ERGODICITY OF SUPERCRITICAL SDES DRIVEN BY $\alpha$-STABLE PROCESSES AND HEAVY-TAILED SAMPLING

XIAOLONG ZHANG, XICHENG ZHANG

Abstract. Let $\alpha \in (0,2)$ and $d \in \mathbb{N}$. Consider the following stochastic differential equation (SDE) driven by $\alpha$-stable process in $\mathbb{R}^d$:

$$dX_t = b(X_t)dt + \sigma(X_t) dL_t^\alpha, \quad X_0 = x \in \mathbb{R}^d,$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ are locally $\gamma$-Hölder continuous with $\gamma \in (0 \vee (1 - \alpha)^+, 1]$, $L_t^\alpha$ is a $d$-dimensional rotationally invariant $\alpha$-stable process. Under some dissipative and non-degenerate assumptions on $b, \sigma$, we show the $V$-uniformly exponential ergodicity for the semigroup $P_t$ associated with $(X_t(x), t \geq 0)$. Our proofs are mainly based on the heat kernel estimates recently established in [25] through showing the strong Feller property and the irreducibility of $P_t$. It is interesting that when $\alpha$ goes to zero, the diffusion coefficient $\sigma$ can grow faster than drift $b$. As applications, we put forward a new heavy-tailed sampling scheme.

Keywords: $\alpha$-stable processes, Ergodicity, Strong Feller property, Irreducibility, Heavy-tailed distribution.

1. Introduction

Let $\pi(x) \propto e^{-U(x)}$ be a probability density function, where $U : \mathbb{R}^d \to \mathbb{R}$ is a potential function. The classical Langevin Monte Carlo (LMC) method is to use the Langevin diffusion processes to approximate the target distribution $\mu(dx) = \pi(x)dx$. More precisely, consider the following stochastic differential equations (SDEs):

$$dX_t = -\nabla U(X_t)dt + \sqrt{2}dB_t, \quad X_0 = x,$$

where $(B_t)_{t \geq 0}$ is a $d$-dimensional standard Brownian motion. Under some regularity assumptions on $U$, it is well known that the above SDE admits a unique solution $(X_t)_{t \geq 0}$ and $\mu$ is an invariant measure of the semigroup $P_t$ generated by $X_t(x)$. Moreover, under the following dissipativity assumption: for some $c_0, c_1 > 0$,

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

it is also well known that the law of $X_t$ exponentially converges to the stationary distribution $\mu$ in some sense as $t \to \infty$ (cf. [3, 26]), i.e., $P_t$ is exponentially ergodic. In particular, it provides a way to generate samples from $\mu$ by the Euler discretization for SDE (1.1) (cf. [3]). Nowadays, there are numerous works about the analysis of the performance of LMC algorithms (see [10, 11, 13] and references therein).

On the other hand, in statistics physics, the Langevin diffusion $X_t$ represents the position of a particle at time $t$ which is influenced by some random force, where the random force is usually the sum of many i.i.d. random pulses with finite variance. By the central limit theorem, the sum of the pulses converges to a Gaussian distribution. While, if the random pulses have infinite variance, then the sum of the pulses would converge to an $\alpha$-stable distribution, a typical class of heavy-tailed distribution. Moreover, it was already shown in [28] that for $d = 1$ and $U(x) = |x|^\beta$, the diffusion process defined by SDE (1.1) is exponentially ergodic if and only if $\beta \geq 1$. Therefore,
using continuous Langevin diffusion (1.1) to simulate the heavy-tailed distribution is not any more a good choice. Thus, instead of SDE (1.1), it is quite natural to consider the following SDE:

$$dX_t = b(X_t)dt + dL^\alpha_t, \quad \alpha \in (0, 2),$$

(1.2)

where \((L^\alpha_t)_{t \geq 0}\) is a \(d\)-dimensional rotationally invariant \(\alpha\)-stable process with generator \(\Delta^{\alpha/2}\), the usual fractional Laplacian, and

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

where \(d \geq 2\) and \(\Delta^{\alpha/2}\) is the Riesz potential (cf. [31, page117]). Formally, one sees that

$$\Delta^{\alpha/2} e^{-U} - \text{div}(be^{-U}) \equiv 0,$$

which implies that \(\mu\) is an invariant measure of SDE (1.2). Rigorous theoretical results for the exponential ergodicity of the above SDE (1.2) were given in [16, Theorem 1.1], and therein, some explicit conditions on \(U\) were provided. This approach called Fractional Langevin Monte Carlo was firstly introduced in [29, 32]. It has been demonstrated to be useful in modern machine learning both in optimization and heavy-tailed sampling [15, 30, 36].

Motivated by the above mentioned works, in this paper we consider the following SDE driven by multiplicative \(\alpha\)-stable noises:

$$dX_t = b(X_t)dt + \sigma(X_t) dL^\alpha_t, \quad X_0 = x,$$

(1.3)

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 paper are two-folds:

(i) Under some dissipativity, non-degeneracy and locally Hölder regularity assumptions on \(b\) and \(\sigma\), we establish the exponential ergodicity of SDE (1.3) for all \(\alpha \in (0, 2)\), which, in particular, covers the supercritical regime \(\alpha \in (0, 1)\).

(ii) For a large class of heavy-tailed distributions, we put forward a new ergodicity SDE driven by multiplicative \(\alpha\)-stable processes to simulate it. Compared with the recent work [16], our conditions on the potential functions are simpler and easier to realize in the computer.

Now we make the following assumptions on \(b\) and \(\sigma\):

\(\textbf{(H}_{\text{loc}})\) For any \(m \in \mathbb{N}\), there is a constant \(C_m \geq 1\) such that for all \(|x| \leq m\) and \(\xi \in \mathbb{R}^d\),

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

Moreover, there are \(\gamma \in ((1 - \alpha)^+, 1]\) and locally bounded measurable functions \(\ell_1\) and \(\ell_2\) on \(\mathbb{R}^d\) such that for all \(x \in \mathbb{R}^d\) and \(|h| \leq 1\),

$$|b(x+h) - b(x)| \leq \ell_1(h) |h|^\gamma, \quad |\sigma(x+h) - \sigma(x)| \leq \ell_2(h) |h|^\gamma.$$

(1.4)

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

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

where \(q\) is defined by (3.1) below.

\textbf{Remark 1.1.} Notice that (1.4) is equivalent to the locally \(\gamma\)-Hölder continuity of \(b\) and \(\sigma\). If \(b, \sigma\) are locally Lipschitz, i.e. \(\gamma = 1\) in (1.4), then one can take \(\varepsilon_0 = 0\) in (1.5). When \(\gamma < 1\), in order to adopt the stopping time to localize the coefficients, we need to use the pathwise uniqueness of strong solutions. Thus we must mollify the coefficients and require \(\varepsilon_0 > 0\) in (1.5) to have some uniform estimates for the approximation coefficients. Examples of \(b, \sigma\) satisfying \(\textbf{(H}_{\text{glo}}^{r,q})\) shall be provided below. In particular, \(b\) and \(\sigma\) are allowed to be polynomial growth. It is interesting that when \(\alpha \in (0, 1)\), \(\sigma\) can grow faster than \(b\) (see Example 1.3 below).
Under the above assumptions, it is well known that there is a unique weak solution to SDE (1.3) (see [7]). The semigroup associated with the Markov process $X_t(x)$ is then defined by

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

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

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

with

$$\mathcal{L}_\sigma f(x) = \frac{\alpha 2^{\alpha} \Gamma\left(\frac{d+\alpha}{2}\right)}{\Gamma\left(\frac{d}{2}\right) \Gamma\left(\frac{d+\alpha}{2}\right)} \int_{\mathbb{R}^d} [f(x + \sigma(x)z) + f(x - \sigma(x)z) - 2f(x)]dz/|z|^{d+\alpha}.$$  \hspace{1cm} (1.6)

In particular,

$$\Delta_\mathbb{R}^d f(x) = \frac{\alpha 2^{\alpha} \Gamma\left(\frac{d+\alpha}{2}\right)}{\Gamma\left(\frac{d}{2}\right) \Gamma\left(\frac{d+\alpha}{2}\right)} \int_{\mathbb{R}^d} [f(x + z) + f(x - z) - 2f(x)]dz/|z|^{d+\alpha}.$$  \hspace{1cm} (1.7)

We call SDE (1.3) being supercritical for $\alpha \in (0, 1)$ since in this case, from the view point of PDE, the drift term plays a dominant role compared with the diffusion term $\mathcal{L}_\sigma$ (see [7]).

Our main result of this paper is

**Theorem 1.2.** Let $r > -\alpha$ and $p \in ((-r) \vee 0, \alpha)$. Define $V_p(x) := 1 + |x|^p$. Suppose that $(H_{\text{loc}})$ and $(H_{\text{glob}}^q)$ hold. Then there is a unique invariant probability measure $\mu$ associated with $(P_t)_{t \geq 0}$. Moreover, we have

(i) If $r \geq 0$, then there are $\lambda, C > 0$ such that for all $t > 0$ and $x \in \mathbb{R}^d$,

$$\sup_{\|\varphi/V_p\| \leq 1} \|P_t\varphi(x) - \mu(\varphi)\| \leq Ce^{-\lambda t}V_p(x).$$

(ii) If $r > 0$, then there are $\lambda, C > 0$ such that for all $t > 0$,

$$\sup_{x \in \mathbb{R}^d} \|P_t(x) - \mu\|_{\text{Var}} \leq Ce^{-\lambda t}.$$

Below we provide two examples to illustrate our main result.

**Example 1.3.** The following example shows that $b$ and $\sigma$ can be polynomial decay and growth:

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

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

$$\langle x, b(x) \rangle + q|\sigma(x)||x|^{2-\alpha} = -|x|^2(1 + |x|^2)^{\beta/2 + q(1 + |x|^2)^{\alpha/2}|x|^{2-\alpha}} \leq -2^{\alpha+\beta}|x|^{2+\beta} + 2^{\alpha+\beta+q}\langle x, b(x) \rangle = -2^{\alpha+\beta+q}|x|^{2+\beta} + c_1.$$  \hspace{1cm} (1.8)

In particular, when $\alpha > 1$, $\beta$ can be less than $-1$ so that $b$ is polynomial decay. Moreover, for fixed $\beta > 0$, when $\alpha \downarrow 0$, $\gamma$ can go to infinity. In other words, $\sigma$ can grow faster than $b$. This is not surprise since one can image that for the SDE driven by compound Poisson processes (in some sense, corresponding to $\alpha = 0$), $\sigma$ can be arbitrarily growth.

**Example 1.4.** The following example shows that $b$ and $\sigma$ can even be exponentially growth:

$$b(x) = -xe^{|x|}, \quad \sigma(x) = e^{\frac{|x|^{1/\beta}}{\beta}}, \quad \beta \in (\alpha, 2).$$

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

$$|b(x + h) - b(x)| \leq (1 + |x|)e^{\frac{|x|+1}{\beta}}|h| =: \ell_1(x)|h|,$$

and

$$|\sigma(x + h) - \sigma(x)| \leq e^{(|x|+1)/\beta}|h| =: \ell_2(x)|h|.$$  \hspace{1cm} (1.7)

Thus for any $|x| > 1$ and $t > 1$, by Young’s inequality, one can choose $\varepsilon_0$ small enough so that

$$\langle x, b(x) \rangle + \varepsilon_0\ell_1(x)|x| + q(\|\sigma(x)\| + \varepsilon_0\ell_2(x))\alpha|x|^{2-\alpha}$$
When the driven noise is Brownian motion, the ergodicity of SDEs (1.3) has been studied extensively in the literature [2, 6, 12, 33, 35], etc., and the dissipativity condition (1.5) usually takes the following form (for example, see [35, (7.1)]):

\[
2 \langle x, b(x) \rangle + \| \sigma(x) \|^2 \leq -c_0 |x|^2 + c_1.
\]

In recent years, there has also been a great interest in the study of the ergodicity for SDEs driven by Lévy processes, see [18, 20, 34, 35] and the references therein. To the best of the authors’ knowledge, there exist at least three well developed methods for studying the exponential ergodicity of a Markov process defined through an SDE: Coupling method (cf. [16, 18, 19, 20, 21, 22]), Functional inequality method (see [2, 5, 33]), Meyn-Tweedie’s minorization condition (cf. [14, 23, 24, 26]). Each method has its own merits and may be adapted to different situations. For the coupling method, in order to construct a successful coupling it is usually assumed that the following one-sided monotone condition holds (see [21]): for some \( R > 0 \),

\[
\langle x - y, \nabla U(x) - \nabla U(y) \rangle \geq \begin{cases} -L|x - y|^2, & |x - y| < R, \\ K|x - y|^2, & |x - y| \geq R. \end{cases}
\]

While, for the functional inequality method, it is usually assumed that the potential function \( U \) is at least \( C^2 \)-smooth (see [2, 33]). The obvious advantage of this method is that the quantitative convergence rate can be calculated from the parameters. Finally, Meyn-Tweedie’s approach is based on the Lyapunov estimates together with verifying the strong Feller property and irreducibility for the associated semigroup.

To prove our main theorem, we adopt the Meyn-Tweedie method. As we know, the strong Feller property reflects the regularization effects of the noise, which 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, e.g., the gradient estimates, Harnack’s inequalities, etc. For the irreducibility, when the driving noise is Brownian motion, there are several ways for choices: by solving a control problem [24], by Girsanov’s transformation [35], by the classical supported theorem of SDE [4]. However, when the driven noise is an \( \alpha \)-stable noise, the above method for irreducibility does not work any more. In [35], we directly verify the positivity of the Dirichlet heat kernel by localization method. Therein, it is restricted to \( \alpha \in (1, 2) \) since in this case, the heat kernel of the operator (1.6) is comparable with the one of the fractional Laplacian \( \Delta^{\frac{\alpha}{2}} \). In a recent work [25], when the conditions in \( (H_{\text{loc}}) \) holds globally, i.e., the constant \( C_m \) does not depend on \( m \) and \( t_1(x), t_2(x) \) are constants, we obtain the following two-sided estimates for the density of SDE (1.3): for any \( T > 0 \) and some \( C > 1 \), and for all \( t \in (0, T] \) and \( x, y \in \mathbb{R}^d \),

\[
C^{-1}t^{\frac{1}{2} + \frac{d}{2} - \alpha} \leq p(t, x, y) \leq Ct^{\frac{1}{2} + \frac{d}{2} - \alpha},
\]

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

\[
\dot{\theta}_t(x) = b \ast \phi_{t/\alpha}(\theta_t(x)), \quad \theta_0(x) = x,
\]

where \( \{\phi_\tau\}_{\tau > 0} \) is a family of standard mollifiers. The above estimates still allow us to show the irreducibility of \( F_t \) by localization. For the existence of invariant measures, we use the standard Krylov-Bogoliubov’s theorem by showing some Lyapunov’s type estimate.

The paper is organized as follows. In Section 2, we recall some basic notions about the ergodicity of Markov processes. These notions are standard in literature (see [12, 27]). 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 works of [9, 35]. In Section 3, we prove our main result Theorem 1.2. For the existence of invariant measure, by the classical Krylov-Bogoliubov’s method, we show the Lyapunov type estimates for \( X_t(x) \). For the uniqueness of invariant measure as well as the
exponential ergodicity, we show the strong Feller property and the irreducibility of $P_t$ by the heat kernel estimates and suitable localization and smoothing arguments. In Section 4, we present an application of our main result to the heavy-tailed sampling. In Appendix, the proof of weak convergence for mollifying SDEs is provided for readers’ convenience.

2. Preliminaries

In this section we prepare some notions and criterion 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 càdlàg 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:

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

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

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

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

(iii) For any $s < t$ and $A \in \mathcal{B}(\mathbb{D})$, it holds that

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

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

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

where $\mathbb{E}_x$ stands for the expectation with respect to $\mathbb{P}_x$.

The following notions are standard and can be found in [12, 26].

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

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

- We call $(P_t)_{t \geq 0}$ ergodic, if there exists a unique invariant probability measure $\mu$ of $(P_t)_{t \geq 0}$, equivalently, for any $x \in \mathbb{R}^N$ and $f \in \mathcal{B}_b(\mathbb{R}^N)$,

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

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

$$\sup_{\|\phi\|_V \leq 1} |P_t \phi(x) - \mu(\phi)| \leq C e^{-\lambda t} V(x), \quad t \geq 0, \ x \in \mathbb{R}^N.$$

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

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

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

The following general result is taken from [14, Theorem 2.5].

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

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

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

$$\mathbb{E}_x |X_t|^p \leq C_3, \quad t \geq T_0, \ x \in \mathbb{R}^N,$$

then $(P_t)_{t \geq 0}$ 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 to 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 \( 0 < \delta < T < \infty \),

\[
\sup_{t \in [\delta, T]} \sup_{r > 0} \rho(t, r) < \infty. \tag{2.2}
\]

In the following we make the following assumptions about the Markov process \( \mathcal{M} \):

(A1) 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

\[
\mathbb{P}_x(X_t \in A) = \int_A p(t, x, y)dy, \quad \forall A \in \mathcal{B}(\mathbb{R}^N), \ t > 0, \ x \in \mathbb{R}^N. \tag{2.3}
\]

Moreover, \( t \mapsto X_t \) is stochastically continuous, equivalently,

\[
\mathbb{P}_x(X_t \neq X_{t-}) = 0, \quad \forall t > 0, \ x \in \mathbb{R}^N. \tag{2.4}
\]

(A2) We suppose that for any \( T > 0 \), there are constants \( \lambda_0, C_0 \geq 1 \) such that for any \( t \in (0, T] \) and \( x \in \mathbb{R}^N \), and for Lebesgue almost all \( y \in \mathbb{R}^N \),

\[
C_0^{-1}\rho(t, \lambda_0 \Gamma_t(x, y)) \leq p(t, x, y) \leq C_0 \rho(t, \Gamma_t(x, y)/\lambda_0), \tag{2.5}
\]

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 (A2), 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 \( \alpha \)-stable case, if \( \alpha \in [1, 2) \), one takes \( \rho(t, r) = t^{1/\alpha}r^{-d-\alpha} \) 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 (1.9) 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 the following general result.

Theorem 2.3. Let \( D_0 \in D \) be two connected domains of \( \mathbb{R}^N \) so that \( D \) has Lebesgue-zero measure boundary. Suppose that (A1) and (A2) hold, and the above \( \Gamma_t(x, y) \) and \( \rho(t, r) \) satisfy the following conditions: 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 \( x' \notin D \),

\[
\Gamma_t(x, y) \leq \ell_0(t), \quad \Gamma_t(x', y) \geq \ell_1(t), \tag{2.6}
\]

and for the \( C_0, \lambda_0 \) in (2.5), \( t \mapsto \rho(t, \ell_1(t)/\lambda_0) \) is increasing on \( (0, t_0] \), and

\[
C_0^{-1}\rho(t, \lambda_0 \ell_0(t)) \geq 2C_0 \rho(t, \ell_1(t)/\lambda_0). \tag{2.7}
\]

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

\[
\inf_{x \in B_{r_2}(x_0)} \mathbb{P}_x \left( X_t \in B_{r_2}(y_0); t < \tau_D \right) \geq c_0 > 0,
\]

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 \( \delta_0 \) be as in the assumptions. We divide the proofs into two steps.

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

\[
\inf_{x \in B_{2\delta}(y_0)} \mathbb{P}_x \left( X_t \in B_{\delta}(y_0); t < \tau_D \right) \geq C_0 \rho(t, \ell_1(t)/\lambda_0) |B_{\delta}(y_0)| > 0. \tag{2.8}
\]
Fix \( s \in (0, t) \) and \( n \in \mathbb{N} \). We define a stopping time \( T_n \) as follows:
\[
T_n = \begin{cases} 
  ks2^{-n} & \text{if } (k-1)s2^{-n} \leq \tau_D < ks2^{-n}, \\
  \infty & \text{if } \tau_D \geq s. 
\end{cases}
\]

Clearly, on \( \{\tau_D < s\} \),
\[
s \geq T_n \downarrow \tau_D \text{ as } n \to \infty.
\]

For any bounded \( A \in \mathcal{F}(D) \), by the Markov property and (2.3), we have
\[
\mathbb{P}_x(X_t \in A; \tau_D < s) = \sum_{k=1}^{2^n} \mathbb{P}_x(X_t \in A; (k-1)s2^{-n} \leq \tau_D < ks2^{-n})
\]
\[
= \sum_{k=1}^{2^n} \mathbb{P}_x(X_t \in A; T_n = ks2^{-n})
\]
\[
= \sum_{k=1}^{2^n} \mathbb{P}_x\left(\mathbb{P}_{X_{T_n}}(X_{t-T_n} \in A); T_n = ks2^{-n}\right)
\]
\[
= \sum_{k=1}^{2^n} \mathbb{E}_x \left(\int_A p(t-T_n, X_{T_n}, y)dy; T_n = ks2^{-n}\right)
\]
\[
= \mathbb{E}_x \left(\int_A p(t-T_n, X_{T_n}, y)dy; \tau_D < s\right)
\]
\[
\stackrel{(2.5)}{\leq} C_0 \mathbb{E}_x \left(\int_A \rho(t-T_n, \Gamma_{t-T_n}(X_{T_n}, y)/\lambda_0)dy; \tau_D < s\right).
\]

Letting \( n \to \infty \) and by (2.2) and the dominated convergence theorem, we get
\[
\mathbb{P}_x(X_t \in A; \tau_D < s) \leq C_0 \mathbb{E}_x \left(\int_A \rho(t-\tau_D, \Gamma_{t-\tau_D}(X_{\tau_D}, y)/\lambda_0)dy; \tau_D < s\right).
\]

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 (2.6) we have
\[
\rho(t-\tau_D, \Gamma_{t-\tau_D}(X_{\tau_D}, y)/\lambda_0) \leq \rho(t-\tau_D, \ell_1(t-\tau_D)/\lambda_0) \leq \rho(t, \ell_1(t)/\lambda_0).
\]

Hence, for any \( s < t \),
\[
\mathbb{P}_x\left(X_t \in B_{\delta}(y_0); \tau_D < s\right) \leq C_0 \int_{B_{\delta}(y_0)} \rho(t, \ell_1(t)/\lambda_0)dy.
\]

On the other hand, by (2.4) and (2.5), we have
\[
\mathbb{P}_x(\tau_D = t) = \mathbb{P}_x(\tau_D = t, X_t \in \partial D) + \mathbb{P}_x(\tau_D = t, X_t \notin \partial D, X_s \in D, s < \tau_D)
\]
\[
\leq \mathbb{P}_x(X_t \in \partial D) + \mathbb{P}_x(X_t \notin \partial D, X_s \in D, s < \tau_D) = 0.
\]

Thus, by the lower bound in (2.5) and the above estimates, we arrive at
\[
\mathbb{P}_x\left(X_t \in B_{\delta}(y_0); t < \tau_D\right) = \mathbb{P}_x\left(X_t \in B_{\delta}(y_0)\right) - \mathbb{P}_x\left(X_t \in B_{\delta}(y_0); \tau_D < t\right)
\]
\[
= \int_{B_{\delta}(y_0)} \rho(t, x, y)dy - \lim_{s \uparrow t} \mathbb{P}_x\left(X_t \in B_{\delta}(y_0); \tau_D < s\right)
\]
\[
\geq \int_{B_{\delta}(y_0)} \left( C_0^{-1} \rho(t, \lambda_0 \Gamma_1(x, y)) - C_0 \rho(t, \ell_1(t)/\lambda_0) \right)dy.
\]

For \( x \in B_{2\delta}(y_0), y \in B_{\delta}(y_0) \), since \( r \to \rho(t, r) \) is decreasing, by (2.6), we obtain
\[
\mathbb{P}_x\left(X_t \in B_{\delta}(y_0); t < \tau_D\right) \geq \left( C_0^{-1} \rho(t, \lambda_0 \ell_0(t)) - C_0 \rho(t, \ell_1(t)/\lambda_0) \right)\mathbb{P}_x|B_{\delta}(y_0)|
\]
\[
\stackrel{(2.7)}{\geq} C_0 \rho(t, \ell_1(t)/\lambda_0)|B_{\delta}(y_0)|.
\]
Thus we get (2.8).

(ii) By (2.3), there is a density \( p^D(t, x, y) \) so that for each \( A \in \mathcal{B}(D) \),

\[
\int_A p^D(t, x, y)dy = P_x(X_t \in A; t < \tau_D).
\]

Moreover, by the Markov property, we have for each \( 0 < s < t \),

\[
\int_A p^D(t, x, y)dy = \int_D p^D(s, x, z)dz \int_A p^D(t - s, z, y)dy.
\]

Indeed, by \( s + \tau_D \circ \theta_s = \tau_D \) on \( \{ s < \tau_D \} \) and (2.1),

\[
P_x(X_t \in A; t < \tau_D) = P_x \left( \mathbb{P}_x \left( X_t \in A; t < \tau_D \mid \mathcal{F}_s \right) \right)
= \mathbb{P}_x \left( \mathbb{P}_x \left( X_{t-s} \circ \theta_s \in A; t < \tau_D \circ \theta_s, s < \tau_D \mid \mathcal{F}_s \right) \right)
= \mathbb{P}_x \left( \mathbb{P}_x \left( X_{t-s} \in A; t - s < \tau_D; s < \tau_D \right) \right)
= \int_D p^D(s, x, z) \int_A p^D(t - s, z, y)dydz.
\]

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 \( \delta := (\text{dist}(D_0, \partial D) \land \delta_0)/4 \).

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

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

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),

\[
\inf_{y_{j-1} \in B_\delta(x_{j-1})} \int_{B_\delta(x_{j})} p^D(t, x_{j-1}, y_j)dy_j \geq \inf_{y_{j-1} \in B_{2\delta}(x_{j})} \int_{B_\delta(x_{j})} p^D(t, x_{j-1}, y_j)dy_j \\
\geq C_0 \rho \bigg( \frac{x_j}{n}, \ell_1 \bigg( \frac{x_j}{n} \bigg) / \lambda_0 \bigg) |B_\delta(y_0)| > 0.
\]

Hence, for all \( x \in B_\delta(x_0) \), by (2.9),

\[
\int_{B_{2\delta}(y_0)} p^D(t, x, y)dy = \int_D p^D(t, x, y_1)dy_1 \cdots \int_D p^D(t, y_{n-1}, y_n)dy_n \int_{B_{2\delta}(y_0)} p^D(t, y_n, y)dy \\
\geq \int_{B_\delta(x_1)} p^D(t, x, y_1)dy_1 \cdots \int_{B_\delta(x_{n+1})} p^D(t, y_{n-1}, y_n)dy_n \int_{B_\delta(x_{n+1})} p^D(t, y_n, y)dy \\
\geq \left( C_0 \rho \bigg( \frac{x_1}{n}, \ell_1 \bigg( \frac{x_1}{n} \bigg) / \lambda_0 \bigg) |B_\delta(y_0)| \right)^{n+1}.
\]

Thus,

\[
\inf_{x \in B_\delta(x_0)} \mathbb{P}_x \left( X_t \in B_{2\delta}(y_0); t < \tau_D \right) \geq \left( C_0 \rho \bigg( \frac{x_1}{n}, \ell_1 \bigg( \frac{x_1}{n} \bigg) / \lambda_0 \bigg) |B_\delta(y_0)| \right)^{n+1} > 0.
\]

For general \( r_1 \in (0, \text{dist}(D_0, D)) \), it follows 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

\( r > -\alpha, \ p \in ((-r) \lor 0, \alpha), \)

and define

\[
V_p(x) := (1 + |x|^2)^{p/2}, \quad q := \frac{2^{4+2\alpha} \alpha \Gamma \left( \frac{d+\alpha}{2} \right)}{\Gamma \left( \frac{d-\alpha}{2} \right) \Gamma \left( \frac{d+\alpha}{2} \right)} \left( \frac{\alpha}{2 - \alpha} + \frac{\alpha}{\alpha - p} \right).
\]

First of all, we show that \( V_p \) is a Lyapunov function of the operator \( \mathcal{L} \).

\[ \text{PROOF} \]

\[ \text{□} \]
Lemma 3.1. Suppose that $b, \sigma$ are locally bounded and for some $c_0, c_1 > 0$,
\[ q\|\sigma(x)\|^{\alpha}|x|^{2-\alpha} + (x, b(x)) \leq -c_0|x|^{2+\epsilon} + c_1, \quad |x| > 1. \quad (3.2) \]
Then there are $\kappa_0, \kappa_1 > 0$ such that for all $x \in \mathbb{R}^d$,
\[ \mathcal{L}\mathcal{V}_p(x) = \mathcal{L}_\sigma \mathcal{V}_p(x) + (b(x), \nabla \mathcal{V}_p(x)) \leq -\kappa_0 \mathcal{V}_p(x)^{1+\frac{\alpha}{2}} + \kappa_1. \quad (3.3) \]

Proof. It suffices to prove (3.3) for $|x| > 1$. By (1.7), we make the following decomposition:
\[ c_{d,\alpha}^{-1} \mathcal{L}_\sigma \mathcal{V}_p(x) = \int_{|z| \leq \frac{|x|}{2\|\sigma(x)\|}} \left[ \mathcal{V}_p(x + \sigma(x)z) + \mathcal{V}_p(x - \sigma(x)z) - 2\mathcal{V}_p(x) \right] \frac{dz}{|z|^{d+\alpha}} \]
\[ + \int_{|z| > \frac{|x|}{2\|\sigma(x)\|}} \left[ \mathcal{V}_p(x + \sigma(x)z) + \mathcal{V}_p(x - \sigma(x)z) - 2\mathcal{V}_p(x) \right] \frac{dz}{|z|^{d+\alpha}} =: I_1 + I_2, \]
where $c_{d,\alpha} := \frac{\alpha^2 \Gamma\left(\frac{d+\alpha}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}$. For $I_1$, denoting by $A(x) := (1 + |x|^2)\mathcal{I} - (2-p)x \otimes x$, we have
\[ \nabla^2 \mathcal{V}_p(x) = p(1 + |x|^2)\frac{\alpha}{2} - 2A(x). \]
Noting that for any $\xi \in \mathbb{R}^d$,
\[ |\langle \xi, A(x)\xi \rangle| \leq (1 + |x|^2)|\xi|^2 + (2-p)|\langle \xi, x \rangle|^2 \leq 3(1 + |x|^2)|\xi|^2. \]
By Taylor’s expansion, for $|z| \leq \frac{|x|}{2\|\sigma(x)\|}$, we have
\[ |\mathcal{V}_p(x + \sigma(x)z) + \mathcal{V}_p(x - \sigma(x)z) - 2\mathcal{V}_p(x)| \leq 3p\|\sigma(x)\|^2 \int_0^1 \int_{|s| < \frac{|x|}{2\|\sigma(x)\|}} \left(1 + |x + s\sigma(x)z|^2\right)^{\frac{\alpha}{2}-1} |z|^{2-\alpha} dz \]
\[ \leq 3p\|\sigma(x)\|^2 |z|^2 \left(1 + \frac{|x|^2}{4}\right)^{\frac{\alpha}{2}-1}. \]
Hence, for $S_d := 2\pi^{\frac{d}{2}}/\Gamma\left(\frac{d}{2}\right)$,
\[ I_1 \leq 3p\|\sigma(x)\|^2 \left(1 + \frac{|x|^2}{4}\right)^{\frac{\alpha}{2}-1} \int_{|z| \leq \frac{|x|}{2\|\sigma(x)\|}} |z|^{2-\alpha} dz \]
\[ = 3pS_d\|\sigma(x)\|^2 \left(1 + \frac{|x|^2}{4}\right)^{\frac{\alpha}{2}-1} \int_{\frac{|x|}{2\|\sigma(x)\|}}^{\frac{|x|}{\|\sigma(x)\|}} r^{1-\alpha} dr \]
\[ \leq \frac{p2^{2+\alpha}S_d}{2-\alpha} \|\sigma(x)\|^\alpha |x|^{p-\alpha}. \]
For $I_2$, noting that
\[ |(1 + |x + z|^2)^\frac{\alpha}{2} - (1 + |x|^2)^\frac{\alpha}{2}| \leq ||x + z|^2 - |x|^2|^\frac{\alpha}{2} \leq 2^{\frac{\alpha}{2}} |x|^\frac{\alpha}{2} |z|^\frac{\alpha}{2} + |z|^p, \]
we have
\[ I_2 \leq 2 \int_{|z| > \frac{|x|}{2\|\sigma(x)\|}} \left[ 2^{\frac{\alpha}{2}} |x|^\frac{\alpha}{2} |\sigma(x)z|^\frac{\alpha}{2} + |\sigma(x)z|^p \right] \frac{dz}{|z|^{d+\alpha}} \leq \frac{2^{3+\alpha}S_d}{\alpha - p} \|\sigma(x)\|^\alpha |x|^{p-\alpha}. \]
Combining the above calculations, we get for $|x| > 1$,
\[ \mathcal{L}_\sigma \mathcal{V}_p(x) \leq \frac{2^{3+\alpha}}{\alpha - p} \|\sigma(x)\|^\alpha |x|^{p-\alpha} \]
\[ \leq p2^{\frac{\alpha}{2}-1} q\|\sigma(x)\|^\alpha |x|^{p-\alpha} \]
Thus, then by (3.2) and $\nabla \mathcal{V}_p(x) = px(1 + |x|^2)^{\frac{\alpha}{2}-1}$, we have
\[ \mathcal{L}_\sigma \mathcal{V}_p(x) + (b(x), \nabla \mathcal{V}_p(x)) \leq p2^{\frac{\alpha}{2}-1} q\|\sigma(x)\|^\alpha |x|^{p-\alpha} + p(1 + |x|^2)^{\frac{\alpha}{2}-1} \langle x, b(x) \rangle \]
\[ \leq p(1 + |x|^2)^{\frac{\alpha}{2}-1} \left( q\|\sigma(x)\|^\alpha |x|^{2-\alpha} + \langle x, b(x) \rangle \right). \]
of Lemma 3.1, there exists a constant $C > 0$ such that in the second inequality we use $(1 + |x|^2)^{\frac{p}{2}} \leq 2^{1-\frac{p}{2}}|x|^{2-p}$ with $|x| > 1$, and $\kappa_0 = pc_0$. □

As a consequence of the above Lyapunov-type estimate, we have the following result (see [35, Lemma 7.1]).

**Lemma 3.2.** Let $X_t(x)$ be any weak solution of SDE (1.3) 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 $t > 0$,

$$
\left[ \mathbb{E} \left( \sup_{s \in [0, t]} V_p(X_s(x))^{\frac{2}{p}} \right) \right]^2 + \mathbb{E} \left( \int_0^t V_p(X_s(x))^{1+\frac{p}{2}} ds \right) \leq C V_p(x) + t,
$$

and for the $\kappa_0$ in (3.3),

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

**Proof.** Let $N(t, dz)$ be the Poisson random measure associated with $L^p$, i.e.,

$$(t, \Gamma) := \sum_{s \in (0, t]} 1_{\Gamma} (L_s^p - L_{s-}^p), \quad \Gamma \in \mathcal{B}(\mathbb{R}^d).$$

Let $\tilde{N}(t, dz) := N(t, dz) - \sigma(z)^{-d-\alpha} dz$ be the compensated Poisson random measure. By Lévy-Itô’s decomposition, one can write

$$L^p_t = \int_{|z|<1} z \tilde{N}(t, dz) + \int_{|z|\geq 1} z N(t, dz).$$

Thus SDE (1.3) can be written as

$$dX_t = b(X_t)dt + \int_{|z|<1} \sigma(X_{t-})z \tilde{N}(dt, dz) + \int_{|z|\geq 1} \sigma(X_{t-})z N(dt, dz).$$

By Itô’s formula (see [1, Theorem 4.4.7]) and (3.3), we have

$$V_p(X_t) = V_p(x) + \int_0^t \left[ (\nabla V_p(X_s), b(X_s)) + \mathcal{L} V_p(X_s) \right] ds + M_t$$

$$\leq V_p(x) + \int_0^t \left[ -\kappa_0 V_p(X_s)^{1+\frac{p}{2}} + \kappa_1 \right] ds + M_t,$$

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

$$\mathbb{E} V_p(X_{t \land \tau_n}) + \kappa_0 \mathbb{E} \int_0^{t \land \tau_n} V_p(X_s)^{1+\frac{p}{2}} ds \leq V_p(x) + \kappa_1 t.$$

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

$$\mathbb{E} V_p(X_t) + \kappa_0 \mathbb{E} \left( \int_0^t V_p(X_s)^{1+\frac{p}{2}} ds \right) \leq V_p(x) + \kappa_1 t.$$

Moreover, starting from (3.7), by stochastic Gronwall’s inequality ([35, Lemma 3.7]), we also have

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

On the other hand, starting from (3.6), as above, we have

$$\mathbb{E} V_p(X_t) = V_p(x) + \int_0^t \mathbb{E} \left[ (\nabla V_p(X_s), b(X_s)) + \mathcal{L} V_p(X_s) \right] ds.$$
For $r \geq 0$, by (3.3) and Jensen’s inequality, we have

$$\frac{d}{dt} \mathbb{E} p(X_t) \leq -\kappa_0 \mathbb{E} \left[ V_p(X_t)^{1+\frac{p}{r}} \right] + \kappa_1 \leq -\kappa_0 \left[ \mathbb{E} V_p(X_t) \right]^{1+\frac{p}{r}} + \kappa_1.$$  

Let $f(t) := \mathbb{E} p(X_t)$. The above inequality then reads

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

If $r = 0$, then

$$\left( e^{\kappa_0 t} f(t) \right)' \leq \kappa_1 e^{\kappa_0 t} \Rightarrow f(t) \leq e^{\kappa_0 (s-t)} f(s) + \frac{\kappa_1}{\kappa_0} (1 - e^{\kappa_0 (s-t)}), \quad t > 0. \tag{3.8}$$

If $r > 0$, then by Young’s inequality,

$$\left( t^{\frac{2p}{r}} f(t) \right)' \leq \frac{2p}{r} t^{\frac{2p}{r}-1} f(t) - \kappa_0 t^{\frac{2p}{r}} f(t)^{1+\frac{p}{r}} + \kappa_1 t^{\frac{2p}{r}} \leq C_t^{\frac{p}{r}} - 1 + \kappa_1 t^{\frac{2p}{r}}.$$  

Integrating both sides from 0 to $t$, we obtain

$$t^{\frac{2p}{r}} f(t) \leq \frac{C_0 t^{\frac{p}{r}}} {p} + \frac{r \kappa_1}{r+2p} t^{\frac{2p}{r}+1} \Rightarrow f(t) \leq \frac{C_0 t^{\frac{p}{r}}} {p} + \frac{r \kappa_1}{r+2p},$$

which together with (3.8) yields (3.5). □

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

$$\phi_\varepsilon(x) := \varepsilon^{-d} \phi(\varepsilon^{-1} x), \quad x \in \mathbb{R}^d, \quad \varepsilon \in (0,1),$$

and

$$b_\varepsilon(x) := b * \phi_\varepsilon(x), \quad \sigma_\varepsilon(x) := \sigma * \phi_\varepsilon(x).$$

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

$$|b_\varepsilon(x) - b_\varepsilon(y)| + ||\sigma_\varepsilon(x) - \sigma_\varepsilon(y)|| \leq C_{m+1} |x-y|^{\gamma}, \quad \forall x, y \in B_m, \tag{3.9}$$

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

$$C_m^{-1} |\xi| \leq |\sigma_\varepsilon(x)\xi| \leq C_m |\xi|, \quad \forall x \in B_m, \quad \xi \in \mathbb{R}^d. \tag{3.10}$$

Moreover, note that by (1.4),

$$|b_\varepsilon(x) - b(x)| \leq \int_{\mathbb{R}^d} |b(x-y) - b(x)| \rho_\varepsilon(y) dy \leq \varepsilon^\gamma \ell_1(x),$$

and

$$|\sigma_\varepsilon(x) - \sigma(x)| \leq \int_{\mathbb{R}^d} |\sigma(x-y) - \sigma(x)| \rho_\varepsilon(y) dy \leq \varepsilon^\gamma \ell_2(x).$$

Hence, by (H_{loc}^\gamma), for any $\varepsilon \in (0,\varepsilon_0^{1/\gamma}),$

$$\langle x, b_\varepsilon(x) \rangle + q(|\sigma_\varepsilon(x)|^\alpha |x|^{2-\alpha} \leq \langle x, b(x) \rangle + \varepsilon^\gamma |x| \ell_1(x) + q(||\sigma(x)|| + \ell_2(x) \varepsilon^\gamma) \alpha |x|^{2-\alpha}$$

$$\leq \langle x, b(x) \rangle + \varepsilon_0 |x| \ell_1(x) + q(||\sigma(x)|| + \ell_2(x) \varepsilon_0) \alpha |x|^{2-\alpha}$$

$$\leq -c_0 |x|^{2+r} + c_1. \tag{3.11}$$

Now, consider the following approximation SDE:

$$dX^\varepsilon_t = b_\varepsilon(X^\varepsilon_t) dt + \sigma_\varepsilon(X^\varepsilon_t) dL^\varepsilon_t, \quad X^\varepsilon_0 = x.$$  

Since the coefficients are smooth, but, may be not bounded, by (3.11) and [7, Theorem 1.1], there is a unique strong solution $X^\varepsilon_t(x)$ to the above SDE. Moreover, for any $f \in C_b(\mathbb{R}^d)$, we have

$$\lim_{\varepsilon \to 0} \mathbb{E} f(X^\varepsilon_t(x)) = \mathbb{E} f(X_t(x)), \tag{3.12}$$

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

Let $\chi$ be a cutoff function with

$$\chi(x) = 1, \quad |x| \leq \frac{1}{2}, \quad \chi(x) = 0, \quad |x| > 1.$$
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 (3.9) and (3.10), it is easy to see that $b_m^\varepsilon$ and $\sigma_m^\varepsilon$ satisfy the following global assumptions:
\[
|b_m^\varepsilon(x) - b_m^\varepsilon(y)| + \|\sigma_m^\varepsilon(x) - \sigma_m^\varepsilon(y)\| \leq C_m|x - y|^\gamma, \quad \forall x, y \in \mathbb{R}^d,
\]
and
\[
C_m^{-1} |\xi| \leq |\sigma_m^\varepsilon(x)| \leq C_m |\xi|, \quad \forall x, \xi \in \mathbb{R}^d,
\]
where the constant does not depend on $\varepsilon$. Let $\theta_\varepsilon(x)$ solve the following ODE:
\[
\dot{\theta}_\varepsilon(x) = (b_m^\varepsilon * \phi_1/\varepsilon)(\theta_\varepsilon(x)), \quad \theta_\varepsilon(0) = x.
\]
We also consider the following SDE with cutoff coefficients:
\[
dX_t^{\varepsilon,m} = b_m^\varepsilon(X_t^{\varepsilon,m})dt + \sigma_m^\varepsilon(X_t^{\varepsilon,m})dL_t^\varepsilon, \quad X_0^{\varepsilon,m} = x.
\]
Let $X_t^{\varepsilon,m}(x)$ be the unique strong solution of the above SDE. By [25, Theorem 1.1], $X_t^{\varepsilon,m}(x)$ admits a density $p_m^\varepsilon(t,x,y)$ with the following estimates: For any $T > 0$ and $m \in \mathbb{N}$, there is a constant $C_0 \geq 1$ depending on $m$, but independent of $\varepsilon \in (0, \varepsilon_1)$, such that for all $t \in (0, T]$ and $x, y \in \mathbb{R}^d$,
\[
p_m^\varepsilon(t,x,y) = \mathcal{C}_0 (t\beta \wedge |x - \theta_\varepsilon(t)|)^{-d-\alpha}.
\]
For any $m \in \mathbb{N}$ and $x \in B_m$, define the exit time of $X_t^{\varepsilon,m}(x)$ from $B_m$ by
\[
\tau_B^{\varepsilon,x} := \inf \{ t > 0 : X_t^{\varepsilon,m}(x) \notin B_m \}.
\]
By the uniqueness of strong solution, we have
\[
X_t^{\varepsilon}(x) = X_t^{\varepsilon,m}(x), \quad t < \tau_B^{\varepsilon,x}.
\]
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 a standard monotonic argument, it suffices to show that for any $t > 0$,
\[
\lim_{x \to y} \sup_{f \in C_b(\mathbb{R}^d), \|f\|_{\infty} \leq 1} |\mathbb{E}(f(X_t(x)) - f(X_t(y)))| = 0.
\]
Since for any $f \in C_b(\mathbb{R}^d)$,
\[
\lim_{\varepsilon \to 0} \mathbb{E}(f(X_t^{\varepsilon}(x)) - f(X_t^{\varepsilon}(y))) = \mathbb{E}(f(X_t(x)) - f(X_t(y))),
\]
we only need to prove that
\[
\lim_{x \to y} \sup_{\varepsilon \in (0, 1)} \sup_{f \in C_b(\mathbb{R}^d), \|f\|_{\infty} \leq 1} |\mathbb{E}(f(X_t^{\varepsilon}(x)) - f(X_t^{\varepsilon}(y)))| = 0.
\]
Given $x, y \in \mathbb{R}^d$, let $m > |x| \vee |y|$. For given $f \in C_b(\mathbb{R}^d)$ with $\|f\|_{\infty} \leq 1$, we have
\[
|\mathbb{E}(f(X_t^{\varepsilon}(x)) - f(X_t^{\varepsilon}(y)))| \leq |\mathbb{E}[f(X_t^{\varepsilon,m}(x))1_{t < r_B^{\varepsilon,x} \wedge r_B^{\varepsilon,y}}]| + 2P(t \geq r_B^{\varepsilon,x} \wedge r_B^{\varepsilon,y})
\]
\[
\leq |\mathbb{E}[f(X_t^{\varepsilon,m}(x)) - f(X_t^{\varepsilon,m}(y))]1_{t < r_B^{\varepsilon,x} \wedge r_B^{\varepsilon,y}})| + 2P(t \geq r_B^{\varepsilon,x} \wedge r_B^{\varepsilon,y})
\]
By (3.18), we have
\[
|\mathbb{E}(f(X_t^{\varepsilon,m}(x)) - f(X_t^{\varepsilon,m}(y)))| = \left| \int_{\mathbb{R}^d} f(z)(p_m^\varepsilon(t,x,z) - p_m^\varepsilon(t,x,z))dz \right|
\]
\[
\leq \|f\|_{\infty} |x - y| \int_{\mathbb{R}^d} \int_{0}^{1} |\nabla p_m^\varepsilon(t,x + r(y - x), z)|dzdr
\]
\[ \leq C_m t^{-\frac{1}{2}} |x-y| \int_0^1 \int_{\mathbb{R}^d} p^z_m(t, x + r(y - x), z)dz dr \]
\[ = C_m t^{-\frac{1}{2}} |x-y|. \]
Moreover, by Chebyshev’s inequality and (3.4),
\[ \mathbb{P}(t \geq \tau_{B_m}^{x,y}) \leq \mathbb{P} \left( \sup_{s \leq t} |X^x_s(x)| \geq m \right) \leq \mathbb{E} \left( \sup_{s \leq t} |X^x_s(x)|^{p/2} \right) / m^{p/2} \]
\[ \leq C (V_p(x) + t)^{1/2} / m^{p/2}. \tag{3.21} \]
Combining the above calculations, we obtain that for any \( f \in C_b(\mathbb{R}^d) \) with \( \|f\|_\infty \leq 1, \)
\[ |\mathbb{E}(f(X_t^x(x)) - f(X_t^y(y)))| \leq C_m t^{-\frac{1}{2}} |x-y| + C_0 (V_p(x) + V_p(y) + t)^{1/2} / m^{p/2}, \]
where \( C_0 \) does not depend on \( m \). Thus we obtain (3.20) by first letting \( x \to y \) and then \( m \to \infty. \)

**Lemma 3.4 (Irreducibility).** For any \( x_0, y_0 \in \mathbb{R}^d \) and \( r, t > 0 \), we have
\[ \inf_{x \in B_r(y_0)} \mathbb{P}(X_t^x(x) \in B_r(y_0)) > 0. \]

**Proof.** Since \( X_t^x(x) \) weakly converges to \( X_t(x) \) as \( \varepsilon \downarrow 0 \), for any open set \( A \subset \mathbb{R}^d \), we have
\[ \mathbb{P}(X_t^x(x) \in A) \geq \liminf_{\varepsilon \to 0} \mathbb{P}(X_t^\varepsilon(x) \in A). \tag{3.22} \]
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
\[ \mathbb{P}(X_t^x(x) \in B_r(y_0)) \geq \mathbb{P}(X_t^x(x) \in B_r(y_0); t < \tau_{B_m}^{x,y}) \geq \mathbb{P}(X_t^{x,m} \in B_r(y_0); t < \tau_{B_m}^{x,y}). \tag{3.23} \]
In order to use Theorem 2.3, we choose
\[ \rho(t, r) = t(1/\alpha + r)^{-d-\alpha}. \]
Clearly, by (3.17), estimate (2.5) is satisfied for \( \Gamma_t(x, y) = |x - \theta_t(y)| \). It remains to find \( t_0 \) and \( \delta_0 \) 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 (3.15) and the definition of \( b^*_m \), for all \( t \in [0, 1] \) and \( y \in \mathbb{R}^d \),
\[ |\theta_t(y) - y| \leq \int_0^t \| (b^*_m \ast \phi_s^{x,y}) (\theta_s(y)) \| ds \leq \sup_{|x| \leq m+1} |b(x)| \cdot t =: C_1 t. \]
Let
\[ \delta_0 \leq \frac{1}{2} \wedge \text{dist}(y_0, D_0), \quad t_0 \leq \frac{1}{2C_1} \wedge \frac{\delta_0}{C_1}. \tag{3.24} \]
For \( x', y \in \text{B}_{\delta_0}(y_0) \subset D_0 \), we have for \( t \in (0, t_0] \),
\[ \Gamma_t(x', y) = |x' - \theta_t(y)| \geq |x' - y_0| - |y - y_0| - |\theta_t(y) - y| \geq 2 - \delta_0 - C_1 t \geq 1. \]
On the other hand, if \( x \in B_{\delta_0}(y_0) \), we have
\[ \Gamma_t(x, y) = |x - \theta_t(y)| \leq |x - y| + |\theta_t(y) - y| \leq 2\delta_0 + C_1 t \leq 3\delta_0. \]
Now, for the \( C_0 \) in (3.17), one can choose \( \delta_0, t_0 > 0 \) small enough such that (3.24) holds and
\[ t \mapsto t/(1/\alpha + 1)^{d+\alpha} \]
is increasing on \( (0, t_0] \), and
\[ \frac{C_0^{-1} t}{(1/\alpha + t_0)^{d+\alpha}} > \frac{2C_0 t}{(1/\alpha + 1)^{d+\alpha}} \quad \text{for } t \in (0, t_0]. \]
In particular, (2.7) is satisfied. Thus by Theorem 2.3, we conclude
\[ \mathbb{P}(X_t^{x,m} \in B_r(y_0); t < \tau_{B_m}^{x,y}) \geq c_0, \]
where $c_0$ is independent of $\varepsilon$. This, together with (3.22) and (3.23), yields
\[ \mathbb{P}(X_t(x) \in B_r(y_0)) \geq c_0. \]
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.

Now, we are in a position to give

**Proof of Theorem 1.2.** First of all, by (3.4) and the standard Krylov-Bogoliubov method (see [12, Theorem 11.7]), there is an invariant probability measure $\mu$ associated with $(P_t)_{t \geq 0}$. The uniqueness follows by the strong Feller property and irreducibility. The exponential convergence (i) and (ii) follow by (3.5), 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 $\mu(dx) = e^{-U(x)}dx/\int_{\mathbb{R}^d} e^{-U(x)}dx$, where $U : \mathbb{R}^d \to \mathbb{R}$ is a continuous function. Suppose that there are $\beta, C_0 > 0$ such that for all $|x| \geq 1$,
\[ U(x) \geq (d + \beta) \log |x| - C_0. \tag{4.1} \]
The above assumption means that $e^{-U(x)}$ has a polynomial decay rate as $|x| \to \infty$, which is usually called the heavy-tailed distribution compared with the exponential or light tailed distribution. Below we want to find an ergodic SDE so that the law of the solution $X_t$ exponentially converges to $\mu$ in some sense as $t \to \infty$. Thus, one can sample $\mu$ theoretically from $X_t$ when $t$ is large.

Fix $\alpha \in (0, 2)$. To construct an ergodic SDE driven by $\alpha$-stable process, we introduce a vector field $B : \mathbb{R}^d \to \mathbb{R}^d$ by
\[ B(x) := c_{d, \alpha} \int_{\mathbb{R}^d} y [\varrho_\alpha(x + y) - \varrho_\alpha(x - y)] |y|^{-d-\alpha} dy, \]
where $c_{d, \alpha} = 2^\alpha \Gamma(d+\alpha)/\Gamma(\frac{d}{2}) \Gamma(\frac{2-\alpha}{2})$ and
\[ \varrho_\alpha(x) := (1 + |x|^2)^{-\frac{d+\alpha}{2}}. \]
The vector field $B$ enjoys the following important property.

**Lemma 4.1.** (i) $B \in C_0^\infty$ satisfies that for some $\kappa_0 = \kappa_0(d, \alpha) > 0$,
\[ |B(x)| \leq \kappa_0 |x|^{1-d-\alpha}, \quad |\nabla B(x)| \leq \kappa_0 |x|^{-d-\alpha}. \tag{4.2} \]
(ii) There is a constant $\kappa_1 > 0$ only depending on $d, \alpha$ such that for all $|x| > 1$,
\[ \langle x, B(x) \rangle \leq -\kappa_1 |x|^{2-d-\alpha}. \tag{4.3} \]
(iii) $\text{div} B(x) = \Delta \tilde{\varphi}_\alpha(x)$.

**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 (4.2), let $\chi : [0, \infty) \to [0, 1]$ be a cutoff function with
\[ \chi(r) = 1, \quad r \in [0, \frac{1}{2}], \quad \chi(r) = 0, \quad r \in [0, \frac{1}{2}]. \]
Define
\[ \varphi_{|x|}(y) := y |y|^{-d-\alpha} \chi \left( \frac{|y|}{|x|} \right), \quad \tilde{\varphi}_{|x|}(y) := y |y|^{-d-\alpha} (1 - \chi \left( \frac{|y|}{|x|} \right)). \]
For $j = 0, 1$, by definition we have
\[ \nabla^j B(x) = c_{d, \alpha} \int_{\mathbb{R}^d} y \left( \nabla^j \varrho_\alpha(x + y) - \nabla^j \varrho_\alpha(x - y) \right) |y|^{-d-\alpha} dy = c_{d, \alpha}(I_1 + I_2), \]
where
\[ I_1 := \int_{\mathbb{R}^d} \varphi_{|x|}(y) \left( \nabla^j \varrho_\alpha(x + y) - \nabla^j \varrho_\alpha(x - y) \right) dy, \]
\[ I_2 := \int_{\mathbb{R}^d} \tilde{\varphi}_{|x|}(y) \left( \nabla^j \varrho_\alpha(x + y) - \nabla^j \varrho_\alpha(x - y) \right) dy. \]

For \( I_1 \), we have
\[
|I_1| \leq \int_{\mathbb{R}^d} \varphi_{|x|}(y) \left| \int_{-1}^{1} |\nabla^{j+1} \varrho_\alpha(x + sy)| dy \right| ds \\
\lesssim \int_{|y| \leq \frac{|x|}{4}} |y|^{2-d-\alpha} \left( \int_{-1}^{1} (1 + |x + sy|)^{-(d+\alpha+1)+} ds \right) dy \\
\lesssim (1 + |x|)^{-(d+\alpha+1)+} \int_{|y| \leq \frac{|x|}{4}} |y|^{2-d-\alpha} dy \\
\lesssim (1 + |x|)^{-(d+\alpha+1)+} |x|^{2-\alpha} \lesssim |x|^{1-j-d-2\alpha}.
\]

For \( I_2 \), by the change of variable, we have
\[
I_2 = \int_{\mathbb{R}^d} \left( \tilde{\varphi}_{|x|}(x - y) - \tilde{\varphi}_{|x|}(x + y) \right) (\nabla^j \varrho_\alpha)(y) dy \\
= |x|^{-\alpha} \int_{\mathbb{R}^d} \left( \tilde{\varphi}_{|x|}(\bar{x} - y) - \tilde{\varphi}_{|x|}(\bar{x} + y) \right) (\nabla^j \varrho_\alpha)(|x| y) dy \\
\]
where \( \bar{x} = x/|x| \) and in the last step we have used
\[
\tilde{\varphi}_{|x|}(|x| y) = |x|^{1-d-\alpha} y y^{d-\alpha} (1 - \chi(|y|)) = |x|^{1-d-\alpha} \tilde{\varphi}_1(y).
\]

For \( j = 0 \), since \( |\tilde{\varphi}_1(y)| \leq |y|^{1-d-\alpha} 1_{|y| > 1/4} \leq 4^{d+\alpha-1} \), it is easy to see that
\[
|I_2| \leq 2 \cdot 4^{d+\alpha-1} |x|^{-\alpha} \int_{\mathbb{R}^d} (1 + |x| |y|)^{-(d+\alpha)} dy \lesssim |x|^{-\alpha-d}.
\]

For \( j = 1 \), since \( \|\text{div}_{y} \tilde{\varphi}_1\|_\infty < \infty \), by the integration by parts, we also have
\[
|I_2| \leq |x|^{-\alpha} \int_{\mathbb{R}^d} \left( \text{div}_{y} \tilde{\varphi}_1(\bar{x} - y) - \text{div}_{y} \tilde{\varphi}_1(\bar{x} + y) \right) \varrho_\alpha(|x| y) dy \\
\leq 2 \|\text{div}_{y} \tilde{\varphi}_1\|_\infty |x|^{-\alpha} \int_{\mathbb{R}^d} (1 + |x| |y|)^{-(d+\alpha)} dy \lesssim |x|^{-\alpha-d}.
\]
Combining the above estimates we obtain (4.2).

(ii) By definition and the symmetry, we have
\[
-\langle x, B(x) \rangle = c_{d,\alpha} \int_{\mathbb{R}^d} \langle x, y \rangle |\varrho_\alpha(x - y) - \varrho_\alpha(x + y)| |y|^{-d-\alpha} dy \\
= 2c_{d,\alpha} \int_{\mathbb{R}^d} \langle x, y \rangle |\varrho_\alpha(x - y) - \varrho_\alpha(x + y)| |y|^{-d-\alpha} dy.
\]

For \( |x| \geq 1 \) and \( \langle x, y \rangle \geq 0 \), by the mean-valued formula, we have
\[
\varrho_\alpha(x - y) - \varrho_\alpha(x + y) = (1 + |x - y|^2)^{-d+\alpha} - (1 + |x + y|^2)^{-d+\alpha} \\
= 2(d+\alpha) \langle x, y \rangle \int_{0}^{1} \left( 1 + s(|x - y|^2 - |x + y|^2) + |x + y|^2 \right)^{-d+\alpha-1} ds
\]
Since for

\[ H = \frac{d}{2} \]

Hence,

\[ \sigma > 1 \]

we have by (1.8) and

\[ x, y \geq 0, |x - y| \leq 1 \]

\[ \int_{x, y \geq 0, |x - y| \leq 1} (x, y)^2 (1 + 4|\beta|^{-2})(x, y) + |x - y|^2 \cdot \frac{4a}{d + a - 1} |y|^{-d - \alpha} dy. \]

Since for |x| \geq 1,

\[ |x - y| \leq 1 \Rightarrow |y|^2 \leq |x|^2 + |x|, \]

we further have

\[ \langle x, B(x) \rangle \geq \frac{4(d + \alpha)c_{d, \alpha}}{|x|^2} \int_{|x - y| \leq 1} (x, y)^2 (2 + 4(1 + |\beta|^{-1})) \cdot \frac{4a}{d + a - 1} |y|^{-d - \alpha} dy \]

\[ \geq \frac{4(d + \alpha)c_{d, \alpha}}{|x|^2} \int_{|x - y| \leq 1} (x, y)^2 |y|^{-d - \alpha} dy \geq \kappa_1 |x|^{2-d-\alpha}, \]

where \( \kappa_1 > 1 \) only depends on \( d, \alpha \). Thus we obtain (4.3).

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

\[ \text{div} B(x) = c_{d, \alpha} \int_{\mathbb{R}^d} \langle y, \nabla_y [\varrho_\alpha(x + y) - \varrho_\alpha(x - y)] \rangle |y|^{-d - \alpha} dy \]

\[ = c_{d, \alpha} \int_{\mathbb{R}^d} \langle y, \nabla_y [\varrho_\alpha(x + y) + \varrho_\alpha(x - y) - 2\varrho_\alpha(x)] \rangle |y|^{-d - \alpha} dy \]

\[ = -c_{d, \alpha} \int_{\mathbb{R}^d} \text{div}(y|y|^{-d-\alpha}) [\varrho_\alpha(x + y) + \varrho_\alpha(x - y) - 2\varrho_\alpha(x)] dy. \]

Since for \( y \neq 0 \),

\[ \text{div}(y|y|^{-d-\alpha}) = d|y|^{-d-\alpha} + \langle y, \nabla |y|^{-d-\alpha} \rangle = -\alpha|y|^{-d-\alpha}, \]

we have by (1.8) and \( c_{d, \alpha} = 2\alpha \Gamma(\frac{d + \alpha}{2}) / \Gamma(\frac{d}{2}) \Gamma(\frac{2 - \alpha}{2}) \),

\[ \text{div} B(x) = \alpha c_{d, \alpha} \int_{\mathbb{R}^d} \frac{[\varrho_\alpha(x + y) + \varrho_\alpha(x - y) - 2\varrho_\alpha(x)]}{|y|^{d+\alpha}} dy = \Delta^{\frac{\alpha}{2}} \varrho_\alpha(x). \]

The proof is complete. \( \square \)

Now we take \( \sigma \) and \( b \) in SDE (1.3) as follows:

\[ \sigma(x) := (\varrho_\alpha(x)e^{U(x)})^{\frac{1}{2}}, \quad b(x) := B(x)e^{U(x)}. \] (4.4)

We have the following main result of this section.

**Theorem 4.2.** Suppose that \( U \) is locally Lipschitz continuous and satisfies (4.1). Then with the above choices of \( \sigma \) and \( b, \mu(dx) = e^{-U(x)}dx / \int_{\mathbb{R}^d} e^{-U(x)}dx \) is the unique invariant probability measure of SDE (1.3) and the conclusions in Theorem 1.2 hold with \( r = \beta - \alpha \).

**Proof.** Let \( \sigma \) and \( b \) be defined by (4.4). By (1.7) and the change of variable, we can write

\[ \mathcal{L}_\sigma f(x) = \alpha^2 \Gamma(\frac{d + \alpha}{2}) / \Gamma(\frac{d}{2}) \Gamma(\frac{2 - \alpha}{2}) \int_{\mathbb{R}^d} \left[ f(x + \sigma(x)z) + f(x - \sigma(x)z) - 2f(x) \right] |z|^{d+\alpha} = \varrho_\alpha(x)e^{U(x)} \Delta^{\frac{\alpha}{2}} f(x). \]
Thus,

\[ \mathcal{L} f(x) = \mathcal{L}_\sigma f(x) + b \cdot \nabla f(x) = \varrho_\alpha(x)e^{U(x)}\Delta^{\frac{\alpha}{2}}f(x) + B(x)e^{U(x)} \cdot \nabla f(x), \tag{4.5} \]

and by (i) of Lemma 4.1,

\[ (\mathcal{L}^*e^{-U})(x) = \Delta^{\frac{\alpha}{2}}\varrho_\alpha(x) - \text{div}B(x) = 0, \]

which implies that \( \mu(dx) = e^{-U(x)}dx/\int_{\mathbb{R}^d} e^{-U(x)}dx \) is an invariant probability measure of SDE (1.3).

To check \((H_{\text{loc}})\) and \((H_{\text{glo}})\), by Remark 1.1, it suffices to verify (1.5) for \( \varepsilon_0 = 0 \). By (4.3) and (4.1), for any \( |x| \geq (2q/\kappa_1)^{1/\alpha} \lor 1 \), we have

\[ \langle x, b(x) \rangle + q\|\sigma(x)\|^{\alpha}|x|^{2-\alpha} = \left[ \langle x, B(x) \rangle + q\varrho_\alpha(x)|x|^{2-\alpha}e^{U(x)} \right] \]
\[ \leq \left[ -\kappa_1|x|^{2-d-\alpha} + q|x|^{2-d-2\alpha} e^{U(x)} \right] \]
\[ \leq -\kappa_1|x|^{2-d-\alpha} e^{U(x)} \leq -\frac{\kappa_1}{2}\varepsilon_0^\alpha |x|^{2+\beta-\alpha}. \]

So, Theorem 1.2 is applicable. The proof is complete. \( \Box \)

**Remark 4.3.** Suppose that \( U(x) = -\ln \varrho_\alpha(x) \). Note that by (4.2),

\[ \|\nabla(B/\varrho_\alpha)\|_{\infty} < \infty. \]

Let \( X_t \) be the unique solution of the following SDE:

\[ dX_t = (B/\varrho_\alpha)(X_t)dt + dL^\alpha_t, \quad X_0 = x. \tag{4.6} \]
Then \( \mu_0(dx) = \varrho_\alpha(x)dx / \int \varrho_\alpha(x)dx \) is the unique invariant probability measure of \( X_t \). It should be noticed that the distributional density \( p_\alpha(x) \) of \( L^\alpha_t \) is comparable with \( \varrho_\alpha(x) \) (cf. [35]), i.e., there is a constant \( \kappa = \kappa(d, \alpha) \geq 1 \) such that for all \( x \in \mathbb{R}^d \),

\[
\kappa^{-1} \varrho_\alpha(x) \leq p_\alpha(x) \leq \kappa \varrho_\alpha(x).
\]

Let \( \mathcal{L}_0 \) be the infinitesimal generator of SDE (4.6), i.e.,

\[
\mathcal{L}_0 = \Delta \tilde{\varphi} + B/\varrho_\alpha \cdot \nabla.
\]

Then by (4.5), one sees that

\[
\mathcal{L} = (\varrho_\alpha e^U) \mathcal{L}_0.
\]

In particular, the solution of SDE (1.3) with coefficients (4.4) is just a time change of SDE (4.6) (see [2, Section 1.15.2]).

**Remark 4.4.** The locally Lipschitz assumption on \( U \) can be relaxed as locally \( \gamma \)-Hölder continuity with \( \gamma \in ((1 - \alpha)^+, 1) \). If it is so, we need to check (1.5) for some \( \varepsilon_0 > 0 \).

5. Appendix

In this appendix we sketch the proof of weak convergence (3.12). First of all, by (3.11) and (3.4), there is a constant \( C > 0 \) such that for all \( \varepsilon \in (0, 1) \) and any \( x \in \mathbb{R}^d \) and \( T > 0 \),

\[
\left( \mathbb{E} \left( \sup_{s \in [0, T]} V_\rho(X_\varepsilon^x(s)) \right)^2 \right)^{\frac{1}{2}} \leq C V_\rho(x) + T.
\]

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

\[
\mathbb{P} \left( |X^{\varepsilon, m}_\eta - X^{\varepsilon, m}_{\eta'}| \geq \delta \right) \leq \mathbb{P} \left( |X^{\varepsilon, m}_\eta - X^{\varepsilon, m}_{\eta'}| \geq \delta; T < \tau_{B_m}^\varepsilon \right) + \mathbb{P} \left( T \geq \tau_{B_m}^\varepsilon \right)
\]

\[
= \mathbb{P} \left( |X^{\varepsilon, m}_\eta - X^{\varepsilon, m}_{\eta'}| \geq \delta; T < \tau_{B_m}^\varepsilon \right) + \mathbb{P} \left( T \geq \tau_{B_m}^\varepsilon \right)
\]

\[
\leq \mathbb{P} \left( |X^{\varepsilon, m}_\eta - X^{\varepsilon, m}_{\eta'}| \geq \delta \right) + 2 \mathbb{P} \left( T \geq \tau_{B_m}^\varepsilon \right).
\]

By SDE (3.16) and (3.13), (3.14), it is by now standard to derive that for fixed \( m \in \mathbb{N} \),

\[
\lim_{\delta \downarrow 0} \sup_{\varepsilon \in (0, 1)} \sup_{\eta, \eta' \leq T} \mathbb{P} \left( |X^{\varepsilon, m}_\eta - X^{\varepsilon, m}_{\eta'}| \geq \delta \right) = 0,
\]

which together with (3.21) yields that for any \( T, \delta > 0 \),

\[
\lim_{\delta \downarrow 0} \sup_{\varepsilon \in (0, 1)} \sup_{\eta, \eta' \leq T} \mathbb{P} \left( |X^{\varepsilon, m}_\eta - X^{\varepsilon, m}_{\eta'}| \geq \delta \right) = 0.
\]

Thus, by Aldous’ criterion (see [17, p.356, Theorem 4.5]), the law \( \mathbb{Q}^{\varepsilon} \) of \( (X^{\varepsilon, L^\alpha}_t), \varepsilon \in (0, 1) \) in \( \mathbb{D} \times \mathbb{D} \) is tight. Let \( \mathbb{Q} \) be any accumulation point of \( \left( \mathbb{Q}^{\varepsilon} \right)_{\varepsilon \in (0, 1)} \). Without loss of generality, we assume that for some subsequence \( \varepsilon_k \to 0 \), \( (Q^k)_{k \in \mathbb{N}} := (Q^{\varepsilon_k})_{k \in \mathbb{N}} \) weakly converges to \( \mathbb{Q} \) as \( k \to \infty \).

By Skorokhod’s representation theorem, there is a probability space \( (\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}) \) and \( \mathbb{D} \times \mathbb{D} \)-valued processes \( (\tilde{X}^k, \tilde{L}^k) \) and \( (\tilde{X}, \tilde{L}) \) such that

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

and

\[
\tilde{\mathbb{P}} \circ (\tilde{X}^k, \tilde{L}^k)^{-1} = Q^k, \text{ } \tilde{\mathbb{P}} \circ (\tilde{X}, \tilde{L})^{-1} = \mathbb{Q}.
\]

Moreover, \( \tilde{L}^k \) and \( \tilde{L} \) are still \( \alpha \)-stable Lévy processes, and

\[
\tilde{X}^k_t = x + \int_0^t b_{\varepsilon_k}(\tilde{X}^k_s)ds + \int_0^t \sigma_{\varepsilon_k}(\tilde{X}^k_s)d\tilde{L}^k_s.
\]

By [17, Theorem 6.22, p.383] and taking limits, one sees that

\[
\tilde{X}_t = x + \int_0^t b(\tilde{X}_s)ds + \int_0^t \sigma(\tilde{X}_s)d\tilde{L}_s,
\]
and \((\tilde{X}, \tilde{L})\) is a weak solution of SDE (1.3). Finally, by the weak uniqueness of [7] we obtain the weak convergence (3.12).

**Acknowledgement:** The authors would like to thank Jian Wang for useful conversation about the heavy-tailed sampling.

**References**

[1] Applebaum D.: *Lévy processes and stochastic calculus*. Cambridge Studies in Advanced Mathematics. Vol. 116, Cambridge University Press, Cambridge, 2009.

[2] Bakry D., Gentil I. and Ledoux M.: *Analysis and geometry of Markov diffusion operators*. Grundlehren der mathematischen Wissenschaften. Vol. 348, Springer, Cham, 2014.

[3] Bubeck S., Eldan R. and Lehec J.: Sampling from a log-concave distribution with projected Langevin Monte Carlo. *Discrete Comput. Geom.* Vol. 59, No. 4, pp. 757-783 (2018).

[4] Carmona P.: Existence and uniqueness of an invariant measure for a chain of oscillators in contact with two heat baths. *Stochastic Process. Appl.* Vol. 129, No. 12, pp. 5278-5311 (2019).

[5] Chen Z. Q., Zhang X. and Zhao G.: Supercritical SDEs driven by multiplicative stable-like Lévy processes. *Trans. Amer. Math. Soc.* Volume 374, Number 11, November 2021, Pages 7621-7655.

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

[7] Chung K. and Zhao Z.: *From Brownian motion to Schrödinger’s equation*. Grundlehren der mathematischen Wissenschaften. Vol. 312, Springer-Verlag, Berlin, 1995.

[8] Dalalyan A. S. and Karagulyan A.: User-friendly guarantees for the Langevin Monte Carlo with inaccurate gradient. *Stochastic Process. Appl.* Vol. 129, No. 12, pp. 4083-4125 (2019).

[9] Dalalyan A. S. and Tsybakov A. B.: Sparse regression learning by aggregation and Langevin Monte-Carlo. *Bernoulli.* Vol. 27, No. 3, pp. 2040-2068 (2021).

[10] Gürbüzbalaban M., Simsekli U. and Zhu L.: The heavy-tail phenomenon in SGD. *Stochastic Partial Differential Equations and Applications-VII*. Vol. 245, Chapman & Hall/CRC, Boca Raton, FL, 2006.

[11] Gökbergen I. and Shiryaev A. N.: *Limit theorems for stochastic processes*. Grundlehren der mathematischen Wissenschaften. Vol. 288, Springer-Verlag, Berlin, 1995.

[12] Huang L., Majka M. B. and Wang J.: Gradient estimates and Wasserstein-type distances for SDEs with Lévy noises. *Stochastic Process. Appl.* Vol. 149, No. 10, pp. 4083-4125 (2017).

[13] Majka M. B.: *Coupling and exponential ergodicity for stochastic differential equations driven by Lévy processes*. *Stochastic Process. Appl.* Vol. 127, No. 9, pp. 3129-3173 (2019).

[14] Masuda H.: *Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps*. *Stochastic Process. Appl.* Vol. 117, No. 1, pp. 35-56 (2007).

[15] Mattingly J.C., Stuart A.M. and Higman D.J.: Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise. *Stoch. Proc. Appl.* Vol. 101, pp. 185-232 (2002).

[16] Menozzi S. and Zhang X.: Heat kernel of supercritical SDEs with unbounded drifts. to appear in *Journal de l’École polytechnique-Mathématiques*, arXiv: 2012.14775 (2020).

[17] Meyn S. and Tweedie R. L.: *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, Second edition, 2009.

[18] Revuz D. and Yor M.: *Continuous martingales and Brownian motion*. Vol. 293, Springer-Verlag, Berlin, 1999.
[28] Roberts G. O. and Tweedie R. L.: Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*. Vol. 2, No. 4, pp. 341-363 (1996).

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

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

[31] Stein E.M.: *Singular integrals and differentiability properties of functions*. Princeton Mathematical Series, No. 30, Princeton University Press, Princeton, N.J., 1970.

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

[33] Wang F.: *Functional inequalities Markov semigroups and spectral theory*. Elsevier, 2006.

[34] Wang J.: *Lp*-Wasserstein distance for stochastic differential equations driven by Lévy processes. *Bernoulli*. Vol. 22, No. 3, pp. 1598-1616 (2016).

[35] Xie L. and Zhang X.: Ergodicity of stochastic differential equations with jumps and singular coefficients. *Ann. Inst. Henri Poincaré Probab. Stat.* Vol. 56, No. 1, pp. 175-229 (2020).

[36] Ye N. and Zhu Z.: Stochastic Fractional Hamiltonian Monte Carlo. *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. pp. 3019-3025 (2018).

XIAOLONG ZHANG: SCHOOL OF MATHEMATICS AND STATISTICS, WUHAN UNIVERSITY, WUHAN, HUBEI 430072, P.R.CHINA, EMAIL: zhangxl.math@whu.edu.cn

XICHENG ZHANG: SCHOOL OF MATHEMATICS AND STATISTICS, WUHAN UNIVERSITY, WUHAN, HUBEI 430072, P.R.CHINA, EMAIL: XICHENGZHANG@gmail.com