Ergodicity of supercritical SDEs driven by α-stable processes and heavy-tailed sampling

XIAOLONG ZHANG[1,a] and XICHENG ZHANG[2,b]

[1] School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, China.
[a] zhangxl.math@whu.edu.cn
[2] School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, China
& School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081, China
[b] XichengZhang@gmail.com

Let $\alpha \in (0,2)$ and $d \in \mathbb{N}$. We consider the stochastic differential equation (SDE) driven by an α-stable process

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sigma(X_{t-})\mathrm{d}L_t^\alpha, \quad X_0 = x \in \mathbb{R}^d,$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ are locally γ-Hölder continuous with $\gamma \in (0 \vee (1-\alpha)^+, 1]$, and L_t^α is a d-dimensional symmetric rotationally invariant α-stable process. Under certain dissipative and non-degenerate assumptions on b and σ, we show the V-uniformly exponential ergodicity for the semigroup P_t associated with $\{X_t(x), t \geqslant 0\}$. Our proofs are mainly based on the heat kernel estimates recently established in [33] to demonstrate the strong Feller property and irreducibility of P_t. Interestingly, when α tends to zero, the diffusion coefficient σ can increase faster than the drift b. As an application, we put forward a new heavy-tailed sampling scheme.

Keywords: α-stable processes; Ergodicity; Heavy-tailed distribution; Irreducibility; Strong Feller property

1. Introduction

Let $\pi(x) \propto \mathrm{e}^{-U(x)}$ be a probability density function, where $U : \mathbb{R}^d \to \mathbb{R}$ represents a potential function used in the field of physics. The classical Langevin Monte Carlo (LMC) method uses Langevin diffusion processes to approximate the target distribution $\mu(\mathrm{d}x) = \pi(x)\mathrm{d}x$. More precisely, we consider the SDE

$$\mathrm{d}X_t = -\nabla U(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t, \quad X_0 = x, \tag{1}$$

where $(B_t)_{t\geqslant 0}$ is a d-dimensional standard Brownian motion. The solution X describes the trajectory of a particle motion in the potential field $U(x)$. Under certain regularity assumptions on U, it is well known that the SDE above admits a unique solution $X_t = X_t(x)$ and that μ is an invariant probability measure of semigroup P_t generated by $X_t(x)$. Moreover, based on the dissipativity assumption

$$\langle x, \nabla U(x)\rangle \geqslant c_0|x|^2 - c_1,$$

where $c_0, c_1 > 0$, it is well known that the law of X_t exponentially converges to the unique stationary distribution μ in a certain sense as $t \to \infty$ (cf. [5,34]), that is, P_t is exponentially ergodic. In particular, a method is provided for generating samples from μ through the Euler discretization for SDE (1) (cf. [5]). Numerous studies have analyzed the performances of LMC algorithms (see [12,13,20] and references therein).

However, in statistical physics, Langevin diffusion X_t represents the position of a particle at time t which is influenced by a random force, where the random force is usually the sum of many i.i.d. random pulses with *finite* variance. According to the central limit theorem, the sum of the pulses converges to a Gaussian distribution. If the random pulses have *infinite* variance, then the sum of the pulses could converge to an α-stable distribution with $\alpha \in (0,2)$, a typical class of heavy-tailed distributions. In fact, let $X_1, X_2, \cdots$ be i.i.d. random variables with common distribution

$$\mathbb{P}(X_1 > x) = \mathbb{P}(X_1 < -x) = x^{-\alpha}/2, \quad x \geqslant 1,$$

where $\alpha \in (0,2)$. Considering the random walk $S_n = X_1 + \cdots + X_n$, by CLT (see [16, p. 158]), we have

$$S_n/n^{1/\alpha} \Rightarrow Y,$$

where Y is a random variable with α-stable distribution, and $\Rightarrow$ represents weak convergence. Moreover, it is well known that heavy-tailed distributions appear in many stochastic systems. For example, they are used to model inputs to computer and communications networks, they are used to describe many risk processes, they also occur naturally in models of epidemiological spread, and as statistical evidence for their appropriateness in physics, geoscience and economics (see [19] and references therein). Note that the important Pareto distributions in economics and their generalizations do indeed provide a very flexible family of heavy-tailed distributions that may be used to model income distributions as well as a wide variety of other social and economic distributions (see [2]). We refer to [4] and [17] for more details on the classes of heavy-tailed distributions and their applications in the insurance and finance fields.

Now, for $d = 1$ and $U(x) = |x|^\beta$ in (1), it was previously shown in [37] that the diffusion process defined by SDE (1) is exponentially ergodic if and only if $\beta \geqslant 1$. Therefore, the use of a continuous Langevin diffusion (1) to simulate the heavy-tailed distribution is no longer a proper choice. Instead of SDE (1), it is quite natural to consider the SDE

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \mathrm{d}L_t^\alpha, \quad \alpha \in (0,2), \tag{2}$$

where $(L_t^\alpha)_{t\geqslant 0}$ is a d-dimensional rotationally invariant α-stable process with the usual fractional Laplace generator $\Delta^{\frac{\alpha}{2}}$, and

$$b(x) = -\frac{\Gamma(\frac{d-2+\alpha}{2})}{\pi^{\frac{d}{2}} 2^{2-\alpha}\Gamma(\frac{2-\alpha}{2})} \mathrm{e}^{U(x)} \int_{\mathbb{R}^d} \frac{\mathrm{e}^{-U(y)}\nabla U(y)}{|x-y|^{d-(2-\alpha)}}\mathrm{d}y = \mathrm{e}^{U(x)}\Delta^{\frac{\alpha-2}{2}}(\nabla \mathrm{e}^{-U})(x),$$

where $d \geqslant 2$ and $\Delta^{\frac{\alpha-2}{2}}$ is the Riesz potential (cf. [40, page117]). Formally, one sees that

$$\Delta^{\frac{\alpha}{2}}\mathrm{e}^{-U} - \mathrm{div}(b\mathrm{e}^{-U}) \equiv 0,$$

which implies that μ is an invariant measure of SDE (2). Rigorous theoretical results for the exponential ergodicity of the above SDE (2) are given in [23, Theorem 1.1], and therein, some explicit conditions on U are provided. This approach called the fractional Langevin Monte Carlo method was firstly introduced in [35,38] and has been demonstrated to be useful in modern machine learning in terms of both optimization and heavy-tailed sampling (see [22,39,45]).

Motivated by the studies above, we consider the SDE driven by multiplicative α-stable noises

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sigma(X_{t-})\mathrm{d}L_t^\alpha, \quad X_0 = x, \tag{3}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ are two Borel measurable functions. The main contributions of this study are twofold:

(i) Under certain dissipativity, non-degeneracy and locally Hölder regularity assumptions on b and σ, we establish the exponential ergodicity of SDE (3) for all $\alpha \in (0,2)$, which covers the supercritical regime $\alpha \in (0,1)$ in particular.
(ii) To simulate a large class of heavy-tailed distributions, we propose a new ergodicity SDE driven by multiplicative stable processes. Compared with the recent study [23], our conditions on the potential functions are simpler and easier to implement by computer.

Now we make certain assumptions on b and σ:

($\mathbf{H}_{\rm loc}$) For any $m \in \mathbb{N}$, there is a constant $C_m \geqslant 1$ such that for all $|x| \leqslant m$ and $\xi \in \mathbb{R}^d$,

$$C_m^{-1}|\xi| \leqslant |\sigma(x)\xi| \leqslant C_m|\xi|.$$

Moreover, there are $\gamma \in ((1-\alpha)^+, 1]$ and locally bounded measurable functions ℓ_1 and ℓ_2 on $\mathbb{R}^d$ such that for all $x \in \mathbb{R}^d$ and $|h| \leqslant 1$,

$$|b(x+h) - b(x)| \leqslant \ell_1(x)|h|^\gamma, \ \|\sigma(x+h) - \sigma(x)\| \leqslant \ell_2(x)|h|^\gamma. \tag{4}$$

($\mathbf{H}_{\rm glo}^{r,q}$) For a given $r > -\alpha$, there is a sufficiently small $\varepsilon_0 \in (0,1]$ and $c_0, c_1 > 0$ such that for all $|x| > 1$,

$$\langle x, b(x)\rangle + \varepsilon_0\ell_1(x)|x| + q(\|\sigma(x)\| + \varepsilon_0\ell_2(x))^\alpha|x|^{2-\alpha} \leqslant -c_0|x|^{2+r} + c_1, \tag{5}$$

where q is defined by (22) below.

Remark 1.1. Notice that (4) is equivalent to the local γ-Hölder continuity of b and σ. If b and σ are locally Lipschitz, that is, $\gamma = 1$ in (4), then one can take $\varepsilon_0 = 0$ in (5). When $\gamma < 1$, in order to adopt the stopping time to localize the coefficients, we must use the pathwise uniqueness of the strong solutions. Thus we must mollify the coefficients and require $\varepsilon_0 > 0$ in (5) to have some uniform estimates for the approximation coefficients. Examples of b and σ satisfying ($\mathbf{H}_{\rm glo}^{r,q}$) are provided below. In particular, b and σ can achieve a polynomial growth. Interestingly, when $\alpha \in (0,1)$, σ can increase faster than b (see Example 1.4 below).

Under the above assumptions, and according to [9, Theorem 1.1] and by a standard localization technique (see [18, p.216, Section 4.6]), there is a unique weak solution to SDE (3). The semigroup associated with the Markov process $X_t(x)$ is defined by

$$P_t f(x) := \mathbb{E} f(X_t(x)), \quad f \in C_b(\mathbb{R}^d).$$

By Itô's formula, one sees that the generator of P_t is given by

$$\mathscr{L} f(x) := \mathcal{L}_\sigma f(x) + \langle b(x), \nabla f(x)\rangle, \tag{6}$$

with

$$\mathcal{L}_\sigma f(x) = c_{d,\alpha}\int_{\mathbb{R}^d} \frac{f(x+\sigma(x)z) + f(x-\sigma(x)z) - 2f(x)}{|z|^{d+\alpha}}\mathrm{d}z, \quad c_{d,\alpha} := \frac{\alpha 2^{\alpha-2}\Gamma(\frac{d+\alpha}{2})}{\pi^{\frac{d}{2}}\Gamma(\frac{2-\alpha}{2})}. \tag{7}$$

In particular,

$$\Delta^{\frac{\alpha}{2}} f(x) = c_{d,\alpha}\int_{\mathbb{R}^d} \frac{f(x+z) + f(x-z) - 2f(x)}{|z|^{d+\alpha}}\mathrm{d}z, \tag{8}$$

which is the usual fractional Laplacian (see [26, Theorem 1.1]). We call SDE (3) supercritical for $\alpha \in (0,1)$, because from the view point of PDE, the drift term in this case plays a dominant role compared with the diffusion term $\mathcal{L}_\sigma$ (see [9]).

The main result of this paper is as follows:

Theorem 1.2. *Let $r > -\alpha$ and $p \in ((-r) \vee 0, \alpha)$. Define $V_p(x) := 1 + |x|^p$. Suppose that* $(\mathbf{H}_{\rm loc})$ *and* $(\mathbf{H}^{r,q}_{\rm glo})$ *hold. Then there is a unique invariant probability measure μ associated with $(P_t)_{t\geqslant 0}$. Moreover,*

(i) *if $r \geqslant 0$, then P_t is V_p-uniformly exponentially ergodic, i.e., there are constants $\lambda, C_p > 0$ such that for all $t > 0$ and $x \in \mathbb{R}^d$,*

$$\|P_t(x.\cdot) - \mu\|_{{\rm Var};V_p} := \sup_{\|\varphi/V_p\|_\infty \leqslant 1} |P_t\varphi(x) - \mu(\varphi)| \leqslant C_p \mathrm{e}^{-\lambda t} V_p(x);$$

(ii) *if $r > 0$, then P_t is uniformly exponentially ergodic, i.e., there are constants $\lambda, C > 0$ such that for all $t > 0$ and $x \in \mathbb{R}^d$,*

$$\|P_t(x.\cdot) - \mu\|_{\rm Var} := \sup_{\|\varphi\|_\infty \leqslant 1} |P_t\varphi(x) - \mu(\varphi)| \leqslant C \mathrm{e}^{-\lambda t}.$$

Remark 1.3. Let $p \geqslant 1$. Since the usual Wasserstein-p metric $\mathscr{W}_p$ is dominated by the weighted total variation distance $\|\cdot\|_{{\rm Var};V_p}$ (see [41, Theorem 6.15]), in case (i) of Theorem 1.2, when $\alpha \in (1,2)$, we also have the exponential convergence of P_t in the $\mathscr{W}_1$-distance.

We provide two examples below to illustrate our main result.

Example 1.4. The example presented below shows that b and σ can undergo polynomial decay and growth:

$$b(x) = -x(1+|x|^2)^{\beta/2}, \ \ \sigma(x) = (1+|x|^2)^{\gamma/2}\mathbb{I}, \ \ \beta \in (-\alpha, \infty), \ \gamma \in (-\infty, 1 + \tfrac{\beta}{\alpha}).$$

For $|x| > 1$, by Young's inequality, we have

$$\begin{aligned}\langle x, b(x)\rangle + q\|\sigma(x)\|^\alpha |x|^{2-\alpha} &= -|x|^2(1+|x|^2)^{\frac{\beta}{2}} + q(1+|x|^2)^{\frac{\alpha\gamma}{2}}|x|^{2-\alpha}\\ &\leqslant -2^{\frac{\beta}{2}\wedge 0}|x|^{2+\beta} + 2^{\frac{\alpha(\gamma\vee 0)}{2}} q|x|^{2-\alpha+\alpha(\gamma\vee 0)} \leqslant -2^{(\frac{\beta}{2}\wedge 0)-1}|x|^{2+\beta} + c_1.\end{aligned}$$

In particular, when $\alpha > 1$, β can be less than -1 such that b undergoes polynomial decay. Moreover, for a fixed $\beta > 0$, when $\alpha \downarrow 0$, γ can go toward infinity. In other words, σ can grow faster than b. This is not surprising because for the SDE driven by compound Poisson processes (in a certain sense, corresponding to $\alpha = 0$), σ can present arbitrary growth.

Example 1.5. The example below shows that b and σ can increase exponentially:

$$b(x) = -x\mathrm{e}^{|x|}, \ \ \sigma(x) = \mathrm{e}^{|x|/\beta}\mathbb{I}, \ \ \beta \in (\alpha, 2).$$

It is easy to see that for $|h| \leqslant 1$,

$$\begin{aligned}|b(x+h) - b(x)| &\leqslant (1+|x|)\mathrm{e}^{|x|+1}|h| =: \ell_1(x)|h|,\\ \|\sigma(x+h) - \sigma(x)\| &\leqslant \mathrm{e}^{(|x|+1)/\beta}|h| =: \ell_2(x)|h|.\end{aligned}$$

Thus for any $|x| > 1$ and $q > 1$, by Young's inequality, a sufficiently small ε_0 can be chosen such that

$$\begin{aligned}&\langle x, b(x)\rangle + \varepsilon_0 \ell_1(x)|x| + q(\|\sigma(x)\| + \varepsilon_0 \ell_2(x))^\alpha |x|^{2-\alpha}\\ &\quad= -|x|^2 \mathrm{e}^{|x|} + \varepsilon_0(1+|x|)|x|\mathrm{e}^{|x|+1} + q(\mathrm{e}^{|x|/\beta} + \varepsilon_0 \mathrm{e}^{(|x|+1)/\beta})^\alpha |x|^{2-\alpha}\\ &\quad\leqslant -|x|^2\mathrm{e}^{|x|} + \tfrac{1}{4}|x|^2\mathrm{e}^{|x|+1} + 2q\mathrm{e}^{\alpha|x|/\beta}|x|^{2-\alpha} \leqslant -\tfrac{1}{2}|x|^2\mathrm{e}^{|x|} + c_1.\end{aligned}$$

However, when $\alpha \downarrow 0$, the diffusion σ can increase faster than the drift b.

The ergodicity of SDE (3) has been extensively studied for Brownian noise (cf. [3,8,14,42,44]), and dissipativity condition (5) usually takes the form (for example, see [44, (7.1)])

$$2\langle x, b(x)\rangle + \|\sigma(x)\|^2 \leqslant -c_0|x|^2 + c_1.$$

In recent years, there has also been significant interest in the ergodicity of SDEs driven by Lévy processes, see [25,28,43,44] and the references therein. To the best of our knowledge, there are at least three well-developed methods for studying the exponential ergodicity of a Markov process defined through an SDE, namely, Coupling method (cf. [23,25,27–30]), Functional inequality method (see [3,7,42]), and Meyn-Tweedie's minorization condition (cf. [15,21,31,32,34]). Each method has its own merits and can be adapted to different situations. To construct a successful coupling when applying a coupling method, it is usually assumed that the one-sided monotone condition holds (see [29]), that is, for some $R > 0$,

$$\langle x-y, \nabla U(x) - \nabla U(y)\rangle \geqslant -L|x-y|^2 \mathbf{1}_{|x-y|<R} + K|x-y|^2\mathbf{1}_{|x-y|\geqslant R}. \tag{9}$$

For the functional inequality method, it is usually assumed that the potential function U is at least C^2-smooth and that the associated process is reversible, that is, the associated generator is symmetric with respect to its invariant measure μ (see [3,42]). As the obvious advantage of this method, the quantitative convergence rate can be calculated from the parameters. Finally, Meyn-Tweedie's approach is based on the Lyapunov estimates and verifying the strong Feller property and irreducibility for the associated semigroup.

To prove our main theorem, we adopt the Meyn-Tweedie method. It is known that the strong Feller property reflects the regularization effects of the noise and is usually related to the continuity of the transition density of a Markov process with respect to the starting point. There are many ways to establish the strong Feller property, such as, the gradient estimates and Harnack's inequalities, etc.. For the irreducibility, when the driving noise is Brownian motion, there are several methods, i.e., by solving the control problem [32], by using Girsanov's transformation [44], or by applying the classical supported theorem of an SDE [6]. However, when the driven noise is an α-stable noise, the above method for irreducibility is no longer effective. In [44], we directly verified the positivity of the Dirichlet heat kernel using a localization method. Therein, this is restricted to $\alpha \in (1,2)$ because the heat kernel of the operator in this case is comparable with that of the fractional Laplacian $\Delta^{\frac{\alpha}{2}}$. In a recent study [33], when the conditions in $(\mathbf{H}_{\mathrm{loc}})$ hold globally, that is, the constant C_m does not depend on m and $\ell_1(x), \ell_2(x)$ are constants, we obtain the following two-sided estimates for the density of SDE (3), i.e., for any $T > 0$ and a certain $C > 1$, and for all $t \in (0,T]$ and $x, y \in \mathbb{R}^d$,

$$C^{-1} t (t^{\frac{1}{\alpha}} + |\theta_t(x) - y|)^{-d-\alpha} \leqslant p(t,x,y) \leqslant Ct(t^{\frac{1}{\alpha}} + |\theta_t(x)-y|)^{-d-\alpha}, \tag{10}$$

where $\theta_t(x)$ solves the following regularized ODE

$$\dot\theta_t(x) = b * \phi_{t^{1/\alpha}}(\theta_t(x)), \quad \theta_0(x) = x, \tag{11}$$

where $(\phi_\varepsilon)_{\varepsilon>0}$ represents a family of standard mollifiers. The above estimates still allow us to demonstrate the irreducibility of P_t by localization. For the existence of invariant measures, we use the standard Krylov-Bogoliubov's theorem by showing certain Lyapunov's type estimates.

The remainder of this paper is organized as follows. In Section 2, we recall some basic notions regarding the ergodicity of the Markov processes. These notions used are standard in the literature (see [14,36]). Moreover, we also prove a general criterion for the irreducibility of a Markov process in terms of the heat kernel estimates, which is motivated by the studies in [11] and [44]. In Section 3, we prove our main result Theorem 1.2. For the existence of invariant measures, using the classical Krylov-Bogoliubov's method, we show the Lyapunov type estimates for $X_t(x)$. For the uniqueness of the invariant measure as well as the exponential ergodicity, we show the strong Feller property and the irreducibility of P_t based on the heat kernel estimates and suitable localization and smoothing arguments. In Section 4, we present an application of our main results to the heavy-tailed sampling. In the Appendix, the proof of weak convergence for mollifying SDEs is provided for convenience.

We conclude this introduction by mentioning the following conventions: With or without a subscript, letter C denotes an unimportant constant, whose value may change in different places. In addition, we use $A \asymp_C B$ and $A \lesssim_C B$ to denote $C^{-1}B \leqslant A \leqslant CB$ and $A \leqslant CB$ for a certain constant $C \geqslant 1$ respectively. Moreover, for $a, b \in \mathbb{R}$, we denote $a \wedge b := \min\{a, b\}$ and $a \vee b := \max\{a, b\}$.

2. Preliminaries

In this section, we prepare some notions and criterions about the ergodicity and irreducibility of general Markov processes in Euclidean spaces. Fix $N \in \mathbb{N}$. Let $\mathbb{D}$ be the space of all cádlág functions from $[0, \infty)$ to $\mathbb{R}^N$, which is endowed with the Skorokhod topology so that $\mathbb{D}$ becomes a Polish space. Let $X_t(\omega) := \omega_t$ be the coordinate process over $\mathbb{D}$. For $s > 0$, let $\theta_s : \mathbb{D} \to \mathbb{D}$ be the shift operator:

$$\theta_s(\omega)(t) = \omega(t+s), \quad t \geqslant 0.$$

Let $(\mathbb{P}_x)_{x\in\mathbb{R}^N}$ be a family of probability measures over $\mathbb{D}$ so that $\mathscr{M} := \{(X_t)_{t\geqslant 0}, (\mathbb{P}_x)_{x\in\mathbb{R}^N}\}$ forms a family of Markov processes with regard to the natural filtration $(\mathscr{F}_t)_{t\geqslant 0}$. More precisely,

(i) For each $x \in \mathbb{R}^N$, $\mathbb{P}_x(X_0 = x) = 1$.

(ii) For each $A \in \mathscr{B}(\mathbb{D})$, $x \mapsto \mathbb{P}_x(A)$ is Borel measurable.

(iii) For any $s > 0$ and $A \in \mathscr{B}(\mathbb{D})$, it holds that

$$\mathbb{P}_x(A \circ \theta_s | \mathscr{F}_s) = \mathbb{P}_{X_s}(A). \tag{12}$$

Let $\mathcal{B}_b(\mathbb{R}^N)$ be the space of all bounded measurable functions over $\mathbb{R}^N$. For $t \geqslant 0$ and $f \in \mathcal{B}_b(\mathbb{R}^N)$, the semigroup associated with $\mathscr{M}$ is defined by

$$P_t f(x) := \mathbb{E}_x f(X_t), \quad x \in \mathbb{R}^N,$$

where $\mathbb{E}_x$ denotes the expectation with respect to $\mathbb{P}_x$. The following notions are standard and can be found in [14, Chapter 11] or [34, Chapter 16].

(1) A probability measure $\mu \in \mathcal{P}(\mathbb{R}^N)$ is called an invariant probability measure of $(P_t)_{t\geqslant 0}$, if

$$\mu(P_t f) = \mu(f), \quad \forall t \geqslant 0, \ f \in \mathcal{B}_b(\mathbb{R}^N).$$

(2) We call $(P_t)_{t\geqslant 0}$ ergodic, if there exists a unique invariant probability measure μ of $(P_t)_{t\geqslant 0}$, equivalently, (see [14, Theorem 11.9]), for any $x\in\mathbb{R}^N$ and $f\in\mathcal{B}_b(\mathbb{R}^N)$,

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t P_s f(x)\mathrm{d}s=\mu(f).$$

(3) Let $V:\mathbb{R}^N\to[1,\infty)$ be a measurable function. We call $(P_t)_{t\geqslant 0}$ be V-uniformly exponentially ergodic if there are constants $C,\lambda>0$ and an invariant probability measure μ such that

$$\sup_{\|\phi/V\|_\infty\leqslant 1}|P_t\phi(x)-\mu(\phi)|\leqslant C\mathrm{e}^{-\lambda t}V(x),\ \ t\geqslant 0,\ x\in\mathbb{R}^N.$$

In particular, if $V\equiv 1$, then we call $(P_t)_{t\geqslant 0}$ be uniformly exponentially ergodic.

(4) $(P_t)_{t\geqslant 0}$ is said to be strong Feller, if $P_t f\in C_b(\mathbb{R}^N)$ for any $t>0$ and $f\in\mathcal{B}_b(\mathbb{R}^N)$.

(5) $(P_t)_{t\geqslant 0}$ is called irreducible, if for any $t>0$, open ball B and $x\in\mathbb{R}^N$, $P_t\mathbf{1}_B(x)>0$.

The following general result is taken from [21, Theorem 2.5].

Theorem 2.1. *Suppose that $(P_t)_{t\geqslant 0}$ is strong Feller and irreducible. If there are $p,\lambda>0$ and $C_1,C_2>0$ such that*

$$\mathbb{E}_x|X_t|^p\leqslant C_1|x|^p\mathrm{e}^{-\lambda t}+C_2,\ \ t\geqslant 0,\ x\in\mathbb{R}^N,$$

then $(P_t)_{t\geqslant 0}$ is V-uniformly exponentially ergodic with $V(x)=1+|x|^p$. If there are $p,T_0,C_3>0$ such that

$$\mathbb{E}_x|X_t|^p\leqslant C_3,\ \ t\geqslant T_0,\ x\in\mathbb{R}^N,$$

then $(P_t)_{t\geqslant 0}$ is uniformly exponentially ergodic.

Next we present a general criterion for irreducibility in terms of the heat kernel estimates. We believe that it could be used in other cases. Let $\rho:(0,\infty)\times(0,\infty)\to(0,\infty)$ be a continuous function with the properties that for any $t>0$, $r\mapsto\rho(t,r)$ is decreasing on $(0,\infty)$, and for any $0<\delta<T<\infty$,

$$\sup_{t\in[\delta,T]}\sup_{r>0}\rho(t,r)<\infty. \tag{13}$$

We make some assumptions about the Markov process $\mathscr{M}$:

(A$_1$) We suppose that for each $t>0$, with respect to $\mathbb{P}_x$, X_t admits a family of transition probability density function $p(t,x,y)$ in $\mathbb{R}^N$ so that

$$\mathbb{P}_x(X_t\in A)=\int_A p(t,x,y)\mathrm{d}y,\ \ \forall A\in\mathscr{B}(\mathbb{R}^N),\ t>0,\ x\in\mathbb{R}^N. \tag{14}$$

Moreover, $t\mapsto X_t$ is stochastically continuous, equivalently,

$$\mathbb{P}_x(X_t\neq X_{t-})=0,\ \ \forall t>0,x\in\mathbb{R}^N. \tag{15}$$

(A$_2$) We suppose that for any $T>0$, there are constants $\lambda_0,C_0\geqslant 1$ such that for any $t\in(0,T]$ and $x\in\mathbb{R}^N$, and for Lebesgue almost all $y\in\mathbb{R}^N$,

$$C_0^{-1}\rho(t,\lambda_0\Gamma_t(x,y))\leqslant p(t,x,y)\leqslant C_0\rho(t,\Gamma_t(x,y)/\lambda_0), \tag{16}$$

where $\Gamma_t(x,y):(0,\infty)\times\mathbb{R}^N\times\mathbb{R}^N\to(0,\infty)$ is a measurable function, and for each $y\in\mathbb{R}^N$,

$$(t,x)\mapsto\Gamma_t(x,y)\text{ is continuous.}$$

Remark 2.2. In $(\mathbf{A}_2)$, for the standard Gaussian case, one usually takes $\rho(t,r)=t^{-d/2}e^{-r^2/t}$ and $\Gamma_t(x,y)=|x-y|$. For the α-stable case, if $\alpha\in[1,2)$, one takes $\rho(t,r)=t(t^{1/\alpha}+r)^{-d-\alpha}$ and $\Gamma_t(x,y)=|x-y|$; and if $\alpha\in(0,1)$, one takes $\Gamma_t(x,y)=|x-\theta_t(y)|$, where $\theta_t(y)$ is the ODE flow associated with drift b (see (11) above). We would like to emphasize that we do not make any continuity assumption about $y\mapsto p(t,x,y)$. Usually, the regularity of the density requires more regular coefficients in the theory of SDEs.

Let D be an open subset of $\mathbb{R}^N$. The exit time of X_t from D is defined by $\tau_D:=\inf\{t>0,X_t\notin D\}$. We have a general result on irreducibility.

Theorem 2.3. *Let $D_0\Subset D$ be two connected domains of $\mathbb{R}^N$ so that D has Lebesgue-zero measure boundary. Suppose that* $(\mathbf{A}_1)$ *and* $(\mathbf{A}_2)$ *hold, and the above $\Gamma_t(x,y)$ and $\rho(t,r)$ satisfy the following conditions stating that there are $t_0,\delta_0>0$ small enough and functions $\ell_0(t)$, $\ell_1(t)$ defined on $(0,t_0]$ such that for all $t\in(0,t_0]$, $y_0\in D_0$, $x,y\in B_{\delta_0}(y_0)\subset D_0$ and $x'\notin D$,*

$$\Gamma_t(x,y)\leqslant\ell_0(t),\ \Gamma_t(x',y)\geqslant\ell_1(t),\tag{17}$$

and for the C_0,λ_0 in (16), $t\mapsto\rho(t,\ell_1(t)/\lambda_0)$ is increasing on $(0,t_0]$, and

$$C_0^{-1}\rho\big(t,\lambda_0\ell_0(t)\big)\geqslant2C_0\rho\big(t,\ell_1(t)/\lambda_0\big).\tag{18}$$

Then for any $x_0,y_0\in D_0$ and $r_1,r_2<\mathrm{dist}(D_0,D^c)$,

$$\inf_{x\in B_{r_1}(x_0)}\mathbb{P}_x\Big(X_t\in B_{r_2}(y_0);t<\tau_D\Big)\geqslant c_0>0,\tag{19}$$

where c_0 only depends on $\rho,\lambda_0,C_0,t_0,\delta_0$ and r_1,r_2, x_0,y_0,D_0,D.

Proof. Let t_0 and δ_0 be as in the assumptions. We divide the proofs into two steps.

(i) In this step we show that for all $y_0\in D_0$, $\delta\in(0,\frac{\delta_0}{2}]$ and $t\in(0,t_0]$,

$$\inf_{x\in B_{2\delta}(y_0)}\mathbb{P}_x\Big(X_t\in B_\delta(y_0);t<\tau_D\Big)\geqslant C_0\rho\big(t,\ell_1(t)/\lambda_0\big)|B_\delta(y_0)|>0.\tag{20}$$

Fix $s\in(0,t)$ and $n\in\mathbb{N}$. We define a stopping time T_n as follows:

$$T_n=\begin{cases}ks2^{-n} & \text{if }(k-1)s2^{-n}\leqslant\tau_D<ks2^{-n},\ k=1,2,\dots,2^n,\\ \infty & \text{if }\tau_D\geqslant s.\end{cases}$$

Clearly, on $\{\tau_D<s\}$,

$$s\geqslant T_n\downarrow\tau_D\quad\text{as }n\to\infty.$$

For any bounded $A\in\mathscr{B}(D)$, by the Markov property and (14), we have

$$\mathbb{P}_x(X_t\in A;\tau_D<s)=\sum_{k=1}^{2^n}\mathbb{P}_x(X_t\in A;(k-1)s2^{-n}\leqslant\tau_D<ks2^{-n})=\sum_{k=1}^{2^n}\mathbb{P}_x(X_t\in A;T_n=ks2^{-n})$$

$$
\begin{aligned}
&= \sum_{k=1}^{2^n} \mathbb{E}_x \Big(\mathbb{P}_{X_{T_n}}(X_{t-T_n} \in A); T_n = ks2^{-n} \Big) \\
&= \sum_{k=1}^{2^n} \mathbb{E}_x \left(\int_A p(t-T_n, X_{T_n}, y) \mathrm{d}y; T_n = ks2^{-n} \right) \\
&= \mathbb{E}_x \left(\int_A p(t-T_n, X_{T_n}, y) \mathrm{d}y; \tau_D < s \right) \\
&\overset{(16)}{\leqslant} C_0 \mathbb{E}_x \left(\int_A \rho\big(t-T_n, \Gamma_{t-T_n}(X_{T_n}, y)/\lambda_0\big) \mathrm{d}y; \tau_D < s \right).
\end{aligned}
$$

Letting $n \to \infty$ and by (13) and the dominated convergence theorem, we get

$$
\mathbb{P}_x(X_t \in A; \tau_D < s) \leqslant C_0 \mathbb{E}_x \left(\int_A \rho\big(t-\tau_D, \Gamma_{t-\tau_D}(X_{\tau_D}, y)/\lambda_0\big) \mathrm{d}y; \tau_D < s \right).
$$

Below we take $A = B_\delta(y_0)$. For $y \in B_\delta(y_0)$, since $X_{\tau_D} \notin D$ and $r \to \rho(t-\tau_D, r)$ is decreasing, $t \mapsto \rho(t, \ell_1(t)/\lambda_0)$ is increasing, by (17) we have

$$
\rho\big(t-\tau_D, \Gamma_{t-\tau_D}(X_{\tau_D}, y)/\lambda_0\big) \leqslant \rho\big(t-\tau_D, \ell_1(t-\tau_D)/\lambda_0\big) \leqslant \rho\big(t, \ell_1(t)/\lambda_0\big).
$$

Hence, for any $s < t$,

$$
\mathbb{P}_x\Big(X_t \in B_\delta(y_0); \tau_D < s\Big) \leqslant C_0 \int_{B_\delta(y_0)} \rho\big(t, \ell_1(t)/\lambda_0\big) \mathrm{d}y.
$$

By (15) and (16), we have

$$
\begin{aligned}
\mathbb{P}_x(\tau_D = t) &= \mathbb{P}_x(\tau_D = t, X_t \in \partial D) + \mathbb{P}_x(\tau_D = t, X_t \notin \partial D, X_s \in D, s < \tau_D) \\
&\leqslant \mathbb{P}_x(X_t \in \partial D) + \mathbb{P}_x(X_t \neq X_{t-}) = 0.
\end{aligned}
$$

Thus, by the lower bound in (16) and the above estimates, we arrive at

$$
\begin{aligned}
\mathbb{P}_x\Big(X_t \in B_\delta(y_0); t < \tau_D\Big) &= \mathbb{P}_x\Big(X_t \in B_\delta(y_0)\Big) - \mathbb{P}_x\Big(X_t \in B_\delta(y_0); \tau_D < t\Big) \\
&= \int_{B_\delta(y_0)} p(t,x,y) \mathrm{d}y - \lim_{s \uparrow t} \mathbb{P}_x\Big(X_t \in B_\delta(y_0); \tau_D < s\Big) \\
&\geqslant \int_{B_\delta(y_0)} \Big(C_0^{-1} \rho\big(t, \lambda_0 \Gamma_t(x,y)\big) - C_0 \rho\big(t, \ell_1(t)/\lambda_0\big) \Big) \mathrm{d}y.
\end{aligned}
$$

For $x \in B_{2\delta}(y_0), y \in B_\delta(y_0)$, since $r \to \rho(t,r)$ is decreasing, by (17), we obtain

$$
\begin{aligned}
\mathbb{P}_x\Big(X_t \in B_\delta(y_0); t < \tau_D\Big) &\geqslant \Big(C_0^{-1} \rho\big(t, \lambda_0 \ell_0(t)\big) - C_0 \rho\big(t, \ell_1(t)/\lambda_0\big) \Big) |B_\delta(y_0)| \\
&\overset{(18)}{\geqslant} C_0 \rho\big(t, \ell_1(t)/\lambda_0\big) |B_\delta(y_0)|,
\end{aligned}
$$

and then get (20).

(ii) By (14), there is a density $p^D(t,x,y)$ so that for each $A \in \mathscr{B}(D)$,

$$\int_A p^D(t,x,y)\mathrm{d}y = \mathbb{P}_x(X_t \in A; t < \tau_D).$$

Moreover, by the Markov property, we have for each $0 < s < t$,

$$\int_A p^D(t,x,y)\mathrm{d}y = \int_D p^D(s,x,z)\mathrm{d}z \int_A p^D(t-s,z,y)\mathrm{d}y. \tag{21}$$

Indeed, by $s + \tau_D \circ \theta_s = \tau_D$ on $\{s < \tau_D\}$ and (12),

$$\begin{aligned}
\mathbb{P}_x(X_t \in A; t < \tau_D) &= \mathbb{E}_x\Big(\mathbb{P}_x\big(X_t \in A; t < \tau_D|\mathscr{F}_s\big)\Big)\\
&= \mathbb{E}_x\Big(\mathbb{P}_x\big(X_{t-s}\circ\theta_s \in A; t < s + \tau_D\circ\theta_s, s < \tau_D|\mathscr{F}_s\big)\Big)\\
&= \mathbb{E}_x\Big(\mathbb{P}_{X_s}\big(X_{t-s} \in A; t-s < \tau_D\big); s < \tau_D\Big)\\
&= \int_D p^D(s,x,z)\int_A p^D(t-s,z,y)\mathrm{d}y\mathrm{d}z.
\end{aligned}$$

Let $x_0 \in D_0$. Since D_0 is connected, there is a continuous curve $C \subset D_0$ connecting x_0 and y_0. Let

$$\delta := (\mathrm{dist}(D_0, \partial D) \wedge \delta_0)/4.$$

It is easy to see that there exists $n \in \mathbb{N}$ large enough and $\{x_i, i = 0, 1, \cdots, n+1\} \subset C$ such that

$$t \leqslant nt_0, \quad x_i \in B_\delta(x_{i-1}),\ i = 1,\cdots,n+1,\ B_\delta(x_{n+1}) \subset B_{r_2}(y_0).$$

For each $j = 1,\cdots,n+1$, since $B_\delta(x_{j-1}) \subset B_{2\delta}(x_j)$, by what we have proved in Step (i),

$$\begin{aligned}
\inf_{y_{j-1}\in B_\delta(x_{j-1})}\int_{B_\delta(x_j)} p^D(\tfrac{t}{n}, y_{j-1}, y_j)\mathrm{d}y_j &\geqslant \inf_{y_{j-1}\in B_{2\delta}(x_j)}\int_{B_\delta(x_j)} p^D(\tfrac{t}{n}, y_{j-1}, y_j)\mathrm{d}y_j\\
&\geqslant C_0\rho\big(\tfrac{t}{n}, \ell_1(\tfrac{t}{n})/\lambda_0\big)|B_\delta(y_0)| > 0.
\end{aligned}$$

Hence, for all $x \in B_\delta(x_0)$, by (21),

$$\begin{aligned}
\int_{B_{r_2}(y_0)} p^D(t,x,y)\mathrm{d}y &= \int_D p^D(\tfrac{t}{n},x,y_1)\mathrm{d}y_1\cdots\int_D p^D(\tfrac{t}{n},y_{n-1},y_n)\mathrm{d}y_n\int_{B_{r_2}(y_0)} p^D(\tfrac{t}{n},y_n,y)\mathrm{d}y\\
&\geqslant \int_{B_\delta(x_1)} p^D(\tfrac{t}{n},x,y_1)\mathrm{d}y_1\cdots\int_{B_\delta(x_n)} p^D(\tfrac{t}{n},y_{n-1},y_n)\mathrm{d}y_n\int_{B_\delta(x_{n+1})} p^D(\tfrac{t}{n},y_n,y)\mathrm{d}y\\
&\geqslant \Big(C_0\rho\big(\tfrac{t}{n},\ell_1(\tfrac{t}{n})/\lambda_0\big)|B_\delta(y_0)|\Big)^{n+1}.
\end{aligned}$$

Thus,

$$\inf_{x\in B_\delta(x_0)}\mathbb{P}_x\Big(X_t \in B_{r_2}(y_0); t < \tau_D\Big) \geqslant \Big(C_0\rho\big(\tfrac{t}{n},\ell_1(\tfrac{t}{n})/\lambda_0\big)|B_\delta(y_0)|\Big)^{n+1} > 0,$$

which implies (19) by the finitely covering technique. □

3. Proof of Theorem 1.2

In this section, we prove Theorem 1.2 by verifying the conditions in Theorem 2.1. Below we fix

$$r > -\alpha, \quad p \in ((-r) \vee 0, \alpha),$$

and define

$$V_p(x) := (1+|x|^2)^{p/2}, \quad q := \frac{2^{3+2\alpha}\Gamma(\frac{d+\alpha}{2})}{\Gamma(\frac{d}{2})\Gamma(\frac{2-\alpha}{2})}\left(\frac{\alpha}{2-\alpha} + \frac{\alpha}{(\alpha-p)p}\right). \tag{22}$$

First of all, we show that V_p is a Lyapunov function of the operator $\mathscr{L}$.

Lemma 3.1. *Suppose that b, σ are locally bounded and for some $c_0, c_1 > 0$,*

$$q\|\sigma(x)\|^\alpha |x|^{2-\alpha} + \langle x, b(x)\rangle \leqslant -c_0|x|^{2+r} + c_1, \quad |x| > 1. \tag{23}$$

Then there are $\kappa_0, \kappa_1 > 0$ such that for all $x \in \mathbb{R}^d$,

$$\mathscr{L}V_p(x) = \mathcal{L}_\sigma V_p(x) + \langle b(x), \nabla V_p(x)\rangle \leqslant -\kappa_0 V_p(x)^{1+\frac{r}{p}} + \kappa_1. \tag{24}$$

Proof. It suffices to prove (24) for $|x| > 1$. By (7), we make the following decomposition:

$$c_{d,\alpha}^{-1}\mathcal{L}_\sigma V_p(x) = \int_{|z|\leqslant \frac{|x|}{2\|\sigma(x)\|}} \Big[V_p(x+\sigma(x)z) + V_p(x-\sigma(x)z) - 2V_p(x)\Big]\frac{\mathrm{d}z}{|z|^{d+\alpha}}$$
$$+ \int_{|z|> \frac{|x|}{2\|\sigma(x)\|}} \Big[V_p(x+\sigma(x)z) + V_p(x-\sigma(x)z) - 2V_p(x)\Big]\frac{\mathrm{d}z}{|z|^{d+\alpha}} =: I_1 + I_2.$$

For I_1, denoting by $A(x) := (1+|x|^2)\mathbb{I} - (2-p)x \otimes x$, we have

$$\nabla^2 V_p(x) = p(1+|x|^2)^{\frac{p}{2}-2}A(x).$$

Note that for any $\xi \in \mathbb{R}^d$,

$$\langle \xi, A(x)\xi\rangle \leqslant (1+|x|^2)|\xi|^2.$$

By Taylor's expansion, for $|z| \leqslant \frac{|x|}{2\|\sigma(x)\|}$, we have

$$V_p(x+\sigma(x)z) + V_p(x-\sigma(x)z) - 2V_p(x) \leqslant p|\sigma(x)z|^2 \int_0^1\int_{-1}^1 (1+|x+r_1r_2\sigma(x)z|^2)^{\frac{p}{2}-1}\mathrm{d}r_1\mathrm{d}r_2$$
$$\leqslant 2p\|\sigma(x)\|^2|z|^2\left(1+\tfrac{|x|^2}{4}\right)^{\frac{p}{2}-1}.$$

Hence, for $S_d := 2\pi^{\frac{d}{2}}/\Gamma(\frac{d}{2})$,

$$I_1 \leqslant 2p\|\sigma(x)\|^2\left(1+\tfrac{|x|^2}{4}\right)^{\frac{p}{2}-1}\int_{|z|\leqslant \frac{|x|}{2\|\sigma(x)\|}} |z|^{2-d-\alpha}\mathrm{d}z$$

$$
\begin{aligned}
&= 2pS_d\|\sigma(x)\|^2\Big(1+\tfrac{|x|^2}{4}\Big)^{\frac{p}{2}-1}\int_0^{\frac{|x|}{2\|\sigma(x)\|}} r^{1-\alpha}\mathrm{d}r\\
&\leqslant \frac{p2^{\alpha-p+1}S_d}{2-\alpha}\|\sigma(x)\|^{\alpha}|x|^{p-\alpha}.
\end{aligned}
$$

For I_2, noting that

$$
|(1+|x+z|^2)^{\frac{p}{2}}-(1+|x|^2)^{\frac{p}{2}}|\leqslant ||x+z|^2-|x|^2|^{\frac{p}{2}}\leqslant 2^{\frac{p}{2}}|x|^{\frac{p}{2}}|z|^{\frac{p}{2}}+|z|^p,
$$

we have

$$
I_2\leqslant 2\int_{|z|>\frac{|x|}{2\|\sigma(x)\|}}\Big[2^{\frac{p}{2}}|x|^{\frac{p}{2}}|\sigma(x)z|^{\frac{p}{2}}+|\sigma(x)z|^p\Big]\frac{\mathrm{d}z}{|z|^{d+\alpha}}\leqslant \frac{2^{3+\alpha}S_d}{\alpha-p}\|\sigma(x)\|^{\alpha}|x|^{p-\alpha}.
$$

Combining the above calculations and recalling (22), we get for $|x|>1$,

$$
\mathcal{L}_\sigma V_p(x)\leqslant c_{d,\alpha}S_d\left(\frac{p2^{\alpha-p+1}}{2-\alpha}+\frac{2^{3+\alpha}}{\alpha-p}\right)\|\sigma(x)\|^{\alpha}|x|^{p-\alpha}\leqslant p2^{\frac{p}{2}-1}q\|\sigma(x)\|^{\alpha}|x|^{p-\alpha}.
$$

Thus by (23) and $\nabla V_p(x)=px(1+|x|^2)^{\frac{p}{2}-1}$, we have

$$
\begin{aligned}
\mathcal{L}_\sigma V_p(x)+\langle b(x),\nabla V_p(x)\rangle &\leqslant p2^{\frac{p}{2}-1}q\|\sigma(x)\|^{\alpha}|x|^{p-\alpha}+p(1+|x|^2)^{\frac{p}{2}-1}\langle x,b(x)\rangle\\
&\leqslant p(1+|x|^2)^{\frac{p}{2}-1}\Big(q\|\sigma(x)\|^{\alpha}|x|^{2-\alpha}+\langle x,b(x)\rangle\Big)\\
&\leqslant p(1+|x|^2)^{\frac{p}{2}-1}\big(-c_0|x|^{2+r}+c_1\big)\\
&\leqslant -\kappa_0(1+|x|^2)^{\frac{p+r}{2}}+\kappa_1=-\kappa_0V_p(x)^{1+\frac{r}{p}}+\kappa_1,
\end{aligned}
$$

where in the second inequality we use $(1+|x|^2)^{1-\frac{p}{2}}\leqslant 2^{1-\frac{p}{2}}|x|^{2-p}$ with $|x|>1$, and $\kappa_0=pc_0$. $\square$

As a consequence of the above Lyapunov-type estimate, we have (see [44, Lemma 7.1]):

Lemma 3.2. *Let $X_t(x)$ be any weak solution of SDE (3) starting from x. Under the assumptions of Lemma 3.1, there exists a constant $C>0$ such that for all $x\in\mathbb{R}^d$ and $t>0$,*

$$
\left[\mathbb{E}\left(\sup_{s\in[0,t]}V_p(X_s(x))^{\frac{1}{2}}\right)\right]^2+\mathbb{E}\left(\int_0^t V_p(X_s(x))^{1+\frac{r}{p}}\mathrm{d}s\right)\lesssim_C V_p(x)+t, \tag{25}
$$

and for the κ_0 in (24),

$$
\mathbb{E}V_p(X_t(x))\leqslant\begin{cases}\mathrm{e}^{-\kappa_0 t}V_p(x)+C, & r=0,\\ C(\mathrm{e}^{-\kappa_0 t/2}t^{-p/r}+1), & r>0.\end{cases} \tag{26}
$$

Proof. Let $N(t,\mathrm{d}z)$ be the Poisson random measure associated with L_t^{α}, i.e.,

$$
N(t,\Gamma):=\sum_{s\in(0,t]}\mathbf{1}_\Gamma(L_s^{\alpha}-L_{s-}^{\alpha}),\ \ \Gamma\in\mathscr{B}(\mathbb{R}^d).
$$

Let $\widetilde{N}(t,\mathrm{d}z) := N(t,\mathrm{d}z) - 2c_{d,\alpha}t|z|^{-d-\alpha}\mathrm{d}z$ be the compensated Poisson random measure, where $c_{d,\alpha}$ is the same constant as in (7). By Lévy-Itô's decomposition, one can write

$$L_t^\alpha = \int_{|z|<1} z\widetilde{N}(t,\mathrm{d}z) + \int_{|z|\geqslant 1} zN(t,\mathrm{d}z).$$

Thus SDE (3) can be written as

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \int_{|z|<1} \sigma(X_{t-})z\widetilde{N}(\mathrm{d}t,\mathrm{d}z) + \int_{|z|\geqslant 1} \sigma(X_{t-})zN(\mathrm{d}t,\mathrm{d}z). \tag{27}$$

By Itô's formula (see [1, Theorem 4.4.7]) and (24), we have

$$V_p(X_t) = V_p(x) + \int_0^t \Big[\langle \nabla V_p(X_s), b(X_s)\rangle + \mathcal{L}_\sigma V_p(X_s)\Big]\mathrm{d}s + M_t \tag{28}$$

$$\leqslant V_p(x) + \int_0^t \Big[-\kappa_0 V_p(X_s)^{1+\frac{r}{p}} + \kappa_1\Big]\mathrm{d}s + M_t, \tag{29}$$

where M_t is a local cádlág martingale. Let τ_n be a sequence of stopping times localizing M_t, i.e., $M_{t\wedge\tau_n}$ is a martingale and $\tau_n \uparrow \infty$ as $n \to \infty$. Then we have

$$\mathbb{E}V_p(X_{t\wedge\tau_n}) + \kappa_0\mathbb{E}\int_0^{t\wedge\tau_n} V_p(X_s)^{1+\frac{r}{p}}\mathrm{d}s \leqslant V_p(x) + \kappa_1 t.$$

In particular, letting $n \to \infty$ and by Fatou's lemma, we obtain that for all $t \geqslant 0$,

$$\mathbb{E}V_p(X_t) + \kappa_0\mathbb{E}\left(\int_0^t V_p(X_s)^{1+\frac{r}{p}}\mathrm{d}s\right) \leqslant V_p(x) + \kappa_1 t.$$

Moreover, starting from (29), by stochastic Gronwall's inequality ([44, Lemma 3.7]), we also have

$$\left[\mathbb{E}\left(\sup_{s\in[0,t]} V_p(X_s(x))^{\frac{1}{2}}\right)\right]^2 \lesssim_C V_p(x) + t,$$

and starting from (28), as above,

$$\mathbb{E}V_p(X_t) = V_p(x) + \int_0^t \mathbb{E}\Big[\langle \nabla V_p(X_s), b(X_s)\rangle + \mathcal{L}_\sigma V_p(X_s)\Big]\mathrm{d}s.$$

For $r \geqslant 0$, by (24) and Jensen's inequality, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}V_p(X_t) \leqslant -\kappa_0\mathbb{E}\Big[V_p(X_t)^{1+\frac{r}{p}}\Big] + \kappa_1 \leqslant -\kappa_0\Big[\mathbb{E}V_p(X_t)\Big]^{1+\frac{r}{p}} + \kappa_1.$$

Let $f(t) := \mathbb{E}V_p(X_t)$. The above inequality implies

$$f(t)' \leqslant -\kappa_0 f(t)^{1+\frac{r}{p}} + \kappa_1.$$

Since $f(t) \geqslant 1$, by the chain rule we have

$$\begin{aligned}(\mathrm{e}^{\kappa_0 t}f(t))' = (\kappa_0 f(t) + f'(t))\mathrm{e}^{\kappa_0 t} \leqslant (\kappa_0 f(t) - \kappa_0 f(t)^{1+\frac{r}{p}} + \kappa_1)\mathrm{e}^{\kappa_0 t} \leqslant \kappa_1\mathrm{e}^{\kappa_0 t}\\ \Rightarrow f(t) \leqslant \mathrm{e}^{\kappa_0(s-t)}f(s) + \tfrac{\kappa_1}{\kappa_0}(1-\mathrm{e}^{\kappa_0(s-t)}),\quad t > s \geqslant 0,\end{aligned} \tag{30}$$

which in particular gives the first estimate in (26) for $r=0$ by taking $s=0$. If $r>0$, then by the chain rule and Young's inequality we also have

$$
\begin{aligned}
(t^{\frac{2p}{r}}f(t))' &\leqslant \tfrac{2p}{r}t^{\frac{2p}{r}-1}f(t)-\kappa_0 t^{\frac{2p}{r}}f(t)^{1+\frac{r}{p}}+\kappa_1 t^{\frac{2p}{r}}\\
&= \tfrac{2p}{r}t^{\frac{p-r}{p+r}}t^{\frac{2p^2}{r(p+r)}}f(t)-\kappa_0 t^{\frac{2p}{r}}f(t)^{1+\frac{r}{p}}+\kappa_1 t^{\frac{2p}{r}}\\
&\leqslant Ct^{\frac{p}{r}-1}+\kappa_1 t^{\frac{2p}{r}}.
\end{aligned}
$$

Integrating both sides from 0 to t, we obtain

$$
t^{\frac{2p}{r}}f(t)\leqslant \tfrac{rC}{p}t^{\frac{p}{r}}+\tfrac{r\kappa_1}{r+2p}t^{\frac{2p}{r}+1}\Rightarrow f(t)\leqslant \tfrac{rC}{p}t^{-\frac{p}{r}}+\tfrac{r\kappa_1}{r+2p}t. \tag{31}
$$

Combining (30) and (31), we obtain

$$
f(2t)\leqslant \mathrm{e}^{-\kappa_0 t}f(t)+\tfrac{\kappa_1}{\kappa_0}\leqslant \mathrm{e}^{-\kappa_0 t}(\tfrac{rC}{p}t^{-\frac{p}{r}}+\tfrac{r\kappa_1}{r+2p}t)+\tfrac{\kappa_1}{\kappa_0},
$$

which in turn gives the second estimate in (26) for $r>0$. □

Let $\phi:\mathbb{R}^d\to\mathbb{R}_+$ be a smooth density function with support in the unit ball. We define

$$
\begin{aligned}
&\phi_\varepsilon(x):=\varepsilon^{-d}\phi(\varepsilon^{-1}x),\ \ x\in\mathbb{R}^d,\ \varepsilon\in(0,1),\\
&b_\varepsilon(x):=b*\phi_\varepsilon(x),\quad \sigma_\varepsilon(x):=\sigma*\phi_\varepsilon(x).
\end{aligned}
$$

By $(\mathbf{H}_{\mathrm{loc}})$, it is easy to see that for any $m\in\mathbb{N}$ and all $\varepsilon\in(0,1)$,

$$
|b_\varepsilon(x)-b_\varepsilon(y)|+\|\sigma_\varepsilon(x)-\sigma_\varepsilon(y)\|\leqslant C_{m+1}|x-y|^\gamma,\ \ \forall x,y\in B_m, \tag{32}
$$

and for some $\varepsilon_1\in(0,1)$ depending on m, and for all $\varepsilon\in(0,\varepsilon_1)$,

$$
C_m^{-1}|\xi|\leqslant|\sigma_\varepsilon(x)\xi|\leqslant C_m|\xi|,\ \ \forall x\in B_m,\ \xi\in\mathbb{R}^d. \tag{33}
$$

Moreover, note that by (4),

$$
\begin{aligned}
|b_\varepsilon(x)-b(x)|&\leqslant\int_{\mathbb{R}^d}|b(x-y)-b(x)|\phi_\varepsilon(y)\mathrm{d}y\leqslant\varepsilon^\gamma\ell_1(x),\\
\|\sigma_\varepsilon(x)-\sigma(x)\|&\leqslant\int_{\mathbb{R}^d}\|\sigma(x-y)-\sigma(x)\|\phi_\varepsilon(y)\mathrm{d}y\leqslant\varepsilon^\gamma\ell_2(x).
\end{aligned}
$$

Hence, by $(\mathbf{H}^{r,q}_{\mathrm{glo}})$, for any $\varepsilon\in(0,\varepsilon_0^{1/\gamma})$,

$$
\begin{aligned}
\langle x,b_\varepsilon(x)\rangle+q|\sigma_\varepsilon(x)|^\alpha|x|^{2-\alpha}&\leqslant\langle x,b(x)\rangle+\varepsilon^\gamma|x|\ell_1(x)+q(\|\sigma(x)\|+\ell_2(x)\varepsilon^\gamma)^\alpha|x|^{2-\alpha}\\
&\leqslant\langle x,b(x)\rangle+\varepsilon_0|x|\ell_1(x)+q(\|\sigma(x)\|+\ell_2(x)\varepsilon_0)^\alpha|x|^{2-\alpha}\\
&\leqslant -c_0|x|^{2+r}+c_1.
\end{aligned} \tag{34}
$$

Now, we consider the approximation SDE

$$
\mathrm{d}X_t^\varepsilon=b_\varepsilon(X_t^\varepsilon)\mathrm{d}t+\sigma_\varepsilon(X_{t-}^\varepsilon)\mathrm{d}L_t^\alpha,\quad X_0^\varepsilon=x. \tag{35}
$$

Since the coefficients are smooth, but, may be not bounded, by (34) and a standard localization technique, there is a unique global strong solution $X_t^\varepsilon(x)$ to the above SDE. Moreover, for any $f \in C_b(\mathbb{R}^d)$, we have

$$\lim_{\varepsilon \to 0} \mathbb{E} f(X_t^\varepsilon(x)) = \mathbb{E} f(X_t(x)), \tag{36}$$

which is proven in Appendix. Notice that if b and σ are locally Lipschitz, then it is not necessary to mollify b and σ, and it suffices to consider the truncated b and σ as done below.

Let $\chi \in C_0^\infty(\mathbb{R}^d)$ be a cutoff function with

$$\chi(x) = 1, \quad |x| \leqslant 1, \quad \chi(x) = 0, \quad |x| > 2.$$

For $m \in \mathbb{N}$, set

$$\chi_m(x) := \chi(x/m), \quad b_m^\varepsilon(x) := b_\varepsilon(x)\chi_m(x), \quad \sigma_m^\varepsilon(x) := \sigma_\varepsilon(x\chi_m(x)).$$

By (32) and (33), it is easy to see that b_m^ε and σ_m^ε satisfy the following global assumptions:

$$|b_m^\varepsilon(x) - b_m^\varepsilon(y)| + \|\sigma_m^\varepsilon(x) - \sigma_m^\varepsilon(y)\| \leqslant C_m |x-y|^\gamma, \quad \forall x, y \in \mathbb{R}^d, \tag{37}$$

$$C_m^{-1}|\xi| \leqslant |\sigma_m^\varepsilon(x)\xi| \leqslant C_m|\xi|, \quad \forall x, \xi \in \mathbb{R}^d, \tag{38}$$

where the constant does not depend on ε. Let $\theta_t(x)$ solve the ODE

$$\dot{\theta}_t(x) = (b_m^\varepsilon * \phi_{t^{1/\alpha}})(\theta_t(x)), \quad \theta_0(x) = x. \tag{39}$$

We also consider the SDE with cutoff coefficients

$$\mathrm{d}X_t^{\varepsilon,m} = b_m^\varepsilon(X_t^{\varepsilon,m})\mathrm{d}t + \sigma_m^\varepsilon(X_{t-}^{\varepsilon,m})\mathrm{d}L_t^\alpha, \quad X_0^{\varepsilon,m} = x. \tag{40}$$

Let $X_t^{\varepsilon,m}(x)$ be the unique strong solution of the above SDE. By [33, Theorem 1.1], $X_t^{\varepsilon,m}(x)$ admits a density $p_m^\varepsilon(t,x,y)$ enjoying the estimates that for any $T > 0$ and $m \in \mathbb{N}$, there is a constant $C_0 \geqslant 1$ depending on m, but *independent of* $\varepsilon \in (0, \varepsilon_1)$, such that for all $t \in (0, T]$ and $x, y \in \mathbb{R}^d$,

$$p_m^\varepsilon(t,x,y) \asymp_{C_0} t(t^{\frac{1}{\alpha}} + |x - \theta_t(y)|)^{-d-\alpha}, \tag{41}$$

$$|\nabla \log p_m^\varepsilon(t,x,\cdot)|(y) \leqslant C_0 t^{-\frac{1}{\alpha}}. \tag{42}$$

For any $m \in \mathbb{N}$ and $x \in B_m$, we define the exit time of $X_t^{\varepsilon,m}(x)$ from B_m by

$$\tau_{B_m}^{\varepsilon,x} := \inf\left\{t > 0 : X_t^\varepsilon(x) \notin B_m\right\}.$$

By the uniqueness of strong solution, we have

$$X_t^\varepsilon(x) = X_t^{\varepsilon,m}(x), \quad t < \tau_{B_m}^{\varepsilon,x}. \tag{43}$$

Lemma 3.3 (Strong Feller property). *For any $t > 0$ and bounded measurable function f, the function $x \mapsto P_t f(x)$ is bounded continuous.*

Proof. By Lusin's theorem, it suffices to show that for any $t>0$,

$$\lim_{x\to y}\sup_{f\in C_b(\mathbb{R}^d),\|f\|_\infty\leqslant 1}|\mathbb{E}(f(X_t(x))-f(X_t(y)))|=0. \tag{44}$$

Indeed, for fixed $y\in\mathbb{R}^d$ and any $\varepsilon>0$, by (44), there is a $\delta>0$ such that for all $|x-y|\leqslant\delta$,

$$\sup_{f\in C_b(\mathbb{R}^d),\|f\|_\infty\leqslant 1}|\mathbb{E}(f(X_t(x))-f(X_t(y)))|\leqslant\varepsilon. \tag{45}$$

For any bounded measurable function f with $\|f\|_\infty\leqslant 1$, by Lusin's theorem, there is a continuous function $f_\varepsilon\in C_b(\mathbb{R}^d)$ with $\|f_\varepsilon\|_\infty\leqslant 1$ (possiblly depending on x,y) such that

$$|\mathbb{E}(f(X_t(x))-f_\varepsilon(X_t(x)))|\leqslant\varepsilon,\ \ |\mathbb{E}(f(X_t(y))-f_\varepsilon(X_t(y)))|\leqslant\varepsilon,$$

which together with (45) yields that for all $|x-y|\leqslant\delta$,

$$|\mathbb{E}(f(X_t(x))-f(X_t(y)))|\leqslant 3\varepsilon.$$

Thus we obtain the strong Feller property. Since for any $f\in C_b(\mathbb{R}^d)$,

$$\lim_{\varepsilon\to 0}\mathbb{E}(f(X^\varepsilon_t(x))-f(X^\varepsilon_t(y)))\overset{(36)}{=}\mathbb{E}(f(X_t(x))-f(X_t(y))),$$

for (44), we only need to prove that

$$\lim_{x\to y}\sup_{\varepsilon\in(0,1)}\sup_{f\in C_b(\mathbb{R}^d),\|f\|_\infty\leqslant 1}|\mathbb{E}(f(X^\varepsilon_t(x))-f(X^\varepsilon_t(y)))|=0. \tag{46}$$

Given $x,y\in\mathbb{R}^d$, let $m>|x|\vee|y|$. For given $f\in C_b(\mathbb{R}^d)$ with $\|f\|_\infty\leqslant 1$, we have

$$\begin{aligned}|\mathbb{E}(f(X^\varepsilon_t(x))-f(X^\varepsilon_t(y)))|&\leqslant\Big|\mathbb{E}\big[(f(X^\varepsilon_t(x))-f(X^\varepsilon_t(y)))\mathbf{1}_{t<\tau^{\varepsilon,x}_{B_m}\wedge\tau^{\varepsilon,y}_{B_m}}\big]\Big|+2\mathbb{P}\big(t\geqslant\tau^{\varepsilon,x}_{B_m}\wedge\tau^{\varepsilon,y}_{B_m}\big)\\&\overset{(43)}{=}\Big|\mathbb{E}\big[(f(X^{\varepsilon,m}_t(x))-f(X^{\varepsilon,m}_t(y)))\mathbf{1}_{t<\tau^{\varepsilon,x}_{B_m}\wedge\tau^{\varepsilon,y}_{B_m}}\big]\Big|+2\mathbb{P}\big(t\geqslant\tau^{\varepsilon,x}_{B_m}\wedge\tau^{\varepsilon,y}_{B_m}\big)\\&\leqslant\Big|\mathbb{E}\big[(f(X^{\varepsilon,m}_t(x))-f(X^{\varepsilon,m}_t(y))\big]\Big|+2\mathbb{P}\big(t\geqslant\tau^{\varepsilon,x}_{B_m}\wedge\tau^{\varepsilon,y}_{B_m}\big).\end{aligned}$$

By (42), we have

$$|\mathbb{E}(f(X^{\varepsilon,m}_t(x))-f(X^{\varepsilon,m}_t(y)))|=\left|\int_{\mathbb{R}^d}f(z)(p^\varepsilon_m(t,x,z)-p^\varepsilon_m(t,y,z))\mathrm{d}z\right|\leqslant C_m t^{-\frac{1}{\alpha}}|x-y|.$$

Moreover, by Chebyshev's inequality and (25) with $p\in(0,\alpha)$,

$$\mathbb{P}(t\geqslant\tau^{\varepsilon,x}_{B_m})\leqslant\mathbb{P}\left(\sup_{s\leqslant t}|X^\varepsilon_s(x)|\geqslant m\right)\leqslant\mathbb{E}\left(\sup_{s\leqslant t}|X^\varepsilon_s(x)|^{p/2}\right)/m^{p/2}\leqslant C(V_p(x)+t)^{1/2}/m^{p/2}. \tag{47}$$

Combining the above calculations, we obtain that for any $f\in C_b(\mathbb{R}^d)$ with $\|f\|_\infty\leqslant 1$,

$$|\mathbb{E}(f(X^\varepsilon_t(x))-f(X^\varepsilon_t(y)))|\leqslant C_m t^{-\frac{1}{\alpha}}|x-y|+C_0(V_p(x)+V_p(y)+t)^{1/2}/m^{p/2},$$

where C_0 does not depend on m. Thus we obtain (46) by first letting $x\to y$ and then $m\to\infty$. $\square$

Lemma 3.4 (Irreducibility). *For any $x_0, y_0 \in \mathbb{R}^d$ and $r, t > 0$, we have*

$$\inf_{x \in B_r(x_0)} \mathbb{P}(X_t(x) \in B_r(y_0)) > 0. \tag{48}$$

Proof. Since $X_t^\varepsilon(x)$ weakly converges to $X_t(x)$ as $\varepsilon \downarrow 0$, for any open set $A \subset \mathbb{R}^d$, we have

$$\mathbb{P}(X_t(x) \in \overline{A}) \geqslant \liminf_{\varepsilon \to 0} \mathbb{P}(X_t^\varepsilon(x) \in A). \tag{49}$$

Below we fix $\varepsilon \in (0,1)$ small enough, and $x, y_0 \in \mathbb{R}^d$, $r > 0$. Let $m > 2$ be big enough such that $\{x\} \cup B_r(y_0) \subset B_{m-2}$. Then we have

$$\mathbb{P}(X_t^\varepsilon(x) \in B_r(y_0)) \geqslant \mathbb{P}\Big(X_t^\varepsilon(x) \in B_r(y_0); t < \tau_{B_m}^{\varepsilon,x}\Big) \overset{(43)}{=} \mathbb{P}\Big(X_t^{\varepsilon,m}(x) \in B_r(y_0); t < \tau_{B_m}^{\varepsilon,x}\Big). \tag{50}$$

In order to use Theorem 2.3, we choose $\rho(t,r) = t(t^{1/\alpha} + r)^{-d-\alpha}$. Clearly, by (41), estimate (16) is satisfied for $\Gamma_t(x,y) = |x - \theta_t(y)|$. It remains to find t_0 and δ_0 small enough as well as functions $\ell_0(t)$ and $\ell_1(t)$ so that the conditions in Theorem 2.3 are satisfied for domains $D_0 = B_{m-2}$ and $D = B_m$. Note that by (39) and the definition of b_m^ε, for all $t \in [0,1]$ and $y \in \mathbb{R}^d$,

$$|\theta_t(y) - y| \leqslant \int_0^t |(b_m^\varepsilon * \phi_{s^{1/\alpha}})(\theta_s(y))| \mathrm{d}s \leqslant \sup_{|x| \leqslant m+1} |b(x)| \cdot t =: C_1 t.$$

Let

$$\delta_0 \leqslant \tfrac{1}{2} \wedge \mathrm{dist}(y_0, D_0), \quad t_0 \leqslant \tfrac{1}{2C_1} \wedge \tfrac{\delta_0}{C_1}. \tag{51}$$

For $x' \notin D$ and $y \in B_{\delta_0}(y_0) \subset D_0$, we have for $t \in (0, t_0]$,

$$\Gamma_t(x', y) = |x' - \theta_t(y)| \geqslant |x' - y_0| - |y - y_0| - |\theta_t(y) - y| \geqslant 2 - \delta_0 - C_1 t \geqslant 1.$$

On the other hand, for $x, y \in B_{\delta_0}(y_0)$, we have

$$\Gamma_t(x,y) = |x - \theta_t(y)| \leqslant |x - y| + |\theta_t(y) - y| \leqslant 2\delta_0 + C_1 t \leqslant 3\delta_0. \tag{52}$$

Now, for the C_0 in (41), one can choose $\delta_0, t_0 > 0$ small enough such that (51) holds and

$$t \mapsto t/(t^{1/\alpha} + 1)^{d+\alpha} \text{ is increasing on } (0, t_0],$$

and

$$\frac{C_0^{-1} t}{(t^{1/\alpha} + 3\delta_0)^{d+\alpha}} > \frac{2C_0 t}{(t^{1/\alpha} + 1)^{d+\alpha}}, \quad t \in (0, t_0]. \tag{53}$$

In particular, (18) is satisfied. Thus by Theorem 2.3, we conclude

$$\mathbb{P}\Big(X_t^{\varepsilon,m}(x) \in B_r(y_0); t < \tau_{B_m}^{\varepsilon,x}\Big) \geqslant c_0,$$

where c_0 is independent of ε. This, together with (49) and (50), yields

$$\mathbb{P}(X_t(x) \in \overline{B_r(y_0)}) \geqslant c_0.$$

The proof is thus complete by the strong Feller property that $x \mapsto \mathbb{P}(X_t(x) \in B_r(y_0))$ is continuous. $\square$

Now, we are in a position to prove Theorem 1.2.

Proof of Theorem 1.2. First of all, by (25) and the standard Krylov-Bogoliubov method (see [14, Theorem 11.7]), there is an invariant probability measure μ associated with $(P_t)_{t\geqslant 0}$. The uniqueness follows by the strong Feller property and irreducibility. The exponential convergence (i) and (ii) follow by (26), Theorem 2.1 and Lemmas 3.3 and 3.4. □

4. Application to heavy-tailed sampling

In this section, we introduce an application of Theorem 1.2 to the heavy-tailed sampling. Let $\mu(\mathrm{d}x) = \mathrm{e}^{-U(x)}\mathrm{d}x/\int_{\mathbb{R}^d}\mathrm{e}^{-U(x)}\mathrm{d}x$, where $U:\mathbb{R}^d\to\mathbb{R}$ is a continuous function. We suppose that there are $\beta, C_0 > 0$ such that for all $|x|\geqslant 1$,

$$U(x) \geqslant (d+\beta)\log|x| - C_0. \tag{54}$$

The above assumption means that $\mathrm{e}^{-U(x)}$ has a polynomial decay rate as $|x|\to\infty$, which characterizes a heavy-tailed distribution, as opposed to the exponential or light-tailed one. Below we want to find an ergodic SDE so that the law of the solution X_t exponentially converges to μ in some sense as $t\to\infty$. Thus, one can sample μ theoretically from X_t when t is large.

Fix $\alpha\in(0,2)$. To construct an ergodic SDE driven by α-stable processes, we introduce a vector field $B:\mathbb{R}^d\to\mathbb{R}^d$ by

$$B(x) := \tilde{c}_{d,\alpha}\int_{\mathbb{R}^d} y[\varrho_\alpha(x+y)-\varrho_\alpha(x-y)]|y|^{-d-\alpha}\mathrm{d}y,$$

where $\tilde{c}_{d,\alpha} = c_{d,\alpha}/\alpha$ and $c_{d,\alpha}$ is the same constant as (7), and

$$\varrho_\alpha(x) := (1+|x|^2)^{-\frac{d+\alpha}{2}}.$$

The vector field B enjoys the following important properties.

Lemma 4.1. (i) *$B\in C_0^\infty$ satisfies that for some $\kappa_0=\kappa_0(d,\alpha)>0$,*

$$|B(x)| \leqslant \kappa_0|x|^{1-d-\alpha}, \quad |\nabla B(x)| \leqslant \kappa_0|x|^{-d-\alpha}. \tag{55}$$

(ii) *There is a constant $\kappa_1>0$ only depending on d,α such that for all $|x|>1$,*

$$\langle x, B(x)\rangle \leqslant -\kappa_1|x|^{2-d-\alpha}. \tag{56}$$

(iii) $\mathrm{div}B(x) = \Delta^{\frac{\alpha}{2}}\varrho_\alpha(x)$.

Proof. (i) Since $\varrho_\alpha\in C_0^\infty(\mathbb{R}^d)$ and for any $j\in\mathbb{N}$, $\|\nabla^j\varrho_\alpha/\varrho_\alpha\|_\infty<\infty$, we clearly have $B\in C_0^\infty$. To show the bounds (55), let $\chi:[0,\infty)\to[0,1]$ be a cutoff function with

$$\chi(r)=1,\ \ r\in[0,\tfrac14],\ \ \chi(r)=0,\ \ r\in[0,\tfrac12]^c.$$

Define

$$\varphi_{|x|}(y) := y|y|^{-d-\alpha}\chi(\tfrac{|y|}{|x|}),\ \ \widetilde{\varphi}_{|x|}(y) := y|y|^{-d-\alpha}(1-\chi(\tfrac{|y|}{|x|})).$$

For $j=0,1$, by definition we have

$$\nabla^j B(x)=\tilde{c}_{d,\alpha}\int_{\mathbb{R}^d} y\Big(\nabla^j \varrho_\alpha(x+y)-\nabla^j \varrho_\alpha(x-y)\Big)|y|^{-d-\alpha}\mathrm{d}y=\tilde{c}_{d,\alpha}(I_1+I_2),$$

where

$$I_1:=\int_{\mathbb{R}^d}\varphi_{|x|}(y)\Big(\nabla^j \varrho_\alpha(x+y)-\nabla^j \varrho_\alpha(x-y)\Big)\mathrm{d}y,$$
$$I_2:=\int_{\mathbb{R}^d}\widetilde{\varphi}_{|x|}(y)\Big(\nabla^j \varrho_\alpha(x+y)-\nabla^j \varrho_\alpha(x-y)\Big)\mathrm{d}y.$$

For I_1, we have

$$\begin{aligned}|I_1|&\leqslant\int_{\mathbb{R}^d}\varphi_{|x|}(y)|y|\left(\int_{-1}^1|\nabla^{j+1}\varrho_\alpha(x+sy)|\mathrm{d}s\right)\mathrm{d}y\\&\lesssim\int_{|y|\leqslant\frac{|x|}{2}}|y|^{2-d-\alpha}\left(\int_{-1}^1(1+|x+sy|)^{-(d+\alpha+1+j)}\mathrm{d}s\right)\mathrm{d}y\\&\lesssim(1+|x|)^{-(d+\alpha+1+j)}\int_{|y|\leqslant\frac{|x|}{2}}|y|^{2-d-\alpha}\mathrm{d}y\\&\lesssim(1+|x|)^{-(d+\alpha+1+j)}|x|^{2-\alpha}\lesssim|x|^{1-j-d-2\alpha}.\end{aligned}$$

For I_2, by the change of variable, we have

$$\begin{aligned}I_2&=\int_{\mathbb{R}^d}\Big(\widetilde{\varphi}_{|x|}(x-y)-\widetilde{\varphi}_{|x|}(x+y)\Big)(\nabla^j\varrho_\alpha)(y)\mathrm{d}y\\&=|x|^d\int_{\mathbb{R}^d}\Big(\widetilde{\varphi}_{|x|}(|x|(\bar{x}-y))-\widetilde{\varphi}_{|x|}(|x|(\bar{x}+y))\Big)(\nabla^j\varrho_\alpha)(|x|y)\mathrm{d}y\\&=|x|^{1-\alpha}\int_{\mathbb{R}^d}\Big(\widetilde{\varphi}_1(\bar{x}-y)-\widetilde{\varphi}_1(\bar{x}+y)\Big)(\nabla^j\varrho_\alpha)(|x|y)\mathrm{d}y,\end{aligned}$$

where $\bar{x}=x/|x|$ and in the last step we have used

$$\widetilde{\varphi}_{|x|}(|x|y)=|x|^{1-d-\alpha}y|y|^{-d-\alpha}\big(1-\chi\big(|y|\big)\big)=|x|^{1-d-\alpha}\widetilde{\varphi}_1(y).$$

For $j=0$, since $|\widetilde{\varphi}_1(y)|\leqslant|y|^{1-d-\alpha}\mathbf{1}_{|y|>1/4}\leqslant 4^{d+\alpha-1}$, it is easy to see that

$$|I_2|\leqslant 2\cdot 4^{d+\alpha-1}|x|^{1-\alpha}\int_{\mathbb{R}^d}|(1+|x||y|)^{-(d+\alpha)}\mathrm{d}y\lesssim|x|^{1-\alpha-d}.$$

For $j=1$, since $\|\mathrm{div}_y\widetilde{\varphi}_1\|_\infty<\infty$, using the integration by parts, we also have

$$\begin{aligned}|I_2|&=|x|^{-\alpha}\left|\int_{\mathbb{R}^d}\Big(\widetilde{\varphi}_1(\bar{x}-y)-\widetilde{\varphi}_1(\bar{x}+y)\Big)\nabla_y\varrho_\alpha(|x|y)\mathrm{d}y\right|\\&=|x|^{-\alpha}\left|\int_{\mathbb{R}^d}\Big(\mathrm{div}_y\widetilde{\varphi}_1(\bar{x}-y)-\mathrm{div}_y\widetilde{\varphi}_1(\bar{x}+y)\Big)\varrho_\alpha(|x|y)\mathrm{d}y\right|\\&\leqslant 2\|\mathrm{div}_y\widetilde{\varphi}_1\|_\infty|x|^{-\alpha}\int_{\mathbb{R}^d}|(1+|x||y|)^{-(d+\alpha)}\mathrm{d}y\lesssim|x|^{-\alpha-d}.\end{aligned}$$

Combining the above estimates we obtain (55).

(ii) By definition and the symmetry, we have

$$-\langle x, B(x)\rangle = \tilde{c}_{d,\alpha}\int_{\mathbb{R}^d}\langle x,y\rangle[\varrho_\alpha(x-y)-\varrho_\alpha(x+y)]|y|^{-d-\alpha}\mathrm{d}y$$
$$=2\tilde{c}_{d,\alpha}\int_{\langle x,y\rangle\geqslant 0}\langle x,y\rangle[\varrho_\alpha(x-y)-\varrho_\alpha(x+y)]|y|^{-d-\alpha}\mathrm{d}y.$$

For $|x|\geqslant 1$ and $\langle x,y\rangle\geqslant 0$, by the mean-valued formula, we have

$$\varrho_\alpha(x-y)-\varrho_\alpha(x+y) = (1+|x-y|^2)^{-\frac{d+\alpha}{2}} - (1+|x+y|^2)^{-\frac{d+\alpha}{2}}$$
$$=2(d+\alpha)\langle x,y\rangle\int_0^1\big(1+s(|x-y|^2-|x+y|^2)+|x+y|^2\big)^{-\frac{d+\alpha}{2}-1}\mathrm{d}s$$
$$=2(d+\alpha)\langle x,y\rangle\int_0^1\big(1-4s\langle x,y\rangle+|x+y|^2\big)^{-\frac{d+\alpha}{2}-1}\mathrm{d}s$$
$$=2(d+\alpha)\langle x,y\rangle\int_0^1\big(1+4s\langle x,y\rangle+|x-y|^2\big)^{-\frac{d+\alpha}{2}-1}\mathrm{d}s$$
$$\geqslant 2(d+\alpha)\langle x,y\rangle\int_0^{|x|^{-2}}\big(1+4s\langle x,y\rangle+|x-y|^2\big)^{-\frac{d+\alpha}{2}-1}\mathrm{d}s$$
$$\geqslant 2(d+\alpha)\langle x,y\rangle\big(1+4|x|^{-2}\langle x,y\rangle+|x-y|^2\big)^{-\frac{d+\alpha}{2}-1}|x|^{-2}.$$

Hence,

$$-\langle x, B(x)\rangle \geqslant \frac{4(d+\alpha)\tilde{c}_{d,\alpha}}{|x|^2}\int_{\langle x,y\rangle\geqslant 0}\langle x,y\rangle^2\big(1+4|x|^{-2}\langle x,y\rangle+|x-y|^2\big)^{-\frac{d+\alpha}{2}-1}|y|^{-d-\alpha}\mathrm{d}y$$
$$\geqslant \frac{4(d+\alpha)\tilde{c}_{d,\alpha}}{|x|^2}\int_{\langle x,y\rangle\geqslant 0,|x-y|\leqslant 1}\langle x,y\rangle^2\big(1+4|x|^{-2}\langle x,y\rangle+|x-y|^2\big)^{-\frac{d+\alpha}{2}-1}\frac{\mathrm{d}y}{|y|^{d+\alpha}}.$$

Since for $|x|\geqslant 1$, $|x-y|\leqslant 1$ implies $\frac{|y|^2}{2}\leqslant\langle x,y\rangle\leqslant|x|^2+|x|$, we further have

$$-\langle x, B(x)\rangle \geqslant \frac{4(d+\alpha)\tilde{c}_{d,\alpha}}{|x|^2}\int_{|x-y|\leqslant 1}\langle x,y\rangle^2\big(2+4(1+|x|^{-1})\big)^{-\frac{d+\alpha}{2}-1}|y|^{-d-\alpha}\mathrm{d}y$$
$$\geqslant \frac{4(d+\alpha)\tilde{c}_{d,\alpha}}{|x|^2 10^{\frac{d+\alpha}{2}+1}}\int_{|x-y|\leqslant 1}\langle x,y\rangle^2|y|^{-d-\alpha}\mathrm{d}y\geqslant\kappa_1|x|^{2-d-\alpha},$$

where $\kappa_1>1$ only depends on d,α. Thus we obtain (56).

(iii) By definition and the integration by parts, we have

$$\mathrm{div}B(x) = \tilde{c}_{d,\alpha}\int_{\mathbb{R}^d}\langle y,\nabla_x[\varrho_\alpha(x+y)-\varrho_\alpha(x-y)]\rangle|y|^{-d-\alpha}\mathrm{d}y$$
$$=\tilde{c}_{d,\alpha}\int_{\mathbb{R}^d}\langle y,\nabla_y[\varrho_\alpha(x+y)+\varrho_\alpha(x-y)-2\varrho_\alpha(x)]\rangle|y|^{-d-\alpha}\mathrm{d}y$$

$$= -\tilde{c}_{d,\alpha}\int_{\mathbb{R}^d}\mathrm{div}(y|y|^{-d-\alpha})[\varrho_\alpha(x+y)+\varrho_\alpha(x-y)-2\varrho_\alpha(x)]\mathrm{d}y.$$

Since for $y\neq 0$,

$$\mathrm{div}(y|y|^{-d-\alpha}) = d|y|^{-d-\alpha}+\langle y,\nabla|y|^{-d-\alpha}\rangle = -\alpha|y|^{-d-\alpha},$$

we have by (8),

$$\mathrm{div}B(x) = \alpha\tilde{c}_{d,\alpha}\int_{\mathbb{R}^d}\frac{[\varrho_\alpha(x+y)+\varrho_\alpha(x-y)-2\varrho_\alpha(x)]}{|y|^{d+\alpha}}\mathrm{d}y = \Delta^{\frac{\alpha}{2}}\varrho_\alpha(x).$$

The proof is complete. □

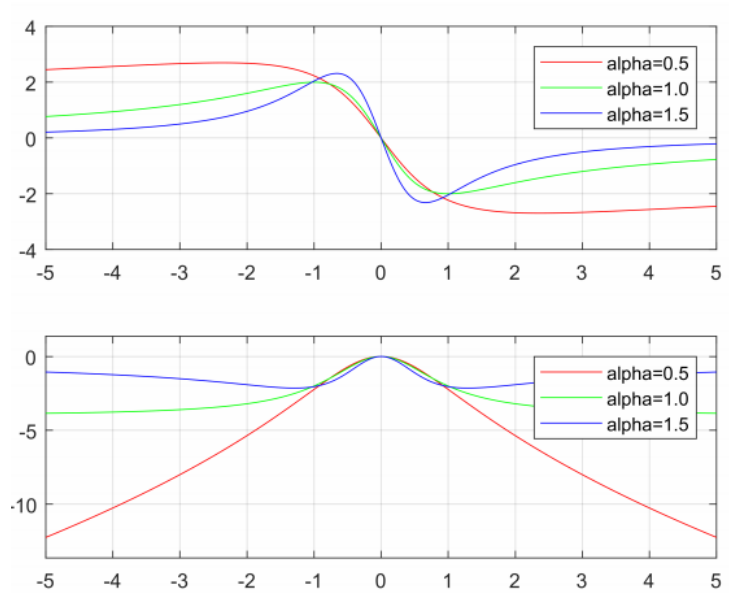

Figure 1. Pictures of $B(x)$ and $\langle x, B(x)\rangle$ for $d=1$.

Now we take σ and b in SDE (3) as follows:

$$\sigma(x) := (\varrho_\alpha(x)\mathrm{e}^{U(x)})^{\frac{1}{\alpha}}\mathbb{I},\quad b(x) := B(x)\mathrm{e}^{U(x)}. \tag{57}$$

We have the main result of this section.

Theorem 4.2. *Suppose that U is locally Lipschitz continuous and satisfies (54). Then with the above choices of σ and b, $\mu(\mathrm{d}x)=\mathrm{e}^{-U(x)}\mathrm{d}x/\int_{\mathbb{R}^d}\mathrm{e}^{-U(x)}\mathrm{d}x$ is the unique invariant probability measure of SDE (3) and the conclusions in Theorem 1.2 hold with $r=\beta-\alpha$.*

Proof. Let σ and b be defined by (57). By (7) and the change of variable, we can write

$$\mathcal{L}_\sigma f(x) = c_{d,\alpha}\int_{\mathbb{R}^d}\frac{f(x+\sigma(x)z)+f(x-\sigma(x)z)-2f(x)}{|z|^{d+\alpha}}\mathrm{d}z = \varrho_\alpha(x)\mathrm{e}^{U(x)}\Delta^{\frac{\alpha}{2}}f(x).$$

Thus,

$$\mathscr{L}f(x)=\mathcal{L}_\sigma f(x)+b\cdot\nabla f(x)=\varrho_\alpha(x)\mathrm{e}^{U(x)}\Delta^{\frac{\alpha}{2}}f(x)+B(x)\mathrm{e}^{U(x)}\cdot\nabla f(x), \tag{58}$$

and by (iii) of Lemma 4.1,

$$\begin{aligned}\int_{\mathbb{R}^d}\mathscr{L}f(x)\mathrm{e}^{-U(x)}\mathrm{d}x&=\int_{\mathbb{R}^d}\Big(\varrho_\alpha(x)\Delta^{\frac{\alpha}{2}}f(x)+B(x)\cdot\nabla f(x)\Big)\mathrm{d}x\\&=\int_{\mathbb{R}^d}(\Delta^{\frac{\alpha}{2}}\varrho_\alpha(x)-\mathrm{div}B(x))f(x)\mathrm{d}x=0,\end{aligned}$$

which implies that $\mu(\mathrm{d}x)=\mathrm{e}^{-U(x)}\mathrm{d}x/\int_{\mathbb{R}^d}\mathrm{e}^{-U(x)}\mathrm{d}x$ is an invariant probability measure of SDE (3).

In order to check $(\mathbf{H}_{\mathrm{loc}})$ and $(\mathbf{H}^{r,q}_{\mathrm{glo}})$, by Remark 1.1, it suffices to verify (5) for $\varepsilon_0=0$. By (56) and (54), for any $|x|\geqslant(2q/\kappa_1)^{1/\alpha}\vee 1$, we have

$$\begin{aligned}\langle x,b(x)\rangle+q\|\sigma(x)\|^\alpha|x|^{2-\alpha}&=\big[\langle x,B(x)\rangle+q\varrho_\alpha(x)|x|^{2-\alpha}\big]\mathrm{e}^{U(x)})\\&\leqslant\big[-\kappa_1|x|^{2-d-\alpha}+q|x|^{2-d-2\alpha}\big]\mathrm{e}^{U(x)}\\&\leqslant-\tfrac{\kappa_1}{2}|x|^{2-d-\alpha}\mathrm{e}^{U(x)}\leqslant-\tfrac{\kappa_1}{2\mathrm{e}^{C_0}}|x|^{2+\beta-\alpha}.\end{aligned}$$

So, Theorem 1.2 is applicable to complete the proof. $\square$

Remark 4.3. Suppose that $U(x)=-\ln\varrho_\alpha(x)$. By (55), $\|\nabla(B/\varrho_\alpha)\|_\infty<\infty$. Let X_t be the unique solution of the following SDE:

$$\mathrm{d}X_t=(B/\rho_\alpha)(X_t)\mathrm{d}t+\mathrm{d}L^\alpha_t,\quad X_0=x. \tag{59}$$

Then $\mu_0(\mathrm{d}x)=\varrho_\alpha(x)\mathrm{d}x/\int\varrho_\alpha(x)\mathrm{d}x$ is the unique invariant probability measure of X_t. It should be noticed that the distributional density $p_\alpha(x)$ of L^α_1 is comparable with $\varrho_\alpha(x)$ (cf. [44]), i.e., there is a constant $\kappa=\kappa(d,\alpha)\geqslant 1$ such that for all $x\in\mathbb{R}^d$, $\kappa^{-1}\varrho_\alpha(x)\leqslant p_\alpha(x)\leqslant\kappa\varrho_\alpha(x)$. Let $\mathscr{L}_0$ be the infinitesimal generator of SDE (59), i.e., $\mathscr{L}_0=\Delta^{\frac{\alpha}{2}}+B/\varrho_\alpha\cdot\nabla$. Then by (58), one sees that $\mathscr{L}=(\varrho_\alpha\mathrm{e}^U)\mathscr{L}_0$. In particular, the solution of SDE (3) with coefficients (57) is just a time change of SDE (59) (see [3, Section 1.15.2]).

Remark 4.4. The locally Lipschitz assumption on U can be relaxed as locally γ-Hölder continuity with $\gamma\in((1-\alpha)^+,1)$. If it is so, we need to check (5) for some $\varepsilon_0>0$.

5. Appendix

In this appendix, we sketch the proof of weak convergence (36). First of all, by (34) and (25), there is a constant $C>0$ such that for all $\varepsilon\in(0,1)$ and any $x\in\mathbb{R}^d$ and $T>0$,

$$\left[\mathbb{E}\left(\sup_{s\in[0,T]}V_p(X^\varepsilon_s(x))^{\frac{1}{2}}\right)\right]^2\lesssim_C V_p(x)+T. \tag{60}$$

For $\theta \in (0,T)$, let η, η' be two stopping times with $0 \leqslant \eta \leqslant \eta' + \theta \leqslant T$. For any $m \in \mathbb{N}$, we have

$$\begin{aligned}\mathbb{P}\left(|X^{\varepsilon}_{\eta} - X^{\varepsilon}_{\eta'}| \geqslant \delta\right) &\leqslant \mathbb{P}\left(|X^{\varepsilon}_{\eta} - X^{\varepsilon}_{\eta'}| \geqslant \delta; T < \tau^{\varepsilon,x}_{B_m}\right) + \mathbb{P}\left(T \geqslant \tau^{\varepsilon,x}_{B_m}\right)\\ &= \mathbb{P}\left(|X^{\varepsilon,m}_{\eta} - X^{\varepsilon,m}_{\eta'}| \geqslant \delta; T < \tau^{\varepsilon,x}_{B_m}\right) + \mathbb{P}\left(T \geqslant \tau^{\varepsilon,x}_{B_m}\right)\\ &\leqslant \mathbb{P}\left(|X^{\varepsilon,m}_{\eta} - X^{\varepsilon,m}_{\eta'}| \geqslant \delta\right) + 2\mathbb{P}\left(T \geqslant \tau^{\varepsilon,x}_{B_m}\right).\end{aligned}$$

By SDE (40) and (37), (38), it is by now standard to derive that for fixed $m \in \mathbb{N}$,

$$\lim_{\theta \downarrow 0} \sup_{\varepsilon \in (0,1)} \sup_{0 \leqslant \eta \leqslant \eta' + \theta \leqslant T} \mathbb{P}\left(|X^{\varepsilon,m}_{\eta} - X^{\varepsilon,m}_{\eta'}| \geqslant \delta\right) = 0,$$

which together with (47) yields that for any $T, \delta > 0$,

$$\lim_{\theta \downarrow 0} \sup_{\varepsilon \in (0,1)} \sup_{0 \leqslant \eta \leqslant \eta' + \theta \leqslant T} \mathbb{P}\left(|X^{\varepsilon}_{\eta} - X^{\varepsilon}_{\eta'}| \geqslant \delta\right) = 0.$$

Thus, by Aldous' criterion (see [24, p.356, Theorem 4.5]), the law $\mathbb{Q}^{\varepsilon}$ of $(X^{\varepsilon}_{\cdot}, L^{\alpha}_{\cdot})$, $\varepsilon \in (0,1)$ in $\mathbb{D} \times \mathbb{D}$ is tight. Let $\mathbb{Q}$ be any accumulation point of $(\mathbb{Q}^{\varepsilon})_{\varepsilon \in (0,1)}$. Without loss of generality, we assume that for some subsequence $\varepsilon_k \to 0$, $(\mathbb{Q}^k)_{k \in \mathbb{N}} := (\mathbb{Q}^{\varepsilon_k})_{k \in \mathbb{N}}$ weakly converges to $\mathbb{Q}$ as $k \to \infty$. By Skorokhod's representation theorem, there is a probability space $(\widetilde{\Omega}, \widetilde{\mathscr{F}}, \widetilde{\mathbb{P}})$ and $\mathbb{D} \times \mathbb{D}$-valued processes $(\widetilde{X}^k, \widetilde{L}^k)$ and $(\widetilde{X}, \widetilde{L})$ such that

$$(\widetilde{X}^k, \widetilde{L}^k) \to (\widetilde{X}, \widetilde{L}) \ \text{ in } \mathbb{D} \times \mathbb{D}, \ \ \widetilde{\mathbb{P}} - a.s.,$$

and

$$\widetilde{\mathbb{P}} \circ (\widetilde{X}^k, \widetilde{L}^k)^{-1} = \mathbb{Q}^k, \ \ \widetilde{\mathbb{P}} \circ (\widetilde{X}, \widetilde{L})^{-1} = \mathbb{Q}.$$

Moreover, $\widetilde{L}^k$ and $\widetilde{L}$ are still α-stable Lévy processes, and

$$\widetilde{X}^k_t = x + \int_0^t b_{\varepsilon_k}(\widetilde{X}^k_s)\mathrm{d}s + \int_0^t \sigma_{\varepsilon_k}(\widetilde{X}^k_s)\mathrm{d}\widetilde{L}^k_s.$$

By [24, Theorem 6.22, p.383] and taking limits, one sees that

$$\widetilde{X}_t = x + \int_0^t b(\widetilde{X}_s)\mathrm{d}s + \int_0^t \sigma(\widetilde{X}_s)\mathrm{d}\widetilde{L}_s,$$

and $(\widetilde{X}, \widetilde{L})$ is a weak solution of SDE (3). Finally, by the weak uniqueness of [9] we obtain the weak convergence (36).

Acknowledgments

The authors would like to thank Zimo Hao and Jian Wang for useful conversations about the heavy-tailed sampling, and the anonymous referees, AE and the Editor for their constructive comments.

Funding

This work is partially supported by NNSFC grants of China (Nos. 12131019, 11731009), and the German Research Foundation (DFG) through the Collaborative Research Centre(CRC) 1283 "Taming uncertainty and profiting from randomness and low regularity in analysis, stochastics and their applications".

References

[1] Applebaum, D. (2009). *Lévy processes and stochastic calculus.* Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge.

[2] Arnold, B. C. (2015). *Pareto Distributions*, 2nd ed. Chapman and Hall/CRC.

[3] Bakry, D., Gentil, I. and Ledoux, M. (2014). *Analysis and geometry of Markov diffusion operators.* Grundlehren der mathematischen Wissenschaften. Springer, Cham.

[4] Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Regular variation.* Cambridge University Press, Cambridge.

[5] Bubeck, S., Eldan, R. and Lehec, J. (2018). Sampling from a log-concave distribution with projected Langevin Monte Carlo. *Discrete Comput. Geom.* **4** 757-783.

[6] Carmona, P. (2007). Existence and uniqueness of an invariant measure for a chain of oscillators in contact with two heat baths. *Stochastic Process. Appl.* **8** 1076-1092.

[7] Chen, M. (2005). *Eigenvalues, inequalities, and ergodic theory.* Probability and its Applications (New York). Springer-Verlag London, Ltd., London.

[8] Cerrai, S. (2001). *Second order PDE's in finite and infinite dimension.* A probabilistic approach. Springer-Verlag, Berlin.

[9] Chen, Z. Q., Zhang, X. and Zhao, G. (2021). Supercritical SDEs driven by multiplicative stable-like Lévy processes. *Trans. Amer. Math. Soc.* **11** 1-33.

[10] Cheng, X., Chatterji, N. S., Abbasi-Yadkori, Y., Bartlett, P. L. and Jordan, M. I. (2020). Sharp convergence rates for Langevin dynamics in the nonconvex setting. arXiv: 1805.01648.

[11] Chung, K. and Zhao, Z. (1995). *From Brownian motion to Schrödinger's equation.* Grundlehren der mathematischen Wissenschaften. Springer-Verlag, Berlin.

[12] Dalalyan, A. S. and Karagulyan, A. (2019). User-friendly guarantees for the Langevin Monte Carlo with inaccurate gradient. *Stochastic Process. Appl.* **12** 5278-5311.

[13] Dalalyan, A. S. and Tsybakov, A. B. (2012). Sparse regression learning by aggregation and Langevin Monte-Carlo. *J. Comput. System Sci.* **5** 1423-1443.

[14] Da Prato, G. and Zabczyk, J. (1992). *Stochastic equations in infinite dimensions*, 2nd ed. Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge.

[15] Down, D., Meyn, S. P. and Tweedie, R. L. (1995). Exponential and uniform ergodicity of Markov processes. *Ann. Probab.* **4** 1671-1691.

[16] Durrett, R. (2010). *Probability: Theory and Examples*, 4th ed. Cambridge University Press.

[17] Embrechts, P., and Klüppelberg, C. and Mikosch, T. (1997). *Modelling extremal events: For insurance and finance.* Springer-Verlag, Berlin.

[18] Ethier, S.N. and Kurtz, T.G. (1986). *Markov Processes: Characterization and Convergence.* Wiley Interscience, New-Jersey.

[19] Foss, S., Korshunov, D. and Zachary, S. (2011). *An introduction to heavy-tailed and subexponential distributions.* Springer Series in Operations Research and Financial Engineering. Springer, New York.

[20] Girolami, M. and Calderhead, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2** 123-214.

[21] Goldys, B. and Maslowski, B. (2006). *Exponential ergodicity for stochastic reaction-diffusion equations.* Stochastic partial differential equations and applications-VII. Chapman & Hall/CRC, Boca Raton, FL.

[22] Gürbüzbalaban, M., Simsekli, U. and Zhu, L. (2021). The heavy-tail phenomenon in SGD. *International Conference on Machine Learning.* PMLR. pp. 3964-3975.

[23] Huang, L., Majka, M. and Wang, J. (2021). Approximation of heavy-tailed distributions via stable-driven SDEs. *Bernoulli.* **3** 2040-2068.

[24] Jacod, J. and Shiryaev, A. N. (2003). *Limit theorems for stochastic processes*, 2nd ed. Grundlehren der mathematischen Wissenschaften. Springer-Verlag, Berlin.

[25] Kulik, A. M. (2009). Exponential ergodicity of the solutions to SDE's with a jump noise. *Stochastic Process. Appl.* **2** 602-632.

[26] Kwaśnicki, M. (2017). Ten equivalent definitions of the fractional Laplace operator. *Fract. Calc. Appl. Anal.* **1** 7-51.

[27] Liang, M. and Wang, J. (2020). Gradient estimates and ergodicity for SDEs driven by multiplicative Lévy noises via coupling. *Stochastic Process. Appl.* **5** 3053-3094.

[28] Liang, M., Majka, M. and Wang, J. (2021). Exponential ergodicity for SDEs and McKean-Vlasov processes with Lévy noise. *Ann. Inst. Henri Poincaré Probab. Stat.* **3** 1665-1701.

[29] Luo, D. and Wang, J. (2019). Refined basic couplings and Wasserstein-type distances for SDEs with Lévy noises. *Stochastic Process. Appl.* **9** 3129-3173.

[30] Majka, M. B. (2017). Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes. *Stochastic Process. Appl.* **12** 4083-4125.

[31] Masuda, H. (2007). Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps. *Stochastic Process. Appl.* **1** 35-56.

[32] Mattingly, J. C., Stuart, A. M. and Higman, D. J. (2002). Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stoch. Proc. Appl.* **2** 185-232.

[33] Menozzi, S. and Zhang, X. (2022). Heat kernel of supercritical nonlocal operators with unbounded drifts. *J. Éc. polytech. Math.* **9** 537-579.

[34] Meyn, S. and Tweedie, R. L. (2009). *Markov chains and stochastic stability*, 2nd ed. Cambridge University Press, Cambridge.

[35] Nguyen, T., Simsekli, U. and Richard, D. (2019). Non-asymptotic analysis of Fractional Langevin Monte Carlo for non-convex optimization. *International Conference on Machine Learning.* PMLR. **97** 4810-4819.

[36] Revuz, D. and Yor, M. (1999). *Continuous martingales and Brownian motion.* Springer-Verlag, Berlin.

[37] Roberts, G. O. and Tweedie, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli.* **4** 341-363.

[38] Simsekli, U. (2017). Fractional langevin monte carlo: Exploring Lévy driven stochastic differential equations for markov chain monte carlo. *International Conference on Machine Learning.* PMLR. 3200-3209.

[39] Simsekli, U., Zhu, L, Teh, Y. W. and Gürbüzbalaban, M. (2020). Fractional underdamped langevin dynamics: Retargeting sgd with momentum under heavy-tailed gradient noise. *International Conference on Machine Learning.* PMLR. 8970-8980.

[40] Stein, E. M. (1970). *Singular integrals and differentiability properties of functions.* Princeton Mathematical Series, Princeton University Press, Princeton, N.J..

[41] Villani, C. (2008). *Optimal transport, old and new.* Springer-Verlag, Tokyo.

[42] Wang, F. (2006). *Functional inequalities Markov semigroups and spectral theory.* Elsevier.

[43] Wang, J. (2016). L^p-Wasserstein distance for stochastic differential equations driven by Lévy processes. *Bernoulli.* **3** 1598-1616.

[44] Xie, L. and Zhang, X. (2020). Ergodicity of stochastic differential equations with jumps and singular coefficients. *Ann. Inst. Henri Poincaré Probab. Stat.* **1** 175-229.

[45] Ye, N. and Zhu, Z. (2018). Stochastic fractional hamiltonian monte carlo. *Proceedings of the 27th International Joint Conference on Artificial Intelligence.* 3019-3025.