{E}f(\widetilde{\Phi}(x, \cdot))|_{x=X_s}$$

and therefore identity (4.1) holds.

To establish the strong Markov property we employ again bound (4.3). This inequality and the Portmanteau theorem imply that the process X is Feller. Since it has also continuous trajectories, it is strongly Markov [22, Theorem 3.3.1]. $\qquad\square$

As mentioned above, our approach for establishing ergodicity is based on Lyapunov functions. The propositions below state that if one is able to construct a "good" Lyapunov function, then SFDE (2.1) possesses all the required ergodic properties. These propositions essentially follow from the corresponding results in [3] and [10].

Recall that by P_t we denoted the Markov semigroup associated with the strong solution to (2.1).

Proposition 4.2. *Suppose that Assumption **A1** holds. Suppose that there exists a measurable function $V \colon \mathcal{C} \to \mathbb{R}_+$ such that $\lim_{\|x\|\to\infty} V(x) = +\infty$ and*

$$\mathsf{E}_x V(X_1) \leq V(x) - \Psi(V(x)) + C, \quad x \in \mathcal{C}, \tag{4.4}$$

where $\Psi \colon \mathbb{R}_+ \to (0, +\infty)$ is a differentiable concave function increasing to infinity. Then SFDE (2.1) has a unique invariant measure π. Furthermore, for any $\rho > 0$, $\varepsilon > 0$ there exist constants $C_1 > 0$, $C_2 > 0$ such that

$$W_{d_\rho}(P_t(x, \cdot), \pi) \leq \frac{C_1(1 + V(x))}{\Psi(H_\Psi^{-1}(C_2 t))^{1-\varepsilon}}, \quad t \geq 0, \, x \in \mathcal{C}. \tag{4.5}$$

Here $H_\Psi(t) := \int_1^t \frac{1}{\Psi(s)}\, ds$, $t \geq 0$, and H_Ψ^{-1} is the inverse function.

Proof. Fix $\rho > 0$. It follows from [10, Sections 5.1 and 5.2] that for some $n_0 \in \mathbb{N}$, $\delta \in (0, \rho)$ we have

$$W_{d_\delta}(P_{n_0}(x, \cdot), P_{n_0}(y, \cdot)) \leq d_\delta(x, y), \quad x, y \in \mathcal{C}. \tag{4.6}$$

and for any $N > 0$ there exists $\gamma \in (0, 1)$ such that

$$W_{d_\delta}(P_{n_0}(x, \cdot), P_{n_0}(y, \cdot)) \leq \gamma d_\delta(x, y), \quad x, y \in \mathcal{C}, \, \|x\| \leq N, \, \|y\| \leq N. \tag{4.7}$$

Consider now an auxiliary skeleton Markov chain with the state space $(\mathcal{C}, d_\delta)$ and transition kernel

$$\widetilde{P}(x, A) := P_{n_0}(x, A), \quad x \in \mathcal{C}, \, A \in \mathcal{B}(\mathcal{C}).$$

Let us check that this chain satisfies all the conditions of [3, Theorem 2.1]. By iterating (4.4) n_0 times, we see that

$$\int_\mathcal{C} V(y)\widetilde{P}(x, dy) = \mathsf{E}_x V(X_{n_0}) \leq V(x) - \Psi(V(x)) + n_0 C, \quad x \in \mathcal{C}.$$

Therefore the first condition of [3, Theorem 2.1] holds. As explained above, the space $(\mathcal{C}, d_\delta)$ is a complete separable metric space, therefore the second condition is also met. It follows from estimates (4.6), (4.7), and our assumption $\lim_{\|x\|\to\infty} V(x) = +\infty$ that the third and the fourth conditions of [3, Theorem 2.1] are also satisfied.

Thus, all conditions of [3, Theorem 2.1] are met. Hence the skeleton chain has a unique invariant measure π, and there exist constants $C_1 > 0$, $C_2 > 0$ such that

$$W_{d_\rho}(\widetilde{P}_n(x, \cdot), \pi) \leq W_{d_\delta}(\widetilde{P}_n(x, \cdot), \pi) \leq \frac{C_1(1 + V(x))}{\Psi(H_\Psi^{-1}(C_2 n))^{1-\varepsilon}}, \quad n \in \mathbb{Z}_+, \, x \in \mathcal{C},$$

where we also used the fact that $d_\rho \leq d_\delta$. Now by a standard argument (see, e.g., [3, p. 550]), we see that the measure π is also a unique invariant measure for our original Markov kernel P_t and that bound (4.5) holds. $\qquad\square$

Proposition 4.3. *Assume that all conditions of Proposition 4.2 are met. Suppose additionally that Assumption **A4** is satisfied. Then the convergence in the Wasserstein metric in* (4.5) *can be replaced by convergence in total variation metric.*

Proof. We begin by observing that, thanks to the additional Assumption **A4**, the Markov semigroup P_t satisfies the Harnack inequality. Namely, it follows from [27, Theorem 4.1] (see also [5, Theorem 1.1]) that for any $t > 1$ and large enough $p > p_0$, there exists $C = C(p)$ such that

$$\left(P_t(x, A)\right)^p \leq P_t(y, A) e^{C(1 + \|x - y\|^2)}, \quad x, y \in \mathcal{C}, \, A \in \mathcal{B}(\mathcal{C}).$$

Therefore for any $x, y \in \mathcal{C}$, $A \in \mathcal{B}(\mathcal{C})$ we have

$$P_t(x, A) - P_t(y, A) \leq P_t(x, A) - \left(P_t(x, A)\right)^p (e^{-C(1 + \|x - y\|^2)} \wedge p^{-1})$$
$$\leq 1 - (e^{-C(1 + \|x - y\|^2)} \wedge p^{-1}).$$

Thus, we have the following bound on the total variation distance.

$$d_{TV}(P_t(x, \cdot), P_t(y, \cdot)) \leq 1 - (e^{-C(1 + \|x - y\|^2)} \wedge p^{-1}), \quad x, y \in \mathcal{C}.$$

Now similar to the proof of Proposition 4.2 we fix $t = 2$ and consider the skeleton Markov chain with the transition kernel

$$\widetilde{P}(x, A) := P_2(x, A), \quad x \in \mathcal{C}, \, A \in \mathcal{B}(\mathcal{C}).$$

It follows from the above that $\widetilde{P}$ satisfies all the assumptions of [15, Theorem 1.15 and Theorem 1.13]. Note that we do not have to check that the skeleton Markov chain is irreducible or aperiodic, see also the related discussion in [9, Remark 3.3].

Thus, if π denotes the invariant measure of P (its existence and uniqueness was already established in Proposition 4.2), then by [15, Theorem 1.15 and Theorem 1.13] we have

$$d_{TV}(P_{2n}(x, \cdot), \pi) = d_{TV}(\widetilde{P}_n(x, \cdot), \pi) \leq \frac{C_1(1 + V(x))}{\Psi(H_\Psi^{-1}(C_2 n))^{1-\varepsilon}}, \quad n \in \mathbb{Z}_+, \, x \in \mathcal{C}.$$

Therefore if $t = 2n + s$, where $n \in \mathbb{Z}_+$ and $s \in [0, 2]$, then for any $x \in \mathcal{C}$ we derive

$$d_{TV}(P_t(x, \cdot), \pi) = d_{TV}(P_{2n+s}(x, \cdot), P_s\pi) \leq d_{TV}(P_{2n}(x, \cdot), \pi) \leq \frac{C_1(1 + V(x))}{\Psi(H_\Psi^{-1}(C_2 t))^{1-\varepsilon}},$$

where we made use of the nonexpanding property of the total variation metric. This completes the proof of the proposition. $\qquad\square$

The following two lemmas describe the behaviour of $D(X_t)$. These lemmas provide very important estimates that will be used in the sequel.

Lemma 4.4. *Suppose that Assumption **A1** holds. Assume that the drift f satisfies the growth condition* (2.2) *with $\beta \in [0, 1)$ and the diffusion g is globally bounded. Then there exists a constant $C > 0$ such that for any $\lambda \geq 0$ we have*

$$\mathsf{E}_x e^{\lambda D(X_1)} \leq e^{C\lambda(|x(0)|^\beta + D(x)^\beta + \lambda + 1)}, \quad x \in \mathcal{C}. \tag{4.8}$$

Moreover, there exist $C > 0$, $\lambda_0 > 0$ such that

$$\mathsf{E}_x e^{\lambda_0 D(X_1)^2} \leq e^{C(|x(0)|^{2\beta} + D(x)^{2\beta} + 1)}, \quad x \in \mathcal{C}. \tag{4.9}$$

Finally, there exist constants C_1, $C_2 > 0$ such that for any $z > 0$ we have

$$\mathsf{P}_x(D(X_1) \geq z) \leq C_1 e^{C_1 |x(0)|^{2\beta} + C_1 D(x)^{2\beta} - C_2 z^2}, \quad x \in \mathcal{C}. \tag{4.10}$$

Proof. We begin by observing that for any $x \in \mathcal{C}$

$$D(X_1) \leq 2 \sup_{0 \leq t \leq 1} |X(t) - X(0)| \leq 2 \int_0^1 |f(X_s)| \, ds + 2 \sup_{0 \leq t \leq 1} |M(t)|, \tag{4.11}$$

where we denoted $M(t) := \int_0^t g(X_s) \, dW(s)$. We make use of the growth condition (2.2) and the estimate $D(X_s) \leq D(x) + D(X_1)$, valid for all $s \in [0, 1]$, to derive

$$\int_0^1 |f(X_s)| \, ds \leq C \int_0^1 (\|X_s\|^\beta + 1) \, ds \leq C(|x(0)|^\beta + D(x)^\beta + D(X_1)^\beta + 1)$$

$$\leq \frac{1}{4} D(X_1) + C|x(0)|^\beta + CD(x)^\beta + C, \tag{4.12}$$

where we also used the fact that $\beta < 1$ and hence for some $C_\beta > 0$ one has $Cz^\beta \leq z/4 + C_\beta$ for all $z \geq 0$. Substituting (4.12) into (4.11), we get

$$D(X_1) \leq C|x(0)|^\beta + CD(x)^\beta + C \sup_{0 \leq t \leq 1} |M(t)| + C. \tag{4.13}$$

To estimate the exponential moments of $\sup_{0 \leq t \leq 1} |M(t)|$ we use the Dambis–Dubins–Schwarz theorem and the global boundedness of g. It follows that

$$M(t) = (B^1(\tau_1), \dots, B^d(\tau_d)), \quad t \in [0, 1],$$

where $B^1, B^2, \dots, B^d$ are (possibly dependent) one-dimensional Brownian motions and

$$\tau_i = \tau_i(t) := \int_0^t \sum_{j=1}^m (g^{ij}(X_s))^2 \, ds \leq C_g, \quad t \in [0, 1], \, i = 1, 2, \dots, d.$$

and the constant C_g does not depend on x. Thus,

$$\sup_{t \in [0,1]} |M(t)| \leq \sum_{i=1}^d \sup_{t \in [0, C_g]} |B^i(t)|.$$

Therefore, we apply the Cauchy–Schwarz inequality to get for any $\lambda \geq 0$, $x \in \mathcal{C}$

$$\mathsf{E}_x \exp\{\lambda \sup_{t \in [0,1]} |M(t)|\} \leq \mathsf{E} \exp\{\lambda \sum_{i=1}^d \sup_{t \in [0, C_g]} |B^i(t)|\} \leq \mathsf{E} \exp\{d\lambda \sup_{t \in [0, C_g]} |B(t)|\}$$

$$= \mathsf{E} \exp\{C\lambda |B(1)|\} \leq \exp\{C(\lambda + \lambda^2)\},$$

where by B we denoted a standard Brownian motion. This together with (4.13) implies (4.8).

Arguing as above, we see that there exists constants $C > 0$, $\lambda_0 > 0$ such that for any $x \in \mathcal{C}$ we have

$$\mathsf{E}_x \exp\{\lambda_0 \sup_{t \in [0,1]} M(t)^2\} < C.$$

This together with (4.13) implies (4.9).

Estimate (4.10) follows directly from (4.9) and the Chebyshev inequality. $\qquad \square$

Lemma 4.5. *Suppose that the assumptions of Lemma 4.4 hold. Let $\psi\colon \mathbb{R}_+ \to \mathbb{R}_+$ be an increasing continuous concave function such that $\psi(t) \leq C_\psi(t+1)$ for some $C_\psi \geq 1$ and any $t \in \mathbb{R}_+$. Then there exist constants $C_1 > 0$, $C_2 > 0$ such that for any $t \in [0,1]$ and $x \in \mathcal{C}$ with $D(x) \leq \psi(|x(0)|)/(4C_\psi)$ we have*

$$\mathsf{P}_x(D(X_t) \geq \psi(|X(t)|)) \leq C_1 e^{C_1|x(0)|^{2\beta} + C_1 D(x)^{2\beta} - C_2 \psi(|x(0)|)^2}. \tag{4.14}$$

Proof. First, let us note that for any $t \in [0,1]$ we have $|X(t)| \geq |X(0)| - D(X_t)$. Hence

$$\psi(|X(t)|) \geq \psi(|X(0)|) - \psi(D(X_t)).$$

Therefore, using the condition $\psi(s) \leq C_\psi(s+1)$, $s \in \mathbb{R}_+$, we derive for any $t \in [0,1]$, $x \in \mathcal{C}$ with $D(x) \leq \psi(|x(0)|)/(4C_\psi)$

$$\begin{aligned}
\mathsf{P}_x\big(D(X_t) \geq \psi(|X(t)|)\big) &\leq \mathsf{P}_x\big((C_\psi + 1)D(X_t) + C_\psi \geq \psi(|X(0)|)\big) \\
&\leq \mathsf{P}_x\big(D(X_t) \geq \frac{\psi(|X(0)|)}{2C_\psi} - 1\big) \\
&\leq \mathsf{P}_x\big(D(X_1) \geq \frac{\psi(|X(0)|)}{4C_\psi} - 1\big),
\end{aligned}$$

where in the last inequality we used the assumption $D(x) \leq \psi(|x(0)|)/(4C_\psi)$ and the estimate $D(X_t) \leq D(X_1) + D(x)$. Now the application of estimate (4.10) yields (4.14). $\square$

4.2. Proof of Theorem 2.1

To prove Theorem 2.1 we use the following Lyapunov function:

$$V(x) := \exp\big\{\lambda|x(0)| + \big(D(x) - \gamma|x(0)|^\beta\big)_+\big\}, \quad x \in \mathcal{C}, \tag{4.15}$$

where the parameters $\lambda > 0$, $\gamma > 0$ are to be set later. To avoid technicalities we assume that the function κ from Assumption **A2** is increasing and concave. Clearly, this is not a restriction at all: if Assumption **A2** is satisfied, then there exists an increasing concave function $\widetilde{\kappa}$ such that **A2** is also satisfied with $\widetilde{\kappa}$ in place of κ. It follows that there exists a constant $C_\kappa \geq 1$ such that $\kappa(t) \leq C_\kappa(t+1)$ for any $t \geq 0$.

First, let us prove that V is a Lyapunov function on the set where $D(x)$ is relatively big compared with $|x(0)|$. The heuristics here is as follows. As we explained in the introduction, it is not typical for the process X to have a large diameter. Thus, $D(X_t)$ will decrease with high probability. This will also cause the decrease of the Lyapunov function V. Formally we have the following lemma.

Lemma 4.6. *Suppose that Assumptions **A1** and **A2** hold. Let V be the Lyapunov function defined in (4.15). Then for any $\lambda > 0$, $\gamma > 0$, $\varepsilon \in (0,1)$ there exist $L > 0$ and $B > 0$ such that*

$$\mathsf{E}_x V(X_1) \leq \varepsilon V(x), \tag{4.16}$$

for any $x \in \mathcal{C}$ such that $D(x) \geq L$ and $D(x) \geq B|x(0)|^\beta$.

Further, for any $\lambda > 0$, $\gamma > 0$, $L > 0$, $R > 0$ we have

$$\sup_{\substack{x \in \mathcal{C}, |x(0)| \leq R \\ D(x) \leq L}} \mathsf{E}_x V(X_1) < \infty. \tag{4.17}$$

Proof. Fix $\lambda > 0$, $\gamma > 0$. Then for any $x \in \mathcal{C}$ we have

$$\begin{aligned}
\mathsf{E}_x V(X_1) &\leq \mathsf{E}_x \exp\{\lambda|X(1)| + D(X_1)\} \\
&\leq \mathsf{E}_x \exp\{\lambda(|X(0)| + D(X_1)) + D(X_1)\} \\
&= e^{\lambda|x(0)|} \mathsf{E}_x e^{(\lambda+1)D(X_1)}, \tag{4.18}
\end{aligned}$$

which immediately yields (4.17).

To establish (4.16), we apply estimate (4.8) to (4.18). We get

$$\mathsf{E}_x V(X_1) \leq \exp\left\{\lambda|x(0)| + C(\lambda+1)(|x(0)|^\beta + D(x)^\beta + \lambda + 2)\right\}$$
$$\leq V(x)\exp\left\{-D(x) + C_1|x(0)|^\beta + C_2 D(x)^\beta + C_3\right\}, \qquad (4.19)$$

where $C_1 := \gamma + C(\lambda+1)$, $C_2 := C(\lambda+1)$, and $C_3 := C(\lambda+1)(\lambda+2)$. Now take large enough $L_0 = L_0(\lambda)$ such that $C_2 z^\beta + C_3 \leq z/3$ whenever $z \geq L_0$. Put $B := 3C_1$. Then it follows from (4.19) that for any $x \in \mathcal{C}$ with $D(x) \geq L_0$ and $D(x) \geq B|x(0)|^\beta$ we get

$$\mathsf{E}_x V(X_1) \leq V(x)e^{-D(x)/3} \leq V(x)e^{-L_0/3},$$

which implies (4.16). $\qquad\square$

The case when the initial diameter is "small" is much more complicated and more precise estimates are needed. In this case $D(X_t)$ stays at the same level, and the decrease of the Lyapunov function V happens due to the decrease of $|X(t)|$. To formalize these ideas we will use the following version of the Gronwall inequality.

Lemma 4.7. *Let $T > 0$, $\theta > 0$, $r \geq 0$. Let $f \colon [0,T] \to \mathbb{R}$ be a continuous function satisfying for any $0 \leq s \leq u \leq T$ the following inequality:*

$$f(u) \leq f(s) - \int_s^u (\theta f(t) - r)dt.$$

Then for any $t \in [0,T]$

$$f(t) \leq e^{-\theta t}f(0) + r/\theta. \qquad (4.20)$$

Proof. Consider the function $g(t) := e^{-\theta t}(f(0) - r/\theta) + r/\theta$, $0 \leq t \leq T$. Clearly, $g(0) = f(0)$ and for any $0 \leq s \leq u \leq T$ we have

$$g(u) = g(s) - \int_s^u (\theta g(t) - r)dt.$$

Hence by [11, Proposition 9.2], we have $f(t) \leq g(t)$ for any $t \in [0,T]$. This implies (4.20). $\qquad\square$

For $B > 0$, $N > 0$ define

$$\mathcal{C}_{B,N} := \{x \in \mathcal{C} : D(x) \leq B|x(0)|^\beta \text{ and } |x(0)| \geq N\}. \qquad (4.21)$$

We start treating this case with the following key lemma.

Lemma 4.8. *Suppose that the assumptions of Lemma 4.6 hold. Then there exist $\nu > 0$, $\rho \in (0,1)$, such that for every $B > 0$ there exists $N > 0$ such that*

$$\mathsf{E}_x e^{\nu|X(1)|} \leq e^{\nu|x(0)|}(1-\rho)$$

for all $x \in \mathcal{C}_{B,N}$.

Proof. Recall the definition of constant M from condition (2.3). With such M in hand, let $\lambda \in (0,1)$, let $\varphi \colon \mathbb{R}^d \to \mathbb{R}_+$ be a smooth function such that $\varphi(y) := \exp(\lambda|y|)$ for $|y| \geq M$. We want to apply a version of Gronwall's lemma to the function $u \mapsto \mathsf{E}_x \varphi(X(u))$.

First of all, we observe that this function is finite. Indeed, thanks to (4.8), we have for any $x \in \mathcal{C}$, $0 \leq u \leq 1$

$$\mathsf{E}_x \varphi(X(u)) \leq C + \mathsf{E}_x e^{|X(u)|} \leq C + e^{|x(0)|}\mathsf{E}_x e^{D(X_1)} < \infty.$$

We make use of assumption **A2** and apply Ito's lemma. We have for $|y| \geq M$

$$\frac{\partial \varphi(y)}{\partial y_i} = \lambda e^{\lambda|y|}|y|^{-1}y_i;$$

$$\frac{\partial^2 \varphi(y)}{\partial y_i \partial y_j} = \lambda^2 e^{\lambda|y|}|y|^{-2}y_iy_j - \lambda e^{\lambda|y|}|y|^{-3}y_iy_j, \quad i \neq j;$$

$$\frac{\partial^2 \varphi(y)}{\partial y_i^2} = \lambda^2 e^{\lambda|y|}|y|^{-2}y_i^2 - \lambda e^{\lambda|y|}|y|^{-3}y_i^2 + \lambda e^{\lambda|y|}|y|^{-1}.$$

Thus, we derive for any $x \in \mathcal{C}$, $0 \leq s \leq u \leq 1$

$$\begin{aligned}
\mathsf{E}_x\varphi(X(u)) \leq{} & \mathsf{E}_x\varphi(X(s)) - \sigma\lambda\int_s^u \mathsf{E}_x\big(\mathbb{1}_{\{D(X_t)\leq\kappa(|X(t)|),|X(t)|>M\}}\,\varphi(X(t))\big)\,dt \\
& + C\lambda\int_s^u \mathsf{E}_x\big(\mathbb{1}_{\{D(X_t)>\kappa(|X(t)|),|X(t)|>M\}}\,(1+\|X_t\|^\beta)e^{\lambda|X(t)|}\big)\,dt \\
& + C\lambda^2\int_s^u \mathsf{E}_x e^{\lambda|X(t)|}\,dt \\
& + C\lambda\int_s^u \mathsf{E}_x\,\mathbb{1}_{\{|X(t)|>M\}}\,e^{\lambda|X(t)|}|X(t)|^{-1}\,dt + C(u-s) + \mathsf{E}_x(M(u)-M(s)) \\
={} & \mathsf{E}_x\varphi(X(s)) - \sigma\lambda I_1 + C\lambda I_2 + C\lambda^2 I_3 \\
& + C\lambda I_4 + C(u-s) + \mathsf{E}_x(M(u)-M(s)),
\end{aligned} \tag{4.22}$$

where we denoted

$$M(u) := \int_0^u \sum_{i,j} \frac{\partial\varphi}{\partial y_i}(X(t))g^{ij}(X_t)dW^j(t).$$

We use the boundedness of g, the definition of φ, and estimate (4.8) to derive

$$\mathsf{E}_x\langle M\rangle_u \leq C + C\int_0^u \mathsf{E}_x e^{2|X(t)|}\,dt \leq C + Ce^{2|x(0)|}\mathsf{E}_x e^{2D(X_1)} < \infty.$$

Thus $(M(t))_{t\geq 0}$ is a martingale and

$$\mathsf{E}_x(M(u)-M(s)) = 0. \tag{4.23}$$

To estimate I_1 we assume that $|x(0)| \geq M_0 := M + \kappa(M)$ in which case $|X(t)| < M$, $t \in [0,1]$ implies $D(X_t) \geq |X(t) - x(0)| > \kappa(M) \geq \kappa(|X(t)|)$. Therefore,

$$\begin{aligned}
\mathsf{E}_x\Big(& \mathbb{1}_{\{D(X_t)\leq\kappa(|X(t)|),|X(t)|\geq M\}}\,\varphi(X(t))\Big) \\
&= \mathsf{E}_x\Big(\mathbb{1}_{\{D(X_t)\leq\kappa(|X(t)|)\}}\,\varphi(X(t))\Big) \\
&= \mathsf{E}_x\varphi(X(t)) - \mathsf{E}_x\big(\mathbb{1}_{\{D(X_t)>\kappa(|X(t)|)\}}\,\varphi(X(t))\big) \\
&\geq \mathsf{E}_x\varphi(X(t)) - \big(\mathsf{P}_x\big(D(X_t)>\kappa(|X(t)|)\big)\big)^{1/2}\big(\mathsf{E}_x[\varphi(X(t))^2]\big)^{1/2}.
\end{aligned} \tag{4.24}$$

We continue this calculation in the following way. Recall that $\kappa(z)/z^\beta \to +\infty$ as $z \to +\infty$. Therefore for any $B > 0$ there exists $N = N(B) > 0$ such that

$$\mathcal{C}_{B,N} \subset \{x \in \mathcal{C} : D(x) \leq \frac{\kappa(|x(0)|)}{4C_\kappa}\},$$

where C_κ was defined in the beginning of Section 4.2. Hence Lemma 4.5 implies that there exist constants C_1, C_2 such that for any $B > 0$ there exists $N = N(B) > 0$ such that for any $t \in [0,1]$

$$\mathsf{P}_x(D(X_t)>\kappa(|X(t)|)) \leq C_1 e^{-C_2\kappa(|x(0)|)^2}, \quad x \in \mathcal{C}_{B,N}. \tag{4.25}$$

By Lemma 4.4, for any $x \in \mathcal{C}$, $t \in [0,1]$ we have

$$
\begin{aligned}
\mathsf{E}_x \varphi(X(t))^2 &\leq C + \mathsf{E}_x e^{2\lambda |X(t)|} \\
&\leq C + e^{2\lambda |x(0)|} \mathsf{E}_x e^{2D(X_1)} \\
&\leq C + e^{2\lambda |x(0)|} e^{C(|x(0)|^\beta + |D(x)|^\beta + 1)}.
\end{aligned}
$$

Using again the fact that $\kappa(z)/z^\beta \to +\infty$ and combining the above estimate with (4.24) and (4.25), we see that there exist constants $C_3 > 0$, $C_4 > 0$ such that for any $B > 0$ there exists $N_1 = N_1(B) > 0$ such that

$$
I_1 \geq \int_s^u \mathsf{E}_x \varphi(X(t)) \, dt - C_3 e^{\lambda |x(0)|} e^{-C_4 \kappa(|x(0)|)^2} (u - s), \quad x \in \mathcal{C}_{B,N_1}. \tag{4.26}
$$

Next, we estimate the integrand in I_2. We estimate this term applying Hölder's inequality to the three factors. The first and third factors are estimated as above. Further, for any $x \in \mathcal{C}$, $t \in [0,1]$

$$
\begin{aligned}
\mathsf{E}_x (1 + \|X_t\|^\beta)^3 &\leq \mathsf{E}_x (1 + |x(0)|^\beta + D(x)^\beta + D(X_1)^\beta)^3 \\
&\leq C\big(1 + |x(0)|^{3\beta} + D(x)^{3\beta} + \mathsf{E}_x[D(X_1)^{3\beta}]\big) \\
&\leq C\big(1 + |x(0)|^{3\beta} + D(x)^{3\beta} + e^{C(|x(0)|^\beta + |D(x)|^\beta + 1)}\big),
\end{aligned}
$$

where in the last inequality we used Lemma 4.4. Thus, there exist constants $C_5 > 0$, $C_6 > 0$ such that for any $B > 0$ there exists $N_2 = N_2(B) > 0$ such that

$$
I_2 \leq C_5 e^{\lambda |x(0)|} e^{-C_6 \kappa(|x(0)|)^2} (u - s), \quad x \in \mathcal{C}_{B,N_2}. \tag{4.27}
$$

For $\widetilde{M} > M$, $x \in \mathcal{C}$ we estimate the integrand in I_4 as follows.

$$
\mathsf{E}_x\big(\mathbb{1}_{\{|X(t)| > M\}} \exp\{\lambda |X(t)|\} |X(t)|^{-1} \big) \leq \widetilde{M}^{-1} \mathsf{E}_x \varphi(X(t)) + \frac{1}{M} e^{\lambda \widetilde{M}}. \tag{4.28}
$$

Combining (4.23), (4.26), (4.27) and (4.28) with (4.22), we see that there exist constants $C_7 > 0$, $C_8 > 0$ such that for any $B > 0$ there exists $N_3 = N_3(B) > 0$ such that for $x \in \mathcal{C}_{B,N_3}$

$$
\begin{aligned}
\mathsf{E}_x \varphi(X(u)) \leq {}& \mathsf{E}_x \varphi(X(s)) - (\sigma\lambda - C\lambda^2 - C\lambda \widetilde{M}^{-1}) \int_s^u \mathsf{E}_x \varphi(X(t)) \, dt \\
&+ C_7 (e^{\lambda |x(0)|} e^{-C_8 \kappa(|x(0)|)^2} + e^{\lambda \widetilde{M}})(u - s).
\end{aligned}
$$

Now, we choose $\widetilde{M}$ large enough and $\lambda > 0$ small enough so that

$$
\theta := \sigma\lambda - C\lambda^2 - C\lambda \widetilde{M}^{-1} > 0.
$$

Clearly θ is independent of B and N_3. Recall also that we have checked in the beginning of the proof that $\mathsf{E}_x \varphi(X(u)) < \infty$ for any $u \in [0,1]$. Thus, by Lemma 4.7 (a version of the Gronwall inequality) we get

$$
\mathsf{E}_x e^{\lambda |X(1)|} \leq \mathsf{E}_x \varphi(X(1)) + e^{\lambda M} \leq e^{\lambda |x(0)|} e^{-\theta} + e^{\lambda |x(0)|} \zeta(|x(0)|), \quad x \in \mathcal{C}_{B,N_3},
$$

where the function $\zeta(z)$ is independent of B and N_3, and tends to 0 as $z \to \infty$. This implies the statement of the lemma. $\qquad \square$

Now we are able to establish a crucial Lyapunov inequality. Recall the definition of $\mathcal{C}_{B,N}$ in (4.21).

Lemma 4.9. *Suppose that Assumptions **A1** and **A2** hold. Then there exist $\lambda > 0$, $\gamma > 0$, $c_1 \in (0,1)$, $c_2 > 0$ such that the function V defined in (4.15) satisfies the following inequality:*

$$
\mathsf{E}_x V(X_1) \leq (1 - c_1) V(x) + c_2, \quad x \in \mathcal{C}. \tag{4.29}
$$

Proof. First we note that for any $\lambda > 0$, $\gamma > 0$, $x \in \mathcal{C}$ we have

$$\mathsf{E}_x V(X_1) \leq \big(\mathsf{E}_x e^{2\lambda|X(1)|}\big)^{1/2} \big(\mathsf{E}_x e^{2(D(X_1)-\gamma|X(1)|^\beta)_+}\big)^{1/2}. \tag{4.30}$$

We take $\nu > 0$ and $\rho \in (0,1)$ as in Lemma 4.8 and put $\lambda := \nu/2$. Then, thanks to Lemma 4.8, for any $B > 0$ there exists $N = N(B) > 0$ such that for any $x \in \mathcal{C}_{B,N}$ we have

$$\big(\mathsf{E}_x e^{2\lambda|X(1)|}\big)^{1/2} \leq e^{\lambda|x(0)|}(1-\rho)^{1/2} \leq V(x)(1-\rho/2). \tag{4.31}$$

Thus, it remains to show that on $\mathcal{C}_{B,N}$ the second factor in the right–hand side of (4.30) is smaller than $(1 + \rho/2)$. Without loss of generality we assume that $N(B)$ is large enough so that $BN^\beta \leq N$ (otherwise we can take larger $N(B)$). Using the inequality $|X(1)| \geq |x(0)| - D(X_1)$, we deduce for any $x \in \mathcal{C}_{B,N}$

$$
\begin{aligned}
\mathsf{E}_x e^{2(D(X_1)-\gamma|X(1)|^\beta)_+} &\leq 1 + \mathsf{E}_x e^{2(D(X_1)-\gamma|X(1)|^\beta)} \\
&\leq 1 + e^{-2\gamma|x(0)|^\beta} \mathsf{E}_x e^{2(D(X_1)+\gamma D(X_1)^\beta)} \\
&\leq 1 + e^{-2(\gamma|x(0)|^\beta - C_\gamma)} \mathsf{E}_x e^{4D(X_1)},
\end{aligned}
$$

where we have also used the fact that for some $C_\gamma > 0$ we have $\gamma z^\beta \leq z + C_\gamma$ for all $z \geq 0$. We continue this estimate, using Lemma 4.4. Recall that on $\mathcal{C}_{B,N}$ we also have $D(x) \leq |x(0)|$, thanks to our additional assumption on $N(B)$. Therefore we derive for any $x \in \mathcal{C}_{B,N}$

$$\mathsf{E}_x e^{2(D(X_1)-\gamma|X(1)|^\beta)_+} \leq 1 + \exp\big\{ -|x(0)|^\beta(2\gamma - C) + C + 2C_\gamma \big\} \tag{4.32}$$

and the constant C depends neither on γ nor on B. Thus, taking $\gamma = \gamma(C)$ large enough, and combining (4.30), (4.31) and (4.32), we see that for any $B > 0$ there exists a constant $N_1(B)$ such that on $\mathcal{C}_{B,N_1}$ we have

$$\mathsf{E}_x V(X_1) \leq V(x)(1 - \rho/4), \quad x \in \mathcal{C}_{B,N_1}. \tag{4.33}$$

Now with such λ and γ in hand we apply Lemma 4.6 with $\varepsilon = 1 - \rho/4$. We get that there exist $B = B(\lambda, \gamma) > 0$, $L = L(\lambda, \gamma) > 0$ such that

$$\mathsf{E}_x V(X_1) \leq V(x)(1 - \rho/4), \quad D(x) \geq (L \vee B|x(0)|^\beta).$$

Together with (4.33) this bound implies that for some $N_2 = N_2(\lambda, \gamma)$, $L_1 = L_1(\lambda, \gamma)$ we have

$$\mathsf{E}_x V(X_1) \leq V(x)(1 - \rho/4), \quad |x(0)| \geq N_2 \text{ or } D(x) \geq L_1.$$

Finally, if $|x(0)| \leq N_2$ and $D(x) \leq L_1$, then by (4.17)

$$\sup_{\substack{|x(0)| \leq N_2 \\ D(x) \leq L_1}} \mathsf{E}_x V(X_1) < \infty.$$

This completes the proof of the lemma. $\qquad\square$

Based on the previous lemmas, we can now complete the proof of Theorem 2.1.

Proof of Theorem 2.1. It follows from Lemma 4.9 that condition (4.4) holds with the function $\Psi(z) := c_1 z$, $z \in \mathbb{R}_+$, where the constant c_1 is defined in (4.29). Hence Theorem 2.1 follows immediately from Propositions 4.2 and 4.3. $\qquad\square$

4.3. Proof of Theorem 2.2

Now we move on to the subgeometric case. We fix till the end of this section the constants α, σ, M and the function κ from Assumption **A3**. As above without loss of generality, we assume that κ is an increasing concave function.

We work with a Lyapunov function

$$V(x) := \exp\big(\lambda_1 |x(0)|^\alpha + \lambda_2 \big(D(x)^2 - \psi(|x(0)|)\big)_+\big), \tag{4.34}$$

where $\lambda_1 > 0$, $\lambda_2 > 0$, and $\psi \colon \mathbb{R}_+ \to \mathbb{R}_+$ is an increasing continuous concave function. We will specify λ_1, λ_2, and the function ψ later. As before, we consider two cases: the "small" diameter case (where we gain from the decrease of the first factor in the Lyapunov function) and the "large" diameter case (where we gain from the decrease of the second factor in the Lyapunov function).

We start with the second case. Recall the definition of λ_0 from Lemma 4.4.

Lemma 4.10. *Suppose that Assumptions **A1** and **A3** hold. Let V be the Lyapunov function defined in (4.34). For any $\lambda_1 > 0$, $\lambda_2 \in (0, \lambda_0/2]$ there exists $R = R(\lambda_1, \lambda_2)$ such that for any $x \in \mathcal{C}$ with $D(x) > (R + \psi(|x(0)|))^{1/2}$ we have*

$$\mathsf{E}_x V(X_1) \le V(x)/2. \tag{4.35}$$

Further, for any $\lambda_1 > 0$, $\lambda_2 \in (0, \lambda_0/2]$, $N > 0$ we have

$$\sup_{x \in \mathcal{C}, \, |x(0)| \le N} \mathsf{E}_x V(X_1) < \infty \tag{4.36}$$

Proof. For any $x \in \mathcal{C}$ we derive

$$\begin{aligned}
\mathsf{E}_x V(X_1) =& \mathsf{E}_x \exp\big(\lambda_1 |X(1)|^\alpha + \lambda_2 \big(D(X_1)^2 - \psi(|X(1)|)\big)_+\big) \\
\le& e^{\lambda_1 |x(0)|^\alpha} \big(\mathsf{E}_x e^{2\lambda_1(D(X_1)+1)}\big)^{1/2} \big(\mathsf{E}_x e^{2\lambda_2 D(X_1)^2}\big)^{1/2} \\
\le& C e^{C(\lambda_1 + \lambda_1^2)} e^{\lambda_1 |x(0)|^\alpha}
\end{aligned} \tag{4.37}$$

where in the second inequality we applied Lemma 4.4 and used the fact that $2\lambda_2 \le \lambda_0$. This immediately implies (4.36). To establish (4.35) we deduce from (4.37) that

$$\mathsf{E}_x V(X_1) \le V(x) e^{-\lambda_2(D(x)^2 - \psi(|x(0)|))_+} C e^{C(\lambda_1 + \lambda_1^2)}, \quad x \in \mathcal{C}.$$

Now we find large $R = R(\lambda_1, \lambda_2)$ such that

$$e^{-\lambda_2 R} C e^{C(\lambda_1 + \lambda_1^2)} < 1/2.$$

By above, if $D(x) > (R + \psi(|x(0)|))^{1/2}$, then $\mathsf{E}_x V(X_1) \le V(x)/2$. $\qquad\square$

Now we move on to the "small" diameter case. Recall the definition of the constant C_κ from the beginning of Section 4.2.

Lemma 4.11. *Suppose that the assumptions of Lemma 4.10 hold. Then there exist $\nu > 0$, $N_0 > 0$, $\rho_1 > 0$, $\rho_2 > 0$ such that for any $x \in \mathcal{C}$ with $|x(0)| \ge N_0$, $D(x) \le \kappa(|x(0)|)/(4C_\kappa)$ we have*

$$\mathsf{E}_x e^{\nu |X(1)|^\alpha} \le e^{\nu |x(0)|^\alpha}(1 - \rho_1 |x(0)|^{2\alpha - 2}) + \rho_2. \tag{4.38}$$

Proof. Recall the definition of M from condition (2.4). Let $\lambda \in (0,1)$. Similar to the proof of Lemma 4.8, we introduce a smooth function $\varphi : \mathbb{R}^d \to \mathbb{R}_+$ such that for $|y| \ge M$ we have $\varphi(y) = \exp(\lambda |y|^\alpha)$ and for $|y| \le M$ we have $\varphi(y) \in [0, e^{\lambda M}]$. Arguing as in the proof of Lemma 4.8 and applying Ito's formula, we have for any $x \in \mathcal{C}$

$$\begin{aligned}
\mathsf{E}_x \varphi(X(1)) \le& \varphi(x) + \alpha\lambda \int_0^1 \mathsf{E}_x \, \mathbb{1}(|X(s)| \ge M) e^{\lambda |X(s)|^\alpha} |X(s)|^{\alpha-2} \langle X(s), f(X_s) \rangle \, ds \\
& + \frac{1}{2}\lambda\alpha \int_0^1 \mathsf{E}_x \, \mathbb{1}(|X(s)| \ge M) e^{\lambda |X(s)|^\alpha} |X(s)|^{\alpha-2} (C_1 \lambda\alpha |X(s)|^\alpha + C_1) ds + C_1 \\
\le& \varphi(x) + \alpha\lambda(I_1 + I_2) + C_1.
\end{aligned}$$

Suppose further that $D(x) \leq \kappa(|x(0)|)/(4C_\kappa)$. First we bound I_1. Using assumption (2.4), we derive

$$I_1 \leq -\sigma \int_0^1 \mathsf{E}_x \, \mathbb{1}(|X(s)| \geq M) e^{\lambda |X(s)|^\alpha} |X(s)|^{2\alpha-2} \, ds$$
$$+ C_2 \int_0^1 \mathsf{E}_x \, \mathbb{1}(D(X_s) > \kappa(|X(s)|), |X(s)| \geq M) e^{\lambda |X(s)|^\alpha} |X(s)|^{\alpha-1} \, ds + C_3.$$

It follows from the Cauchy–Schwarz inequality that for any $s \in [0,1]$

$$\mathsf{E}_x \, \mathbb{1}(D(X_s) > \kappa(|X(s)|), |X(s)| \geq M) e^{\lambda |X(s)|^\alpha} |X(s)|^{\alpha-1}$$
$$\leq M^{\alpha-1} \Big(\mathsf{P}_x(D(X_s) > \kappa(|X(s)|)) \Big)^{1/2} \Big(\mathsf{E}_x \exp(2\lambda |X(s)|^\alpha) \Big)^{1/2}$$
$$\leq C_4 M^{\alpha-1} e^{-C_5 \kappa(|x(0)|)^2} e^{\lambda |x(0)|^\alpha} e^{C_6(1+\lambda+\lambda^2)},$$

where in the last inequality we use the bound $|X(s)|^\alpha \leq |x(0)|^\alpha + D(X_1) + 1$, and estimates (4.8) and (4.14).

To bound I_2 we take a large $\widetilde{M} > M$ to get

$$I_2 \leq \int_0^1 \mathsf{E}_x \, \mathbb{1}(X(s) \geq \widetilde{M}) e^{\lambda |X(s)|^\alpha} |X(s)|^{2\alpha-2} (C_1 \lambda \alpha + C_1 \widetilde{M}^{-\alpha}) ds + C_7 e^{\widetilde{M}}.$$

Note that the constants C_1, C_2, ... C_7 above do not depend on $\lambda \in [0,1]$ or $\widetilde{M}$. Combining all the previous estimates, we deduce

$$\mathsf{E}_x \varphi(X(1)) \leq \varphi(x) - \theta \int_0^1 \mathsf{E}_x \, \mathbb{1}(|X(s)| \geq M) \varphi(X(s)) |X(s)|^{2\alpha-2} \, ds + C_8 e^{\widetilde{M}}$$
$$+ C_9 e^{-C_5 \kappa(|x(0)|)^2} e^{\lambda |x(0)|^\alpha}. \tag{4.39}$$

where $\theta := \alpha\lambda(\sigma - C_1\lambda\alpha - C_1\widetilde{M}^{-\alpha})$.

Recall that C_1 does not depend on $\widetilde{M}$ and λ. Thus, we take λ to be small enough and $\widetilde{M}$ to be large enough so that $\theta > 0$.

Suppose additionally that $|x(0)| \geq 1$. We derive for $s \in [0,1]$

$$\mathsf{E}_x \big[\mathbb{1}(|X(s)| \geq M) \varphi(X(s)) |X(s)|^{2\alpha-2} \big]$$
$$\geq \mathsf{E}_x \big[\mathbb{1}(|X(s)| \geq M) \varphi(X(s)) (1 + |X(s)|)^{2\alpha-2} \big]$$
$$\geq \mathsf{E}_x \big[e^{\lambda |X(s)|^\alpha} (1 + |X(s)|)^{2\alpha-2} \big] - C_{10}$$
$$\geq e^{\lambda |x(0)|^\alpha} |x(0)|^{2\alpha-2} \mathsf{E}_x \big[e^{-D(X_1)-1} (1 + |x(0)|^{-1} + D(X_1)|x(0)|^{-1})^{2\alpha-2} \big] - C_{10}$$
$$\geq e^{\lambda |x(0)|^\alpha} |x(0)|^{2\alpha-2} \mathsf{E}_x \big[e^{-D(X_1)-1} (2 + D(X_1))^{2\alpha-2} \big] - C_{10}. \tag{4.40}$$

We continue the calculations using Jensen's inequality and estimate (4.8) with $\beta = 0$. We get

$$\mathsf{E}_x \big[e^{-D(X_1)-1} (2 + D(X_1))^{2\alpha-2} \big] \geq \Big(\mathsf{E}_x \big[e^{D(X_1)+1} (2 + D(X_1))^{2-2\alpha} \big] \Big)^{-1} \geq C_{11} > 0.$$

Combining this estimate with (4.40) and (4.39), we get

$$\mathsf{E}_x e^{\lambda |X(1)|^\alpha} \leq e^{\lambda |x(0)|^\alpha} - C_{12} e^{\lambda |x(0)|^\alpha} |x(0)|^{2\alpha-2} + C_9 e^{\lambda |x(0)|^\alpha} e^{-C_5 \kappa(|x(0)|)^2} + C_{13}. \tag{4.41}$$

Recall our assumption on growth of κ: $\kappa^2(z)/\log(z) \to \infty$ as $z \to \infty$. Thus, if N_0 is large enough, and $|x(0)| > N_0$, then inequality (4.38) with $\nu = \lambda$ follows directly from (4.41). $\qquad \square$

Now we are ready to give the proof of Theorem 2.2. Recall the definitions of N_0, ν from Lemma 4.11. The next lemma is crucial.

Lemma 4.12. *Suppose that the assumptions of Lemma 4.10 hold. Then there exist $\lambda_1 > 0$, $\lambda_2 > 0$, $c_1 > 0$, $c_2 > 0$, $N > 0$ and a function ψ such that*

$$\mathsf{E}_x V(X_1) \leq V(x)(1 - c_1|x(0)|^{2\alpha-2}) + c_2, \quad x \in \mathcal{C}, |x(0)| \geq N. \qquad (4.42)$$

Proof. First we fix the function ψ such that $\psi(t)/\log(t) \to \infty$ and $\kappa(t)^2/\psi(t) \to \infty$ as $t \to \infty$. Such ψ exists due to our assumptions on the growth of κ. Take C_ψ such that $\psi(t) \leq t + C_\psi$, $t \geq 0$.

We begin in the same way as in Lemma 4.10. We derive for any $x \in \mathcal{C}$

$$\mathsf{E}_x V(X_1) = \mathsf{E}_x \exp\left(\lambda_1 |X(1)|^\alpha + \lambda_2 \left(D(X_1)^2 - \psi(|X(1)|)\right)_+\right)$$

$$\leq \left(\mathsf{E}_x e^{2\lambda_1 |X(1)|^\alpha}\right)^{1/2} \left(\mathsf{E}_x e^{2\lambda_2 (D(X_1)^2 - \psi(|X(1)|))_+}\right)^{1/2}. \qquad (4.43)$$

First we deal with the second factor in the right-hand side of (4.43). We have

$$\mathsf{E}_x e^{2\lambda_2 (D(X_1)^2 - \psi(|X(1)|))_+} \leq 1 + \mathsf{E}_x e^{2\lambda_2 (D(X_1)^2 - \psi(|X(1)|))}$$

$$\leq 1 + e^{-2\lambda_2 \psi(|x(0)|)} \mathsf{E}_x e^{2\lambda_2 [D(X_1)^2 + \psi(D(X_1))]}$$

$$\leq 1 + e^{-2\lambda_2 (\psi(|x(0)|) - C_\psi - 1)} \mathsf{E}_x e^{4\lambda_2 D(X_1)^2}. \qquad (4.44)$$

We choose $\lambda_2 := \lambda_0/4$. Then (4.44) and (4.9) imply

$$\mathsf{E}_x e^{2\lambda_2 (D(X_1)^2 - \psi(|X(1)|))_+} \leq 1 + C e^{-2\lambda_2 \psi(|x(0)|)}, \quad x \in \mathcal{C}. \qquad (4.45)$$

The first factor in the right–hand side of inequality (4.43) has been already estimated in Lemma 4.11. We put $\lambda_1 := \nu/2$. If $|x(0)| \geq N_0$ and $D(x) \leq \kappa(|x(0)|)/(4C_\kappa)$, then (4.38), (4.43), (4.45) imply for such x

$$\mathsf{E}_x V(X_1) \leq e^{\lambda_1 |x(0)|^\alpha}(1 - C_1|x(0)|^{2\alpha-2} + C_2 e^{-2\lambda_2 \psi(|x(0)|)}) + C_3.$$

Since $\psi(z)/\log(z) \to \infty$ as $z \to \infty$, we get for large enough N_1 and all $x \in \mathcal{C}$ with $|x(0)| \geq N_1$ and $D(x) \leq \kappa(|x(0)|)/(4C_\kappa)$

$$\mathsf{E}_x V(X_1) \leq e^{\lambda_1 |x(0)|^\alpha}(1 - C_4|x(0)|^{2\alpha-2}) + C_5 \leq V(x)(1 - C_4|x(0)|^{2\alpha-2}) + C_5. \quad (4.46)$$

Now let us consider the second case: $D(x) \geq \kappa(|x(0)|)/(4C_\kappa)$. We choose R as in Lemma 4.10. Due to our assumptions on the growth of ψ, for large enough N_2 and any $z > N_2$ we have $\kappa(z)/(4C_\kappa) \geq (R + \psi(z))^{1/2}$. Thus, by Lemma 4.10 we have for all $x \in \mathcal{C}$ with $|x(0)| \geq N_2$ and $D(x) \geq \kappa(|x(0)|)/(4C_\kappa)$

$$\mathsf{E}_x V(X_1) \leq V(x)/2.$$

This together with (4.46) proves the lemma. $\qquad \square$

Proof of Theorem 2.2. Let V be a Lyapunov function defined in (4.34) with the parameters specified in Lemma 4.12. Let $\Psi \colon \mathbb{R}_+ \to \mathbb{R}_+$ be a differentiable concave increasing function such that

$$\Psi(z) = \frac{z}{(\log z)^{(2-2\alpha)/\alpha}}$$

for large enough z (more precisely, for $z \geq N$, where $N > 0$ is the same as in Lemma 4.12). It follows from (4.42) that for $x \in \mathcal{C}$ with $|x(0)| \geq N$

$$\mathsf{E}_x V(X_1) \leq V(x) - C_1 V(x)(\log V(x))^{(2\alpha-2)/\alpha} + C_2.$$

This together with (4.36) implies that for any $x \in \mathcal{C}$

$$\mathsf{E}_x V(X_1) \leq V(x) - C_1 \Psi(V(x)) + C_3.$$

Thus, condition (4.4) holds with the function Ψ defined above. Therefore Theorem 2.2 follows now from Propositions 4.2 and 4.3. $\qquad \square$

References

[1] D. Bakry, P. Cattiaux, A. Guillin (2008). Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Analysis* **254** 727–759. MR-2381160

[2] V.I. Bogachev, A.V. Kolesnikov (2012). The Monge–Kantorovich problem: Achievements, connections, and perspectives. *Russian Math. Surveys* **67** 785–890. MR-3058744

[3] O. Butkovsky (2014). Subgeometric rates of convergence of Markov processes in the Wasserstein metric. *Ann. Appl. Probab.* **24** 526–552. MR-3178490

[4] R. Douc, G. Fort, A. Guillin (2009). Subgeometric rates of convergence of f–ergodic strong Markov processes. *Stoch. Process. Appl.* **119** 897–923. MR-2499863

[5] A. Es–Sarhir, M.-K. von Renesse, M. Scheutzow (2009). Harnack inequality for functional SDEs with bounded memory. *Electron. Commun. Probab* **14** 560–565. MR-2570679

[6] D. Gusak, A. Kukush, A. Kulik, Y. Mishura, A. Pilipenko (2010). *Theory of Stochastic Processes: With Applications to Financial Mathematics and Risk Theory*. N.Y.: Springer. MR-2572942

[7] A.A. Gushchin, U. Küchler (2000). On stationary solutions of delay differential equations driven by a Lévy process. *Stoch. Process. Appl.* **88** 195–211. MR-1767844

[8] M. Hairer (2010). *Convergence of Markov processes*. Lecture notes, Univ. Warwick. Available at www.hairer.org/notes/Convergence.pdf.

[9] M. Hairer, J.C. Mattingly (2011). Yet another look at Harris' ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI. Progress in Probability* **63** 109–117. Birkhäuser, Basel. MR-2857021

[10] M. Hairer, J.C. Mattingly, M. Scheutzow (2011). Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations. *Probab. Theory Related Fields.* **149** 223–259. MR-2773030

[11] D. Herzog, J.C. Mattingly (2015). Noise–induced stabilization of planar flows I. *Electron. J. Probab.* **111** 1–43. MR-3418543

[12] K. Ito, M. Nisio (1964). On stationary solutions of a stochastic differential equations. *J. Math. Kyoto. Univ.* **4** 1–75. MR-0177456

[13] S.A. Klokov, A.Yu. Veretennikov (2005). On subexponential mixing rate for Markov processes. *Theory Probab. Appl.* **49** 110–122. MR-2141328

[14] U. Küchler, B. Mensch (1992). Langevins stochastic differential equation extended by a time–delayed term. *Stochastics.* **40** 23–42. MR-1275126

[15] A. Kulik (2015). *Introduction to Ergodic Rates for Markov Chains and Processes With Applications to Limit Theorems*. Potsdam Univ. Press, Potsdam. MR-3617049

[16] A. Kulik, M. Scheutzow (2015). Generalized couplings and convergence of transition probabilities. *Probab. Theory Related Fields*, DOI: 10.1007/s00440-017-0779-8, *to appear*.

[17] S. Meyn, R.L. Tweedie (2009). Markov Chains and Stochastic Stability, 2nd ed. Cambridge Univ. Press, Cambridge. MR-2509253

[18] S.-E. A. Mohammed, M. Scheutzow (1990). Lyapunov exponents and stationary solutions for affine stochastic delay equations. *Stochastics* **29** 259–283. MR-1041039

[19] B. Oksendal (2013). *Stochastic differential equations: an introduction with applications*. Berlin: Springer. MR-0804391

[20] M. Reiss, M. Riedle, O. van Gaans (2006). Delay differential equations driven by Lévy processes: stationarity and Feller properties. *Stoch. Process. Appl.* **116** 1409–1432. MR-2260741

[21] M.-K. von Renesse, M. Scheutzow (2010). Existence and uniqueness of solutions of stochastic functional differential equations. *Random Oper. Stoch. Equ.* **18** 267–284. MR-2718125

[22] D. Revuz, M. Yor (2013). *Continuous martingales and Brownian motion*. Berlin: Springer. MR-1725357

[23] M. Scheutzow (1984). Qualitative behaviour of stochastic delay equations with a bounded memory. *Stochastics.* **12** 41–80. MR-0738934

[24] M. Scheutzow (2005). Exponential growth rates for stochastic delay differential equations. *Stoch. Dyn.* **5** 163–174. MR-2147280

[25] A.Yu. Veretennikov (1988). Bounds for the mixing rate in the theory of stochastic equations. *Theory Probab. Appl.* **32** 273–281. MR-0902757

[26] A.Yu. Veretennikov (2000). On polynomial mixing and convergence rate for stochastic difference and differential equations. *Theory Probab. Appl.* **44** 361–374. MR-1751475

[27] F.-Y. Wang, C. Yuan (2011). Harnack inequalities for functional SDEs with multiplicative noise and applications. *Stoch. Process. Appl.* **121** 2692–2710. MR-2832420

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards

- Free for authors, free for readers

- Quick publication (no backlog)

- Secure publication (LOCKSS[1])

- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]

- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)

- Submit your best articles to EJP-ECP

- Choose EJP-ECP over for-profit journals

---

[1]LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2]EJMS: Electronic Journal Management System http://www.vtex.lt/en/ejms.html
[3]IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4]BS: Bernoulli Society http://www.bernoulli-society.org/
[5]Project Euclid: https://projecteuclid.org/
[6]IMS Open Access Fund: http://www.imstat.org/publications/open.htm