APPROXIMATION OF THE INVARIANT MEASURE OF STABLE SDES BY AN EULER–MARUYAMA SCHEME

PENG CHEN, CHANG-SONG DENG, RENÉ L. SCHILLING, AND LIHU XU

ABSTRACT. We propose two Euler-Maruyama (EM) type numerical schemes in order to approximate the invariant measure of a stochastic differential equation (SDE) driven by an α-stable Lévy process ($1 < \alpha < 2$): an approximation scheme with the α-stable distributed noise and a further scheme with Pareto-distributed noise. Using a discrete version of Duhamel's principle and Bismut's formula in Malliavin calculus, we prove that the error bounds in Wasserstein-1 distance are in the order of $\eta^{1-\epsilon}$ and $\eta^{\frac{2}{\alpha}-1}$, respectively, where $\epsilon \in (0,1)$ is arbitrary and η is the step size of the approximation schemes. For the Pareto-driven scheme, an explicit calculation for Ornstein–Uhlenbeck α-stable process shows that the rate $\eta^{\frac{2}{\alpha}-1}$ cannot be improved.

1. INTRODUCTION

We study the solution $(X_t)_{t\geq 0}$ of the following stochastic differential equation (SDE) driven by an α-stable Lévy process:

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \mathrm{d}Z_t, \quad X_0 = x, \tag{1.1}$$

where $x \in \mathbb{R}^d$ is the starting point, $(Z_t)_{t\geq 0}$ is a d-dimensional, rotationally invariant α-stable Lévy process with index $\alpha \in (1,2)$, and $b : \mathbb{R}^d \to \mathbb{R}^d$ is a function satisfying **Assumption A** below.

The Euler-Maruyama (EM) scheme of the SDE (1.1), with a step size $\eta \in (0,1)$, is defined by

$$Y_0 = x, \quad Y_{k+1} = Y_k + \eta b(Y_k) + (Z_{(k+1)\eta} - Z_{k\eta}), \quad k = 0, 1, 2, \dots, \tag{1.2}$$

see, e.g. [44, 20]. It is easy to see that $(Y_k)_{k\geq 0}$ is a Markov chain. A drawback of the scheme (1.2) is that there is no explicit representation for the probability density of α-stable noise $Z_{(k+1)\eta} - Z_{k\eta}$, $\alpha \in (1,2)$, making the numerical simulation is complicated and numerically expensive, see the very recent monograph [32, Section 1.9] for a detailed discussion about the difficulties arising in the multivariate stable distribution simulations. See also [5, 29, 31] for sampling stable distributed random variables. In contrast, the Pareto distribution has a simple probability density and thus can be easily sampled by the classical acceptance and rejection method. Since the stable and the Pareto distribution have the same tail behaviour, and inspired by the stable central limit theorem (see, e.g. [16, 6]), we replace the stable noise in (1.2) with a Pareto distributed noise, and consider the following EM scheme:

Let $\widetilde{Z}_1, \widetilde{Z}_2, \cdots$ be an iid sequence of d-dimensional random vectors, which are Pareto distributed, i.e.

$$\widetilde{Z}_1 \sim p(z) = \frac{\alpha}{\sigma_{d-1}|z|^{\alpha+d}}\,\mathbb{1}_{(1,\infty)}(|z|); \tag{1.3}$$

2020 *Mathematics Subject Classification.* 60H10; 37M25; 60G51; 60H07; 60H35; 60G52.
Key words and phrases. Euler-Maruyama method; invariant measure; convergence rate; Wasserstein distance.

we denote by $\sigma_{d-1} = 2\pi^{\frac{d}{2}}/\Gamma(\frac{d}{2})$ the surface area of the unit sphere $\mathbb{S}^{d-1} \subset \mathbb{R}^d$. We will approximate the SDE (1.1) by the following approximation scheme:

$$\tilde{Y}_0 = x, \quad \tilde{Y}_{k+1} = \tilde{Y}_k + \eta b(\tilde{Y}_k) + \frac{\eta^{1/\alpha}}{\sigma} \widetilde{Z}_{k+1}, \quad k = 0, 1, 2, \ldots, \tag{1.4}$$

where $\eta > 0$ is the step size, $\sigma^\alpha = \alpha/(\sigma_{d-1} C_{d,\alpha})$, and

$$C_{d,\alpha} = |\xi|^\alpha \left(\int_{\mathbb{R}^d \setminus \{0\}} (1 - \cos\langle \xi, y \rangle) \frac{\mathrm{d}y}{|y|^{\alpha+d}} \right)^{-1} = \alpha 2^{\alpha-1} \pi^{-d/2} \frac{\Gamma\left(\frac{d+\alpha}{2}\right)}{\Gamma\left(1 - \frac{\alpha}{2}\right)}, \tag{1.5}$$

see e.g. [4, Example 2.4.d)] and [2, III.18.23]. It is easy to see that $(\tilde{Y}_k)_{k\geq 0}$ is a Markov chain.

We aim to study the error bounds in the Wasserstein-1 distance for the above two schemes, in particular for large time.

1.1. **Motivation, contribution and method.**

The EM approximation of SDEs is a classical research topic, both in probability theory and in numerical analysis, and over the past decades there have been many contributions, see for instance [1, 14, 22, 41, 9, 34] for SDEs driven by a Brownian motion, and [19, 38, 17, 36, 28, 20] for SDEs driven by Lévy noise. Most of these papers focus on error bounds of the solution to the SDE and the EM approximation in a time interval $[0, T]$ for some finite $T > 0$; typically, there appears a constant C_T (depending on T) in the error bounds, which tends to ∞ as $T \to \infty$.

The recent use of Langevin samplings in machine learning, has caused a surge of interest error bounds for the invariant measures of the solution of the SDE and of the EM discretization, see e.g. [30, 42, 50, 7, 18]. We refer the reader to [23, 39, 43] for discrete schemes for the invariant measure of SDEs driven by a Brownian motion. Panloup [35] uses certain recursive procedures to compute the invariant measure of Lévy-driven SDEs, but he does not determine the convergence rate. To the best of our knowledge, our paper is the first contribution studying the bound between the invariant measures of solutions to SDEs driven by stable noise and their EM discretizations.

A further motivation of our research is to show that the EM scheme with Pareto distributed innovations can indeed be used to approximate the invariant measures of SDEs driven by an α-stable noise with $\alpha \in (1, 2)$. In order to speed up the EM scheme, actual implementations of the discretization (1.4), use iid random variables $(\widetilde{Z}_k)_{k\geq 1}$ with Pareto distribution rather than stable innovations. The advantage of this approach is that the Pareto distribution has an explicitly given density (see (1.3)) which allows for a much simpler sampling than stable random variables. We also show that the convergence rate $\eta^{2/\alpha - 1}$ is optimal for the Ornstein–Uhlenbeck process on $\mathbb{R}$.

For $\alpha = 2$, the stable process Z_t is (essentially) a d-dimensional standard Brownian motion and the convergence rate for the corresponding invariant measure is $\sqrt{\eta}$ (up to a logarithmic correction), see for instance [13]. Our optimal rate $\eta^{\frac{2}{\alpha}-1}$ will tend to $O(1)$ rather than $\sqrt{\eta}$ as $\alpha \uparrow 2$; this type of "phase transition" has been observed in many situations, e.g. in the stable law CLT [16, 48]. This is due to the fact that α-stable distributions with $\alpha \in (0, 2)$ do not have second moments, while the 2-stable distribution is the Gaussian law having arbitrary moments.

Our approach in proving the main results is via a discrete version of Duhamel principle and Bismut's formula in Malliavin calculus. More precisely, we split the stochastic process $(X_t)_{t\geq 0}$ into smaller pieces $(X_t)_{(k-1)\eta \leq t \leq k\eta}$ for $k \geq 1$ and replace $(X_t)_{(k-1)\eta \leq t \leq k\eta}$ with $\tilde{Y}_k$ and Y_k, respectively. This procedure is reminiscent to Lindeberg's method for the CLT. In order to bound the error caused by these replacements, we use the semigroup P_t given by $(X_t)_{t\geq 0}$ and study its regularity using Malliavin's calculus for jump processes. In order to bound the second-order derivative of P_t, we need to adopt the framework of the time-change argument established in [49] and use the Bismut formula.

1.2. **Notation.** Whenever we want to emphasize the starting point $X_0 = x$ for a given $x \in \mathbb{R}^d$, we will write X_t^x instead of X_t; we use this also for Y_k^y and $\tilde{Y}_k^y$ for a given $y \in \mathbb{R}^d$. By P_t, Q_k and $\tilde{Q}_k$ we denote the Markov semigroups of X_t, Y_k and $\tilde{Y}_k$, respectively, i.e.

$$P_t f(x) = \mathbb{E} f(X_t^x), \quad Q_k f(x) = \mathbb{E} f(Y_k^x), \quad \text{and} \quad \tilde{Q}_k f(x) = \mathbb{E} f(\tilde{Y}_k^x).$$

for a bounded measurable function $f : \mathbb{R}^d \to \mathbb{R}$, $x \in \mathbb{R}^d$, $t \geq 0$ and $k = 0, 1, 2, \ldots$.

As usual, $\mathcal{C}(\mathbb{R}^d, \mathbb{R})$ denote the continuous functions $f : \mathbb{R}^d \to \mathbb{R}$, and $\mathcal{C}^2(\mathbb{R}^d, \mathbb{R})$ [$\mathcal{C}_b^2(\mathbb{R}^d, \mathbb{R})$] are the twice continuously differentiable functions [which are bounded together with all their derivatives]; $\nabla f(x) \in \mathbb{R}^d$ and $\nabla^2 f(x) \in \mathbb{R}^{d\times d}$ are the gradient and the Hessian. For $v, v_1, v_2, x \in \mathbb{R}^d$, the directional derivatives are given by

$$\nabla_v f(x) = \langle \nabla f(x), v \rangle = \lim_{\varepsilon \to 0} \frac{f(x+\varepsilon v) - f(x)}{\varepsilon},$$
$$\nabla_{v_2} \nabla_{v_1} f(x) = \left\langle \nabla^2 f(x), v_1 v_2^\top \right\rangle_{\mathrm{HS}} = \lim_{\varepsilon \to 0} \frac{\nabla_{v_1} f(x+\varepsilon v_2) - \nabla_{v_1} f(x)}{\varepsilon},$$

where $\langle A, B \rangle_{\mathrm{HS}} := \sum_{i,j=1}^d A_{ij} B_{ij}$ for $A, B \in \mathbb{R}^{d\times d}$. The Hilbert-Schmidt norm of a matrix $A \in \mathbb{R}^{d\times d}$ is $\|A\|_{\mathrm{HS}} = \sqrt{\sum_{i,j=1}^d A_{ij}^2}$.

The directional derivatives are similarly defined for (sufficiently smooth) vector-valued functions $f = (f_1, f_2, \cdots, f_d)^\top : \mathbb{R}^d \to \mathbb{R}^d$: let $v, v_1, v_2, x \in \mathbb{R}^d$, then $\nabla_v f(x) = (\nabla_v f_1, \nabla_v, \ldots, \nabla_v f_d)^\top$, $\nabla_{v_2}\nabla_{v_1} f(x) = (\nabla_{v_2}\nabla_{v_1} f_1, \ldots, \nabla_{v_2}\nabla_{v_1} f_d)^\top$.

For $f \in \mathcal{C}_b^2(\mathbb{R}^d, \mathbb{R})$, we will use the supremum and the supremum Hilbert-Schmidt norm

$$\|\nabla f\|_\infty = \sup_{x\in\mathbb{R}^d} |\nabla f(x)|, \quad \|\nabla^2 f\|_{\mathrm{HS},\infty} = \sup_{x\in\mathbb{R}^d} \|\nabla^2 f(x)\|_{\mathrm{HS}}.$$

The Wasserstein-1 distance between two probability measures μ_1 and μ_2 on $\mathbb{R}^d$ is defined as

$$W_1(\mu_1, \mu_2) = \inf_{(X,Y)\in \mathsf{C}(\mu_1,\mu_2)} \mathbb{E}|X - Y|, \tag{1.6}$$

where $\mathsf{C}(\mu_1, \mu_2)$ is the set of all coupling realizations of μ_1, μ_2, i.e. all random variables with values in $\mathbb{R}^{2d}$ with marginals μ_1, μ_2. We also have the following dual description of the Wasserstein distance

$$W_1(\mu_1, \mu_2) = \sup_{h\in \mathrm{Lip}(1)} |\mu_1(h) - \mu_2(h)|,$$

where $\mathrm{Lip}(1) = \{h : \mathbb{R}^d \to \mathbb{R};\ |h(y) - h(x)| \leq |y - x|\}$ and $\mu_i(h) = \int_{\mathbb{R}} h(x)\, \mu_i(\mathrm{d}x)$, $i = 1, 2$.

We will frequently need the following weight function

$$V_\beta(x) = (1 + |x|^2)^{\beta/2}, \quad x \in \mathbb{R}^d,\ \beta \geq 0.$$

Finally, we write $\lfloor x \rfloor$ for the largest integer which is less than or equal to $x \in \mathbb{R}$, and throughout $C_{d,\alpha}$ is the constant (1.5).

1.3. **Assumptions and main results.** Throughout this paper, we make the following assumption:

Assumption A. The function $b : \mathbb{R}^d \to \mathbb{R}^d$ is twice continuously differentiable and there exist constants $\theta_1, \theta_2 > 0$ and $\theta_3, K \geq 0$ such that

$$\langle b(x) - b(y), x - y \rangle \leq -\theta_1 |x - y|^2 + K \quad \forall x, y \in \mathbb{R}^d \tag{1.7}$$

and

$$|\nabla_v b(x)| \le \theta_2 |v|, \quad |\nabla_{v_1}\nabla_{v_2} b(x)| \le \theta_3 |v_1|\,|v_2| \quad \forall v, v_1, v_2, x \in \mathbb{R}^d. \tag{1.8}$$

Remark 1.1. Note that (1.8) immediately implies the following linear growth condition

$$|b(x) - b(0)| \le \theta_2 |x|, \quad x \in \mathbb{R}^d. \tag{1.9}$$

Under **Assumption A**, we will show that both $(X_t)_{t\ge 0}$ and $(\tilde{Y}_k)_{k\ge 0}$ are ergodic; we write μ and $\tilde{\mu}_\eta$, respectively, for their invariant measures, see Propositions 1.5 and 1.7 below. Throughout the paper the constants C, c_1, c_2, c_3, c_4 and λ may depend on $\theta_1, \theta_2, \theta_3, K, \alpha, d, |b(0)|$ and β for some constant $\beta \in [1, \alpha)$, but we often suppress this in our notation; moreover, the exact values of the constants may vary from line to line. Our *main results* are the following two theorems:

Theorem 1.2. *Let $(X_t)_{t\ge 0}$ and $(\tilde{Y}_k)_{k\ge 0}$ be defined by (1.1) and (1.4) (step size η), and denote by μ and $\tilde{\mu}_\eta$, their invariant measures. Under **Assumption A**, there exists a constant C such that the following two statements hold:*

(1) *For every $N \ge 2$ and step size $\eta < \min\{1,\ \theta_1/(8\theta_2^2),\ 1/\theta_1\}$, one has*

$$W_1\big(\mathrm{law}(X_{\eta N}), \mathrm{law}(\tilde{Y}_N)\big) \le C(1+|x|)\eta^{2/\alpha - 1}.$$

(2) *For every step size $\eta < \min\{1,\ \theta_1/\theta_2^2,\ 1/\theta_1\}$, one has*

$$W_1\big(\mu, \tilde{\mu}_\eta\big) \le C\eta^{2/\alpha - 1}.$$

Theorem 1.3. *Let $(X_t)_{t\ge 0}$ and $(Y_k)_{k\ge 0}$ be defined by (1.1) and (1.2) (step size η), and denote by μ and μ_η, their invariant measures. Under **Assumption A**, for any $\beta \in [1, \alpha)$, there exists a constant C depending on β such that the following two statements hold:*

(1) *For every $N \ge 2$ and step size $\eta < \min\{1,\ \theta_1/(8\theta_2^2),\ 1/\theta_1\}$, one has*

$$W_1\big(\mathrm{law}(X_{\eta N}), \mathrm{law}(Y_N)\big) \le C(1+|x|^\beta)\eta^{1+\frac{1}{\alpha}-\frac{1}{\beta}}.$$

(2) *For every step size $\eta < \min\{1,\ \theta_1/\theta_2^2,\ 1/\theta_1\}$, one has*

$$W_1\big(\mu, \mu_\eta\big) \le C\eta^{1+\frac{1}{\alpha}-\frac{1}{\beta}}.$$

Remark 1.4. The rate $\eta^{2/\alpha-1}$ in the first theorem is optimal for the one-dimensional Ornstein–Uhlenbeck process, see Proposition B.1 below.

The proofs of Theorems 1.2 and 1.3 are presented in Section 2. In Section 3, we use a time-change argument and the Bismut formula to prove Lemma 2.1, which is the key to the proof of our main result. Appendix A includes the proofs of the propositions in this section for the completeness. Finally, in Appendix B, the exact convergence rate $\eta^{2/\alpha-1}$ is reached for the Ornstein–Uhlenbeck process on $\mathbb{R}$, which shows that the rate in Theorem 1.2 (2) is sharp.

1.4. **Auxiliary propositions.** Here we collect a few auxiliary properties of $(X_t)_{t\ge 0}$ and $(Y_k)_{k\ge 0}$. The proofs are standard, but we include them in Appendix A to be self-contained. Recall that $V_\beta(x) = (1+|x|^2)^{\beta/2}$.

Proposition 1.5. *Let **Assumption A** hold and denote by $(X_t)_{t\ge 0}$ the solution to the SDE (1.1). Then, $(X_t)_{t\ge 0}$ admits a unique invariant probability measure μ such that for $1 \le \beta < \alpha$*

$$\sup_{|f|\le V_\beta} \big|\mathbb{E}[f(X_t^x)] - \mu(f)\big| \le c_1 V_\beta(x)\mathrm{e}^{-c_2 t}, \quad t > 0, \tag{1.10}$$

for some constants $c_1, c_2 > 0$. In particular, there exists a constant $C > 0$ such that

$$\mathbb{E}|X_t^x|^\beta \le C(1+|x|^\beta), \quad t > 0. \tag{1.11}$$

Proposition 1.6. *Under **Assumption A**, there exist for every $t > 0$ and $x, y \in \mathbb{R}^d$ constants $C > 0$ and $\lambda > 0$ such that*

$$W_1\left(\mathrm{law}(X_t^x), \mathrm{law}(X_t^y)\right) \le C\mathrm{e}^{-\lambda t}|x-y|.$$

Proposition 1.7. *Let **Assumption A** hold and denote by $(Y_k)_{k\ge 0}$ and $(\tilde{Y}_k)_{k\ge 0}$ the Markov chains defined by (1.2) and (1.4), respectively. Assume that the step size satisfies $\eta < \min\{1,\ \theta_1/\theta_2^2,\ 1/\theta_1\}$. Then*

(1) *the chain $(Y_k)_{k\ge 0}$ admits a unique invariant measure μ_η, such that for all $x \in \mathbb{R}^d$ and $k > 0$,*

$$\sup_{|f|\le V_1} |\mathbb{E} f(Y_k^x) - \mu_\eta(f)| \le c_1 V_1(x)\mathrm{e}^{-c_2 k}, \tag{1.12}$$

for some constants $c_1, c_2 > 0$.

(2) *the chain $(\tilde{Y}_k)_{k\ge 0}$ admits a unique invariant measure $\tilde{\mu}_\eta$, such that for all $x \in \mathbb{R}^d$ and $k > 0$,*

$$\sup_{|f|\le V_1} |\mathbb{E} f(\tilde{Y}_k^x) - \tilde{\mu}_\eta(f)| \le c_3 V_1(x)\mathrm{e}^{-c_4 k}, \tag{1.13}$$

for some constants $c_3, c_4 > 0$.

Lemma 1.8. *Let **Assumption A** hold and denote by $(Y_k)_{k\ge 0}$ and $(\tilde{Y}_k)_{k\ge 0}$ the Markov chains defined by (1.2) and (1.4), respectively. If the step size satisfies $\eta < \min\left\{1, \frac{\theta_1}{8\theta_2^2}, \frac{1}{\theta_1}\right\}$, then there is a constant $C > 0$, which is independent of η, such that*

$$\mathbb{E}|Y_k^x|^\beta \le C(1+|x|^\beta), \tag{1.14}$$

$$\mathbb{E}|\tilde{Y}_k^x| \le C(1+|x|), \tag{1.15}$$

hold for any $\beta \in [1, \alpha)$, $x \in \mathbb{R}^d$ and $k > 0$.

2. Proof of Theorems 1.2 and 1.3

We begin with several auxiliary lemmas which will be used to prove Theorems 1.2 and 1.3.

2.1. Auxiliary lemmas.

The first auxiliary lemma is about the regularity of the semigroup induced by $(X_t)_{t\ge 0}$. We will prove it using Malliavin's calculus in Section 3 further down.

Lemma 2.1. *Let $h \in \mathrm{Lip}(1)$ and X_t^x be the solution to the SDE (1.1). For all vectors $v, v_1, v_2 \in \mathbb{R}^d$ and $t \in (0, 1]$, we have*

$$|\nabla_v P_t h(x)| \le \mathrm{e}^{\theta_2}\, |v| \tag{2.1}$$

and

$$|\nabla_{v_2}\nabla_{v_1} P_t h(x)| \le C t^{-1/\alpha}\, |v_1|\, |v_2|\,, \tag{2.2}$$

for some constant $C > 0$.

Using the inequalities (1.11) and (1.14), we can obtain the following estimates:

Lemma 2.2. *Let $(X_t)_{t\ge 0}$ be the solution to the SDE (1.1) and $(Y_k)_{k\ge 0}$ be the Markov chains defined by (1.2). If the step size satisfies $\eta < \min\left\{1, \frac{\theta_1}{8\theta_2^2}, \frac{1}{\theta_1}\right\}$, then the following estimates* *hold for all $t \in (0,1]$, $\beta \in [1,\alpha)$:*

$$\mathbb{E}|Y_1^x - x|^\beta \le C(1+|x|^\beta)\eta^{\beta/\alpha}, \tag{2.3}$$

$$\mathbb{E}|X_t^x - x|^\beta \le C(1+|x|^\beta)t^{\beta/\alpha}, \tag{2.4}$$

$$\mathbb{E}|X_\eta^x - Y_1^x|^\beta \le C(1+|x|^\beta)\eta^{\beta+\frac{\beta}{\alpha}}. \tag{2.5}$$

Proof. The first inequality follows immediately from

$$\mathbb{E}|Y_1^x - x|^\beta = \mathbb{E}\left|\eta b(x) + Z_\eta\right|^\beta \le 2\left[\eta^\beta |b(x)|^\beta + \mathbb{E}|Z_\eta|^\beta\right] \le C(1+|x|^\beta)\eta^{\beta/\alpha}.$$

From the Hölder inequality and (1.11), we obtain

$$\begin{aligned}
\mathbb{E}|X_t^x - x|^\beta &= \mathbb{E}\left|\int_0^t b(X_s^x)\,\mathrm{d}s + Z_t\right|^\beta \\
&\le 2\mathbb{E}\left|\int_0^t b(X_s^x)\,\mathrm{d}s\right|^\beta + 2|Z_t|^\beta \\
&\le 2t^{\beta-1}\int_0^t \mathbb{E}|b(X_s^x)|^\beta\,\mathrm{d}s + 2\mathbb{E}|Z_1|^\beta t^{\frac{\beta}{\alpha}} \\
&\le C(1+|x|^\beta)t^{\beta/\alpha},
\end{aligned}$$

which implies the second inequality.

For the last inequality, the Hölder inequality, (1.9) and (2.4) imply

$$\begin{aligned}
\mathbb{E}|X_\eta^x - Y_1^x|^\beta &= \mathbb{E}\left|\int_0^\eta [b(X_s^x) - b(x)]\,\mathrm{d}s\right|^\beta \\
&\le \eta^{\beta-1}\int_0^\eta \mathbb{E}|b(X_s^x) - b(x)|^\beta\,\mathrm{d}s \\
&\le \theta\eta^{\beta-1}\int_0^\eta \mathbb{E}|X_s^x - x|^\beta\,\mathrm{d}s \\
&\le C(1+|x|^\beta)\eta^{\beta-1}\int_0^\eta s^{\frac{\beta}{\alpha}}\,\mathrm{d}s \\
&\le C(1+|x|^\beta)\eta^{\beta+\frac{\beta}{\alpha}}.
\end{aligned}$$

□

In order to prove Theorem 1.2, we need the following two lemmas. The first is just an intermediate step for the proof of the second lemma, which is the key to proving Theorem 1.2. Notice that the fractional Laplacian operator $(-\Delta)^{\alpha/2}$ is the infinitesimal generator of the rotationally invariant α-stable Lévy process process $(Z_t)_{t\ge 0}$, which is defined as a principal value (p.v.) integral: for any $f \in \mathcal{C}^2(\mathbb{R}^d,\mathbb{R})$,

$$(-\Delta)^{\alpha/2} f(x) = C_{d,\alpha}\cdot \text{p.v.}\int_{\mathbb{R}^d} (f(x+y) - f(x))\frac{\mathrm{d}y}{|y|^{\alpha+d}}. \tag{2.6}$$

Lemma 2.3. *Let $\alpha \in (1,2)$ and $f : \mathbb{R}^d \to \mathbb{R}$ satisfying $\|\nabla f\|_\infty < \infty$ and $\|\nabla^2 f\|_{\mathrm{HS},\infty} < \infty$. For all $x, y \in \mathbb{R}^d$ one has*

$$\left|(-\Delta)^{\alpha/2} f(x) - (-\Delta)^{\alpha/2} f(y)\right| \le \frac{C_{d,\alpha}\|\nabla^2 f\|_{\mathrm{HS},\infty}\sigma_{d-1}}{(2-\alpha)(\alpha-1)}|x-y|^{2-\alpha}. \tag{2.7}$$

Proof. From the definition of the fractional Laplacian (2.6) and the symmetry of the representing measure we have for any $R > 0$

$$\begin{aligned}(-\Delta)^{\alpha/2} f(x) &= C_{d,\alpha} \int_{\mathbb{S}^{d-1}} \int_0^\infty \frac{f(x+r\theta) - f(x) - r\langle \theta, \nabla f(x)\rangle \mathbb{1}_{(0,R)}(r)}{r^{\alpha+1}} \,\mathrm{d}r\,\mathrm{d}\theta \\ &= C_{d,\alpha} \int_{\mathbb{S}^{d-1}} \int_0^R \int_0^r \frac{\langle \theta, \nabla f(x+\theta s) - \nabla f(x)\rangle}{r^{\alpha+1}} \,\mathrm{d}s\,\mathrm{d}r\,\mathrm{d}\theta \\ &\quad + C_{d,\alpha} \int_{\mathbb{S}^{d-1}} \int_R^\infty \int_0^r \frac{\langle \theta, \nabla f(x+\theta s)\rangle}{r^{\alpha+1}} \,\mathrm{d}s\,\mathrm{d}r\,\mathrm{d}\theta\end{aligned}$$

Then, for all $x, y \in \mathbb{R}^d$,

$$\begin{aligned}&\left|(-\Delta)^{\alpha/2} f(x) - (-\Delta)^{\alpha/2} f(y)\right| \\ &\leq C_{d,\alpha} \int_{\mathbb{S}^{d-1}} \int_0^R \int_0^r \frac{|\nabla f(x+\theta s) - \nabla f(x) - \nabla f(y+\theta s) + \nabla f(y)|}{r^{\alpha+1}} \,\mathrm{d}s\,\mathrm{d}r\,\mathrm{d}\theta \\ &\quad + C_{d,\alpha} \int_{\mathbb{S}^{d-1}} \int_R^\infty \int_0^r \frac{|\nabla f(x+\theta s) - \nabla f(y+\theta s)|}{r^{\alpha+1}} \,\mathrm{d}s\,\mathrm{d}r\,\mathrm{d}\theta.\end{aligned}$$

For the first integral we have

$$\begin{aligned}&C_{d,\alpha} \int_{\mathbb{S}^{d-1}} \int_0^R \int_0^r \frac{|\nabla f(x+\theta s) - \nabla f(x) - \nabla f(y+\theta s) + \nabla f(y)|}{r^{\alpha+1}} \,\mathrm{d}s\,\mathrm{d}r\,\mathrm{d}\theta \\ &\leq C_{d,\alpha} \int_{\mathbb{S}^{d-1}} \int_0^R \int_0^r \frac{|\nabla f(x+\theta s) - \nabla f(x)| + |\nabla f(y+\theta s) - \nabla f(y)|}{r^{\alpha+1}} \,\mathrm{d}s\,\mathrm{d}r\,\mathrm{d}\theta \\ &\leq 2C_{d,\alpha} \|\nabla^2 f\|_{\mathrm{HS},\infty} \int_{\mathbb{S}^{d-1}} \int_0^r \frac{s}{r^{\alpha+1}} \,\mathrm{d}s\,\mathrm{d}r\,\mathrm{d}\theta = \frac{C_{d,\alpha} \|\nabla^2 f\|_{\mathrm{HS},\infty} \sigma_{d-1}}{2-\alpha} R^{2-\alpha},\end{aligned}$$

and for the second term we get

$$\begin{aligned}&C_{d,\alpha} \int_{\mathbb{S}^{d-1}} \int_R^\infty \int_0^r \frac{|\nabla f(x+\theta s) - \nabla f(y+\theta s)|}{r^{\alpha+1}} \,\mathrm{d}s\,\mathrm{d}r\,\mathrm{d}\theta \\ &\leq C_{d,\alpha} \|\nabla^2 f\|_{\mathrm{HS},\infty} \int_{\mathbb{S}^{d-1}} \int_R^\infty \int_0^r \frac{|x-y|}{r^{\alpha+1}} \,\mathrm{d}s\,\mathrm{d}r\,\mathrm{d}\theta \\ &= \frac{C_{d,\alpha} \|\nabla^2 f\|_{\mathrm{HS},\infty} \sigma_{d-1}}{\alpha-1} |x-y| R^{1-\alpha}.\end{aligned}$$

Hence, the assertion follows upon taking $R = |x-y|$. □

Lemma 2.4. *Let $(X_t)_{t\geq 0}$ and $(\tilde{Y}_k)_{k\geq 0}$ be defined by (1.1) and (1.4), respectively. There exists a constant $C > 0$ such that for all $x \in \mathbb{R}^d$, $\eta \in (0,1)$, $f : \mathbb{R}^d \to \mathbb{R}$ satisfying $\|\nabla f\|_\infty < \infty$ and $\|\nabla^2 f\|_{\mathrm{HS},\infty} < \infty$,*

$$|P_\eta f(x) - \tilde{Q}_1 f(x)| \leq C(1+|x|)\left(\|\nabla f\|_\infty + \|\nabla^2 f\|_{\mathrm{HS},\infty}\right)\eta^{2/\alpha}.$$

Proof. From (1.1) and (1.4), we see

$$\begin{aligned}\mathbb{E}[f(X_\eta^x) - f(\tilde{Y}_1)] &= \mathbb{E}\left[f\left(x + \int_0^\eta b(X_r^x)\,\mathrm{d}r + Z_\eta\right) - f\left(x + \eta b(x) + \frac{\eta^{1/\alpha}}{\sigma}\widetilde{Z}\right)\right] \\ &= \mathsf{J}_1 + \mathsf{J}_2,\end{aligned}$$

where

$$\mathsf{J}_1 := \mathbb{E}\left[f\left(x+\int_0^\eta b(X_r^x)\,\mathrm{d}r + Z_\eta\right) - f\left(x+\eta b(x)+Z_\eta\right)\right],$$

$$\mathsf{J}_2 := \mathbb{E}\left[f\left(x+\eta b(x)+Z_\eta\right) - f\left(x+\eta b(x)\right)\right] - \mathbb{E}\left[f\left(x+\eta b(x)+\frac{\eta^{1/\alpha}}{\sigma}\widetilde{Z}\right) - f\left(x+\eta b(x)\right)\right].$$

We can bound J_1 using (1.8) and (2.4) with $\beta = 1$:

$$\begin{aligned}
|\mathsf{J}_1| &\le \|\nabla f\|_\infty \mathbb{E}\left|\int_0^\eta b(X_r^x)\,\mathrm{d}r - \eta b(x)\right| \\
&\le \|\nabla f\|_\infty \int_0^\eta \mathbb{E}|b(X_r^x)-b(x)|\,\mathrm{d}r \\
&\le \theta_2\|\nabla f\|_\infty \int_0^\eta \mathbb{E}|X_r^x - x|\,\mathrm{d}r \\
&\le C\theta_2(1+|x|)\|\nabla f\|_\infty \int_0^\eta r^{1/\alpha}\,\mathrm{d}r \\
&\le C(1+|x|)\|\nabla f\|_\infty \eta^{1+1/\alpha}.
\end{aligned}$$

For the first term of J_2 we use Dynkin's formula (see e.g. [8]) to get

$$\mathbb{E}\big[f\big(x+\eta b(x)+Z_\eta\big) - f\big(x+\eta b(x)\big)\big] = \int_0^\eta \mathbb{E}\big[(-\Delta)^{\alpha/2} f\big(x+\eta b(x)+Z_r\big)\big]\,\mathrm{d}r.$$

For the second part of J_2 we use that $C_{d,\alpha} = \alpha\sigma_{d-1}^{-1}\sigma^{-\alpha}$ and Taylor's formula to see

$$\begin{aligned}
&\mathbb{E}\left[f\left(x+\eta b(x)+\frac{\eta^{1/\alpha}}{\sigma}\widetilde{Z}\right) - f\big(x+\eta b(x)\big)\right] \\
&= \frac{\eta^{1/\alpha}}{\sigma}\mathbb{E}\left[\int_0^1 \left\langle \nabla f\left(x+\eta b(x)+\frac{\eta^{1/\alpha}}{\sigma}t\widetilde{Z}\right),\, \widetilde{Z}\right\rangle \mathrm{d}t\right] \\
&= \frac{\eta^{1/\alpha}}{\sigma}\int_{|z|\ge 1}\int_0^1 \alpha\left\langle \nabla f\big(x+\eta b(x)+\frac{\eta^{1/\alpha}}{\sigma}tz\big),\, z\right\rangle \frac{\mathrm{d}t\,\mathrm{d}z}{\sigma_{d-1}|z|^{\alpha+d}} \\
&= \frac{\alpha\eta}{\sigma_{d-1}\sigma^\alpha}\int_{|z|\ge\sigma^{-1}\eta^{1/\alpha}}\int_0^1 \big\langle \nabla f\big(x+\eta b(x)+tz\big),\, z\big\rangle \frac{\mathrm{d}t\,\mathrm{d}z}{|z|^{\alpha+d}} \\
&= \eta(-\Delta)^{\alpha/2} f(x+\eta b(x)) - \mathsf{R},
\end{aligned}$$

where

$$\mathsf{R} := \eta C_{d,\alpha}\int_{|z|<\sigma^{-1}\eta^{1/\alpha}}\int_0^1 \big\langle \nabla f\big(x+\eta b(x)+tz\big),\, z\big\rangle \frac{\mathrm{d}t\,\mathrm{d}z}{|z|^{\alpha+d}}.$$

Together, the above estimates yield

$$|\mathsf{J}_2| \le |\mathsf{R}| + \left|\int_0^\eta \mathbb{E}\big[(-\Delta)^{\alpha/2} f\big(x+\eta b(x)+Z_r\big)\big]\,\mathrm{d}r - \eta(-\Delta)^{\alpha/2} f(x+\eta b(x))\right|.$$

Further, we have

$$
\begin{aligned}
|\mathsf{R}| &= \eta C_{d,\alpha}\left|\int_{|z|<\sigma^{-1}\eta^{1/\alpha}}\int_0^1 \left\langle \nabla f\big(x+\eta b(x)+tz\big)-\nabla f\big(x+\eta b(x)\big),\ z\right\rangle \frac{\mathrm{d}t\,\mathrm{d}z}{|z|^{\alpha+d}}\right|\\
&\le \eta C_{d,\alpha}\int_{|z|<\sigma^{-1}\eta^{1/\alpha}}\int_0^1 \left|\nabla f\big(x+\eta b(x)+tz\big)-\nabla f\big(x+\eta b(x)\big)\right| \frac{\mathrm{d}t\,\mathrm{d}z}{|z|^{\alpha+d-1}}\\
&\le \frac{1}{2}\eta C_{d,\alpha}\|\nabla^2 f\|_{\mathrm{HS},\infty}\int_{|z|<\sigma^{-1}\eta^{1/\alpha}}\frac{\mathrm{d}z}{|z|^{\alpha+d-2}} \le C\|\nabla^2 f\|_{\mathrm{HS},\infty}\,\eta^{2/\alpha}.
\end{aligned}
$$

By Lemma 2.3, we also have

$$
\begin{aligned}
&\left|\int_0^\eta \mathbb{E}\left[(-\Delta)^{\alpha/2} f\big(x+\eta b(x)+Z_r\big)\right]\mathrm{d}r-\eta(-\Delta)^{\alpha/2}f(x+\eta b(x))\right|\\
&\quad\le \int_0^\eta \mathbb{E}\left|(-\Delta)^{\alpha/2} f\big(x+\eta b(x)+Z_r\big)\right]-(-\Delta)^{\alpha/2}f(x+\eta b(x))\Big|\,\mathrm{d}r\\
&\quad\le C\|\nabla^2 f\|_{\mathrm{HS},\infty}\int_0^\eta \mathbb{E}\left[|Z_r|^{2-\alpha}\right]\mathrm{d}r\\
&\quad= C\|\nabla^2 f\|_{\mathrm{HS},\infty}\int_0^\eta \mathbb{E}\left[|Z_1|^{2-\alpha}\right] r^{2/\alpha-1}\,\mathrm{d}r\\
&\quad\le C\mathbb{E}\left[|Z_1|^{2-\alpha}\right]\|\nabla^2 f\|_{\mathrm{HS},\infty}\,\eta^{2/\alpha}.
\end{aligned}
$$

The proof follows if we combine all estimates. □

In order to prove Theorem 1.3, we need two more lemmas. The first is just an intermediate step for the proof of the second lemma, which is the key to proving Theorem 1.3.

Lemma 2.5. *Assume that f satisfies $\|\nabla f\|_\infty<\infty$ and $\|\nabla^2 f\|_{\mathrm{HS},\infty}<\infty$. For any $\beta\in[1,2]$ and $x,y\in\mathbb{R}^d$, we have*

$$
|\nabla f(x)-\nabla f(y)|\le \big(2\|\nabla f\|_\infty+\|\nabla^2 f\|_{\mathrm{HS},\infty}\big)\,|x-y|^{\beta-1}.
$$

Proof. For $|x-y|>1$ we have

$$
|\nabla f(x)-\nabla f(y)|\le 2\|\nabla f\|_\infty \le \|\nabla f\|_\infty|x-y|^{\beta-1},
$$

and for $|x-y|\le 1$ we have

$$
|\nabla f(x)-\nabla f(y)|\le \|\nabla^2 f\|_{\mathrm{HS},\infty}|x-y|\le \|\nabla^2 f\|_{\mathrm{HS},\infty}|x-y|^{\beta-1}. \qquad \square
$$

Lemma 2.6. *Let $(X_t)_{t\ge0}$ and $(Y_k)_{k\ge0}$ be defined by (1.1) and (1.2), respectively. There exists a constant $C>0$ such that for all $x\in\mathbb{R}^d$, $\eta\in(0,1)$, $\beta\in[1,\alpha)$ and $f:\mathbb{R}^d\to\mathbb{R}$ satisfying $\|\nabla f\|_\infty<\infty$ and $\|\nabla^2 f\|_{\mathrm{HS},\infty}<\infty$,*

$$
|P_\eta f(x)-Q_1 f(x)|\le C(1+|x|^\beta)\left(\|\nabla f\|_\infty+\|\nabla^2 f\|_{\mathrm{HS},\infty}\right)\eta^{2+\frac{1}{\alpha}-\frac{1}{\beta}}.
$$

Proof. We use a Taylor expansion to get

$$
\begin{aligned}
&\mathbb{E}f(X_\eta^x)-\mathbb{E}f(Y_1^x)\\
&=\mathbb{E}\left\langle \nabla f(Y_1^x), X_\eta^x-Y_1^x\right\rangle+\mathbb{E}\int_0^1\left\langle \nabla f\left(Y_1^x+r(X_\eta^x-Y_1^x)\right)-\nabla f(Y_1^x), X_\eta^x-Y_1^x\right\rangle\mathrm{d}r\\
&=\mathbb{E}\left\langle \nabla f\left(x+\eta b(x)+Z_\eta\right)-\nabla f\left(x+\eta b(x)\right), X_\eta^x-Y_1^x\right\rangle+\mathbb{E}\left\langle \nabla f\left(x+\eta b(x)\right), X_\eta^x-Y_1^x\right\rangle\\
&\qquad+\mathbb{E}\int_0^1\left\langle \nabla f\left(Y_1^x+r(X_\eta^x-Y_1^x)\right)-\nabla f(Y_1^x), X_\eta^x-Y_1^x\right\rangle\mathrm{d}r\\
&=:\mathsf{I}+\mathsf{II}+\mathsf{III}.
\end{aligned}
$$

For the first term I we have

$$\mathsf{I} = \mathbb{E}\left\langle \nabla f\left(x+\eta b(x)+Z_\eta\right) - \nabla f\left(x+\eta b(x)\right), X_\eta^x - Y_1^x\right\rangle \left(\mathbb{1}_{(0,1]}(|Z_\eta|) + \mathbb{1}_{(1,\infty)}(|Z_\eta|)\right)$$
$$=: \mathsf{I}_1 + \mathsf{I}_2.$$

We need the following estimates for the truncated moment of order $\lambda > \alpha$ and the tail of the α-stable random variable Z_η and $\eta \le 1$:

$$\mathbb{P}\left(|Z_\eta| > 1\right) \le c\eta \quad \text{and} \quad \mathbb{E}\left[|Z_\eta|^\lambda \mathbb{1}_{(0,1]}(Z_\eta)\right] \le C\eta.$$

Both estimates follow from a straightforward calculation using the standard estimate $q_\alpha(\eta,x) \le C\eta/(\eta^{1/\alpha}+|x|)^{\alpha+d}$ for the density of Z_η, see e.g. [3, Theorem 2.1]. Since $\frac{\beta-1}{\beta} > \alpha$, we can use the Hölder inequality and (2.5) to get

$$\begin{aligned}
|\mathsf{I}_1| &\le \mathbb{E}\left[\left|\nabla f\left(x+\eta b(x)+Z_\eta\right) - \nabla f\left(x+\eta b(x)\right)\right| \mathbb{1}_{(0,1]}(|Z_\eta|) \left|X_\eta^x - Y_1^x\right|\right] \\
&\le \|\nabla^2 f\|_{\mathrm{HS},\infty} \mathbb{E}\left[|Z_\eta| \mathbb{1}_{(0,1]}(|Z_\eta|) \left|X_\eta^x - Y_1^x\right|\right] \\
&\le \|\nabla^2 f\|_{\mathrm{HS},\infty} \left(\mathbb{E}\left[|Z_\eta|^{\frac{\beta}{\beta-1}} \mathbb{1}_{(0,1]}(|Z_\eta|)\right]\right)^{\frac{\beta-1}{\beta}} \left(\mathbb{E}|\left|X_\eta^x - Y_1^x\right|^\beta\right)^{\frac{1}{\beta}} \\
&\le C(1+|x|)\|\nabla^2 f\|_{\mathrm{HS},\infty} \eta^{\frac{\beta-1}{\beta}} \eta^{1+\frac{1}{\alpha}} \\
&= C(1+|x|)\|\nabla^2 f\|_{\mathrm{HS},\infty} \eta^{2+\frac{1}{\alpha}-\frac{1}{\beta}},
\end{aligned}$$

whereas by the Hölder inequality

$$\begin{aligned}
|\mathsf{I}_2| &\le \mathbb{E}\left[\left|\nabla f\left(x+\eta b(x)+Z_\eta\right) - \nabla f\left(x+\eta b(x)\right)\right| \mathbb{1}_{(1,\infty)}(|Z_\eta|) \left|X_\eta^x - Y_1^x\right|\right] \\
&\le 2\|\nabla f\|_\infty \mathbb{E}\left[\mathbb{1}_{(1,\infty)}(|Z_\eta|) \left|X_\eta^x - Y_1^x\right|\right] \\
&\le 2\|\nabla f\|_\infty \left(\mathbb{E}\mathbb{1}_{(1,\infty)}(|Z_\eta|)\right)^{\frac{\beta-1}{\beta}} \left(\mathbb{E}\left|X_\eta^x - Y_1^x\right|^\beta\right)^{\frac{1}{\beta}} \\
&\le C(1+|x|)\|\nabla f\|_\infty \eta^{\frac{\beta-1}{\beta}} \eta^{1+\frac{1}{\alpha}} \\
&= C(1+|x|)\|\nabla f\|_\infty \eta^{2+\frac{1}{\alpha}-\frac{1}{\beta}}.
\end{aligned}$$

Hence, we have

$$|\mathsf{I}| \le C(1+|x|)\left(\|\nabla f\|_\infty + \|\nabla^2 f\|_{\mathrm{HS},\infty}\right) \eta^{2+\frac{1}{\alpha}-\frac{1}{\beta}}.$$

For II we use Itô's formula and the definitions (1.1), (1.2) of X_η^x and Y_1^x to see

$$\begin{aligned}
\mathsf{II} &= \mathbb{E}\left\langle \nabla f\left(x+\eta b(x)\right), \int_0^\eta \left[b(X_s^x) - b(x)\right] \mathrm{d}s \right\rangle \\
&= \left\langle \nabla f\left(x+\eta b(x)\right), \int_0^\eta \mathbb{E}\left[b(X_s^x) - b(x)\right] \mathrm{d}s \right\rangle \\
&= \left\langle \nabla f\left(x+\eta b(x)\right), \int_0^\eta \int_0^s \mathbb{E}\left[\langle \nabla b(X_r^x), b(X_r^x)\rangle + (-\Delta)^{\frac{\alpha}{2}} b(X_r^x)\right] \mathrm{d}r\, \mathrm{d}s \right\rangle \\
&\le C\|\nabla f\|_\infty (1+|x|)\eta^2.
\end{aligned}$$

In the last inequality we use the estimate (1.9) (for $b(x_r^x)$) and Lemma 2.3 (for $(-\Delta)^{\alpha/2} b(X_r^x)$), combined with the moment estimate (2.4).

Finally, III is estimated by Lemma 2.5 and (2.5),

$$\begin{aligned} \mathrm{III} &\le C\left(\|\nabla f\|_\infty + \|\nabla^2 f\|_{\mathrm{HS},\infty}\right)\mathbb{E}\left|X_\eta^x - Y_1^x\right|^\beta \\ &\le C(1+|x|^\beta)\left(\|\nabla f\|_\infty + \|\nabla^2 f\|_{\mathrm{HS},\infty}\right)\eta^{\beta+\frac{\beta}{\alpha}} \\ &\le C(1+|x|^\beta)\left(\|\nabla f\|_\infty + \|\nabla^2 f\|_{\mathrm{HS},\infty}\right)\eta^2. \end{aligned}$$

This finishes the proof. □

2.2. Proof of Theorems 1.2 and 1.3.

Proof of Theorem 1.2. Thanks to the discrete version of the classical Duhamel principle, it is easy to check that for $h \in \mathrm{Lip}(1)$

$$P_{N\eta}h(x) - \tilde{Q}_N h(x) = \sum_{i=1}^N \tilde{Q}_{i-1}(P_\eta - \tilde{Q}_1)P_{(N-i)\eta}h(x). \tag{2.8}$$

Then we have

$$\begin{aligned} W_1\left(\mathrm{law}(X_{\eta N}), \mathrm{law}(\tilde{Y}_N)\right) &= \sup_{h\in\mathrm{Lip}(1)} |P_{N\eta}h(x) - \tilde{Q}_N h(x)| \\ &\le \sum_{i=1}^{N-1} \sup_{h\in\mathrm{Lip}(1)} \left|\tilde{Q}_{i-1}(P_\eta - \tilde{Q}_1)P_{(N-i)\eta}h(x)\right| \\ &\quad + \sup_{h\in\mathrm{Lip}(1)} \left|\tilde{Q}_{N-1}(P_\eta - \tilde{Q}_1)h(x)\right|. \end{aligned} \tag{2.9}$$

First, we bound the last term. By (2.4) with $\beta = 1$, (1.4) and (1.8), for $h \in \mathrm{Lip}(1)$ and $\eta < 1$,

$$\begin{aligned} \left|(P_\eta - \tilde{Q}_1)h(x)\right| &= \left|\mathbb{E}h(X_\eta^x) - \mathbb{E}h(\tilde{Y}_1)\right| \\ &\le \left|\mathbb{E}h(X_\eta^x) - h(x)\right| + \left|\mathbb{E}h(\tilde{Y}_1) - h(x)\right| \\ &\le \mathbb{E}|X_\eta^x - x| + \mathbb{E}|\tilde{Y}_1 - x| \\ &\le C(1+|x|)\,\eta^{1/\alpha} + \eta\,|b(x)| + \sigma^{-1}\eta^{1/\alpha}\mathbb{E}|\widetilde{Z}_1| \\ &\le C(1+|x|)\,\eta^{1/\alpha} + \eta^{1/\alpha}(|b(0)| + \theta_2|x|) + \sigma^{-1}\eta^{1/\alpha}\mathbb{E}|\widetilde{Z}_1| \\ &\le C(1+|x|)\,\eta^{1/\alpha}. \end{aligned}$$

Together with (1.15) we get

$$\sup_{h\in\mathrm{Lip}(1)} \left|\tilde{Q}_{N-1}(P_\eta - \tilde{Q}_1)h(x)\right| \le C(1+\mathbb{E}|\tilde{Y}_{N-1}^x|)\eta^{1/\alpha} \le C(1+|x|)\eta^{2/\alpha-1}. \tag{2.10}$$

Next, we bound the first term in (2.9); we distinguish between two cases:
Case 1: $N \le \eta^{-1} + 1$. By Lemmas 2.4 and 2.1,

$$\begin{aligned} \left|(P_\eta - \tilde{Q}_1)P_{(N-i)\eta}h(x)\right| &\le C(1+|x|)\left(\|\nabla P_{(N-i)\eta}h\|_\infty + \|\nabla^2 P_{(N-i)\eta}h\|_{\mathrm{HS},\infty}\right)\eta^{2/\alpha} \\ &\le C(1+|x|)[(N-i)\eta]^{-1/\alpha}\eta^{2/\alpha}. \end{aligned}$$

Combining this with (1.15), we get

$$\begin{aligned} \sup_{h\in\mathrm{Lip}(1)} \left|Q_{i-1}(P_\eta - \tilde{Q}_1)P_{(N-i)\eta}h(x)\right| &\le C(1+\mathbb{E}|\tilde{Y}_{i-1}^x|)[(N-i)\eta]^{-1/\alpha}\eta^{2/\alpha} \\ &\le C(1+|x|)[(N-i)\eta]^{-1/\alpha}\eta^{2/\alpha}. \end{aligned} \tag{2.11}$$

Since $N-1 \le \eta^{-1}$,

$$\sum_{i=1}^{N-1}[(N-i)\eta]^{-1/\alpha} = \eta^{-1/\alpha}\sum_{i=1}^{N-1} i^{-1/\alpha} \le \eta^{-1/\alpha}\int_0^{N-1} r^{-1/\alpha}\,\mathrm{d}r$$
$$= \frac{\alpha}{\alpha-1}\,\eta^{-1/\alpha}(N-1)^{-1/\alpha+1} \le \frac{\alpha}{\alpha-1}\,\eta^{-1}.$$

This gives the upper bound

$$\sum_{i=1}^{N-1}\sup_{h\in \mathrm{Lip}(1)}\left|\tilde{Q}_{i-1}\big(P_\eta-\tilde{Q}_1\big)P_{(N-i)\eta}h(x)\right| \le C(1+|x|)\eta^{2/\alpha}\sum_{i=1}^{N-1}[(N-i)\eta]^{-1/\alpha}$$
$$\le C\frac{\alpha}{\alpha-1}\,(1+|x|)\eta^{2/\alpha-1}.$$

Case 2: $N>\eta^{-1}+1$. By Proposition 1.6, for any $x,y\in\mathbb{R}^d$, there exist constants $C>0$ and $\lambda>0$ such that

$$|P_t h(x)-P_t h(y)| \le C\mathrm{e}^{-\lambda t}|x-y|, \quad h\in \mathrm{Lip}(1),\ t\ge 0.$$

This implies that

$$\sup_{h\in \mathrm{Lip}(1)}\left|Q_{i-1}\big(P_\eta-\tilde{Q}_1\big)P_{(N-i)\eta}h(x)\right| = \sup_{h\in \mathrm{Lip}(1)}\left|\tilde{Q}_{i-1}\big(P_\eta-\tilde{Q}_1\big)P_1P_{(N-i)\eta-1}h(x)\right|$$
$$\le C\mathrm{e}^{-\lambda[(N-i)\eta-1]}\sup_{g\in \mathrm{Lip}(1)}\left|\tilde{Q}_{i-1}\big(P_\eta-\tilde{Q}_1\big)P_1 g(x)\right|,$$

where $i\le \lfloor N-\eta^{-1}\rfloor$. By Lemmas 2.4 and 2.1,

$$\left|\big(P_\eta-\tilde{Q}_1\big)P_1 g(x)\right| \le C(1+|x|)\left(\|\nabla P_1 g\|_\infty+\|\nabla^2 P_1 g\|_{\mathrm{HS},\infty}\right)\eta^{2/\alpha} \le C(1+|x|)\,\eta^{2/\alpha}.$$

Combining this with (1.15), we get

$$\sum_{i=1}^{\lfloor N-\eta^{-1}\rfloor}\sup_{h\in \mathrm{Lip}(1)}\left|\tilde{Q}_{i-1}\big(P_\eta-\tilde{Q}_1\big)P_{(N-i)\eta}h(x)\right| \le C\eta^{2/\alpha}\sum_{i=1}^{\lfloor N-\eta^{-1}\rfloor}\mathrm{e}^{-\lambda[(N-i)\eta-1]}\left(1+\mathbb{E}|\tilde{Y}^x_{i-1}|\right)$$
$$\le C(1+|x|)\eta^{2/\alpha}\sum_{i=1}^{\lfloor N-\eta^{-1}\rfloor}\mathrm{e}^{-\lambda[(N-i)\eta-1]}.$$

Observe that

$$\sum_{i=1}^{\lfloor N-\eta^{-1}\rfloor}\mathrm{e}^{-\lambda[(N-i)\eta-1]} = \sum_{i=\lfloor\eta^{-1}\rfloor}^{N-1}\mathrm{e}^{-\lambda(i\eta-1)} \le \mathrm{e}^{\lambda}\int_{\lfloor\eta^{-1}\rfloor-1}^{N-1}\mathrm{e}^{-\lambda\eta r}\,\mathrm{d}r$$
$$\le \mathrm{e}^{\lambda}\eta^{-1}\int_0^\infty \mathrm{e}^{-\lambda r}\,\mathrm{d}r = \lambda^{-1}\mathrm{e}^{\lambda}\eta^{-1}.$$

Thus, we get

$$\sum_{i=1}^{\lfloor N-\eta^{-1}\rfloor}\sup_{h\in \mathrm{Lip}(1)}\left|\tilde{Q}_{i-1}\big(P_\eta-\tilde{Q}_1\big)P_{(N-i)\eta}h(x)\right| \le C\lambda^{-1}\mathrm{e}^{\lambda}(1+|x|)\eta^{2/\alpha-1}.$$

For $i\ge\lfloor N-\eta^{-1}\rfloor+1$, by almost the same as the calculation in the first case, we find

$$\sum_{i=\lfloor N-\eta^{-1}\rfloor+1}^{N-1}[(N-i)\eta]^{-1/\alpha} \le \frac{\alpha}{\alpha-1}\,\eta^{-1}.$$

Combining this with (2.11), we obtain

$$\sum_{i=\lfloor N-\eta^{-1}\rfloor+1}^{N-1} \sup_{h\in\mathrm{Lip}(1)} \left|\tilde{Q}_{i-1}\big(P_\eta - \tilde{Q}_1\big)P_{(N-i)\eta}h(x)\right|$$
$$\leq C(1+|x|)\,\eta^{2/\alpha} \sum_{i=\lfloor N-\eta^{-1}\rfloor+1}^{N-1} [(N-i)\eta]^{-1/\alpha} \leq C\frac{\alpha}{\alpha-1}\,(1+|x|)\,\eta^{2/\alpha-1}.$$

We have just shown estimates for $\sum_{i=1}^{\lfloor N-\eta^{-1}\rfloor}\ldots$ and $\sum_{i=\lfloor N-\eta^{-1}\rfloor+1}^{N-1}\ldots$. Adding them up we arrive at

$$\sum_{i=1}^{N-1} \sup_{h\in\mathrm{Lip}(1)} \left|\tilde{Q}_{i-1}\big(P_\eta - \tilde{Q}_1\big)P_{(N-i)\eta}h(x)\right|$$
$$= \left(\sum_{i=1}^{\lfloor N-\eta^{-1}\rfloor} + \sum_{i=\lfloor N-\eta^{-1}\rfloor+1}^{N-1}\right) \sup_{h\in\mathrm{Lip}(1)} \left|\tilde{Q}_{i-1}\big(P_\eta - \tilde{Q}_1\big)P_{(N-i)\eta}h(x)\right| \leq C(1+|x|)\eta^{2/\alpha-1}.$$

Both Case 1 and Case 2 lead to an estimate of the form

$$\sum_{i=1}^{N-1} \sup_{h\in\mathrm{Lip}(1)} \left|\tilde{Q}_{i-1}\big(P_\eta - \tilde{Q}_1\big)P_{(N-i)\eta}h(x)\right| \leq C(1+|x|)\,\eta^{2/\alpha-1}.$$

Substituting this and (2.10) into (2.9), the first assertion of Theorem 1.2 follows.

It remains to prove Part (2). It is easy to see from (1.10) and (1.13), that we have

$$\lim_{k\to\infty} W_1\big(\mu, \mathrm{law}(X_{\eta k})\big) = \lim_{k\to\infty} W_1\big(\mathrm{law}(\tilde{Y}_k), \tilde{\mu}_\eta\big) = 0.$$

By the triangle inequality and Part (1) with $x=0$,

$$\begin{aligned} W_1\big(\mu, \tilde{\mu}_\eta\big) &\leq W_1\big(\mu, \mathrm{law}(X_{\eta N})\big) + W_1\big(\mathrm{law}(X_{\eta N}), \mathrm{law}(\tilde{Y}_N)\big) + W_1\big(\mathrm{law}(\tilde{Y}_N), \tilde{\mu}_\eta\big) \\ &\leq W_1\big(\mu, \mathrm{law}(X_{\eta N})\big) + C\,\eta^{2/\alpha-1} + W_1\big(\mathrm{law}(\tilde{Y}_N), \tilde{\mu}_\eta\big). \end{aligned}$$

Letting $N\to\infty$ finishes the proof. □

Proof of Theorem 1.3. In the above proof, replacing Lemma 2.4 and (1.13) with Lemma 2.6 and (1.12), the proof of Theorem 1.3 is similar to that of Theorem 1.2. □

3. Malliavin calculus and the proof of Lemma 2.1

3.1. Jacobi flow associated with the SDE (1.1). The *Jacobian flow* is the derivative of X_t^x with respect to the initial value x; the Jacobian flow in direction $v\in\mathbb{R}^d$ is defined by

$$\nabla_v X_t^x := \lim_{\epsilon\to0}\frac{X_t^{x+\epsilon v}-X_t^x}{\epsilon}, \quad t\geq 0.$$

This limit exists and satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\nabla_v X_t^x = \nabla_{\nabla_v X_t^x} b(X_t^x), \quad \nabla_v X_0^x = v. \tag{3.1}$$

Similarly, for $v_1,v_2\in\mathbb{R}^d$, we can define $\nabla_{v_2}\nabla_{v_1}X_t^x$, which satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\nabla_{v_2}\nabla_{v_1}X_t^x = \nabla_{\nabla_{v_2}\nabla_{v_1}X_t^x} b(X_t^x) + \nabla_{\nabla_{v_2}X_t^x}\nabla_{\nabla_{v_1}X_t^x} b(X_t^x), \quad \nabla_{v_1}\nabla_{v_1}X_0^x = 0. \tag{3.2}$$

Then, we first have the following estimates of $\nabla_{v_1}X_t^x$ and $\nabla_{v_2}\nabla_{v_1}X_t^x$.

Lemma 3.1. *For any starting point $x \in \mathbb{R}^d$ and all directions $v_1, v_2 \in \mathbb{R}^d$ the following (deterministic) estimates hold:*

$$|\nabla_{v_1} X_t^x| \le \mathrm{e}^{\theta_2} |v_1| , \qquad t \in (0,1], \tag{3.3}$$

$$|\nabla_{v_2}\nabla_{v_1} X_t^x| \le \frac{\theta_3}{2\sqrt{2\theta_2}} \mathrm{e}^{4\theta_2} |v_1| \, |v_2| , \qquad t \in (0,1]. \tag{3.4}$$

Proof. By (3.1) and (1.8), we have

$$\frac{\mathrm{d}}{\mathrm{d}t} |\nabla_{v_1} X_t^x|^2 = 2 \left\langle \nabla_{v_1} X_t^x, \nabla_{\nabla_{v_1} X_t^x} b(X_t^x) \right\rangle \le 2\theta_2 |\nabla_{v_1} X_t^x|^2,$$

and Gronwall's inequality yields for $t \in (0,1]$

$$|\nabla_{v_1} X_t^x|^2 \le \mathrm{e}^{2\theta_2 t} |v_1|^2 \le \mathrm{e}^{2\theta_2} |v_1|^2 .$$

This proves the first assertion. Writing $\zeta(t) := \nabla_{v_2}\nabla_{v_1} X_t^x$, we see from (3.2), (1.8), (3.3), the Cauchy–Schwarz inequality, and the elementary estimate $2AB \le A^2 + B^2$ that

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} |\zeta(t)|^2 &= 2 \left\langle \zeta(t), \nabla_{\zeta(t)} b(X_t^x) \right\rangle + 2 \left\langle \zeta(t), \nabla_{\nabla_{v_2} X_t^x} \nabla_{\nabla_{v_1} X_t^x} b(X_t^x) \right\rangle \\
&\le 2\theta_2 |\zeta(t)|^2 + 2\theta_3 \mathrm{e}^{2\theta_2} |v_1| \, |v_2| \, |\zeta(t)| \\
&\le 4\theta_2 |\zeta(t)|^2 + \frac{\theta_3^2}{2\theta_2} \mathrm{e}^{4\theta_2} |v_1|^2 |v_2|^2 .
\end{aligned}$$

Since $\zeta(0) = 0$, we can use again Gronwall's inequality and get for all $t \in (0,1]$

$$|\zeta(t)|^2 \le \frac{\theta_3^2}{2\theta_2} \mathrm{e}^{4\theta_2} |v_1|^2 |v_2|^2 \int_0^t \mathrm{e}^{4\theta_2 (t-s)} \, \mathrm{d}s \le \frac{\theta_3^2}{8\theta_2^2} \mathrm{e}^{8\theta_2} |v_1|^2 |v_2|^2 . \qquad \square$$

3.2. **Bismut's formula.** (See also [33]). Let $u \in L^2_{loc}([0,\infty) \times (\Omega, \mathcal{F}, \mathbb{P}); \mathbb{R}^d)$, i.e. we have $\mathbb{E} \int_0^t |u(s)|^2 \, \mathrm{d}s < \infty$ for all $t > 0$. Let $\{W_t\}_{t \ge 0}$ be a d-dimensional standard Brownian motion and assume that u is adapted to the filtration $(\mathcal{F}_t)_{t \ge 0}$ with $\mathcal{F}_t := \sigma(W_s : 0 \le s \le t)$; i.e. $u(t)$ is $\mathcal{F}_t$ measurable for $t \ge 0$. Define

$$U = \int_0^{\bullet} u(s) \, \mathrm{d}s. \tag{3.5}$$

For a $t > 0$, let $F_t : C([0,t], \mathbb{R}^d) \to \mathbb{R}^m$ be an $\mathcal{F}_t$ measurable map. If the following limit exists

$$D_U F_t(W) = \lim_{\epsilon \to 0} \frac{F_t(W + \epsilon U) - F_t(W)}{\epsilon}$$

in $L^2((\Omega, \mathcal{F}, \mathbb{P}); \mathbb{R}^m)$, then $F_t(W)$ is said to be m-dimensional *Malliavin differentiable* and $D_U F_t(W)$ is called the *Malliavin derivative* of $F_t(W)$ in the direction U.

Let $\phi \in \mathcal{C}_b^2(\mathbb{R}^d, \mathbb{R})$ and both $F_t(W)$ and $G_t(W)$ be d-dimensional Malliavin differentiable functionals. Then we have the following product and chain rules:

$$D_U \big(\langle F_t(W), G_t(W) \rangle \big) = \langle D_U F_t(W), G_t(W) \rangle + \langle F_t(W), D_U G_t(W) \rangle,$$

and

$$D_U \nabla \phi(F_t(W)) = \nabla_{D_U F_t(W)} \nabla \phi(F_t(W)).$$

The following integration by parts formula is often called *Bismut's formula*. For a Malliavin differentiable $F_t(W)$ such that $F_t(W), D_U F_t(W) \in L^2((\Omega, \mathcal{F}, \mathbb{P}); \mathbb{R})$, we have

$$\mathbb{E}[D_U F_t(W)] = \mathbb{E}\left[F_t(W) \int_0^t \langle u(s), \mathrm{d}W_s \rangle \right]. \tag{3.6}$$

3.3. **Time-change method for the SDE** (1.1). (See also [49]). We will now turn to the SDE driven by a rotationally invariant α-stable Lévy process with $\alpha \in (1,2)$. We can express such drivers as subordinated Brownian motion. More precisely, let $\{S_t\}_{t\geq 0}$ be an independent $\frac{\alpha}{2}$-stable subordinator. Then, $Z_t := W_{S_t}$ is a rotationally invariant α-stable Lévy process, see e.g. [40]. This means that we can re-write (1.1) as

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \mathrm{d}W_{S_t}, \quad X_0 = x. \tag{3.7}$$

Let $\mathbb{W}$ be the space of all continuous functions from $[0,\infty)$ to $\mathbb{R}^d$ vanishing at $t = 0$; we equip $\mathbb{W}$ with the topology of locally uniform convergence, and the Wiener measure $\mu_{\mathbb{W}}$; therefore, the coordinate process

$$W_t(w) = w_t$$

are a standard d-dimensional Brownian motion. Let $\mathbb{S}$ be the space of all increasing, càdlàg (right continuous with finite left limits) functions from $[0,\infty)$ to $[0,\infty)$ vanishing at $t = 0$; we equip $\mathbb{S}$ with the Skorohod metric and the probability measure $\mu_{\mathbb{S}}$ so that for any $l \in \mathbb{S}$ the coordinate process

$$S_t(l) := l_t$$

is an $\frac{\alpha}{2}$-stable subordinator. On the product measure space

$$(\Omega, \mathcal{F}, \mathbb{P}) := (\mathbb{W} \times \mathbb{S}, \mathcal{B}(\mathbb{W}) \otimes \mathcal{B}(\mathbb{S}), \mu_{\mathbb{W}} \times \mu_{\mathbb{S}}),$$

we define

$$L_t(w,l) := w_{l_t}.$$

The process $\{L_t\}_{t\geq 0}$ is a rotationally invariant α-stable Lévy process on $(\Omega, \mathcal{F}, \mathbb{P})$. We will use the following two natural filtrations associated with the Lévy process L_t and the Brownian motion W_t:

$$\mathcal{F}_t := \sigma\{L_s(w,l); s \leq t\} \quad \text{and} \quad \mathcal{F}_t^{\mathbb{W}} := \sigma\{W_s(w); s \leq t\}.$$

In particular, we can regard the solution X_t^x of the SDE (3.7) as an $(\mathcal{F}_t)$-adapted functional on Ω, and therefore,

$$\mathbb{E} f(X_t^x) = \int_{\mathbb{S}} \int_{\mathbb{W}} f(X_t^x(w_l))\, \mu_{\mathbb{W}}(\mathrm{d}w)\, \mu_{\mathbb{S}}(\mathrm{d}l).$$

For every fixed $l \in \mathbb{S}$, we denote by X_t^l the solution to the SDE

$$\mathrm{d}X_t^l = b(X_t^l)\,\mathrm{d}t + \mathrm{d}W_{l_t}, \quad X_0^l = x. \tag{3.8}$$

We will now fix a path $l \in \mathbb{S}$, and consider the SDE (3.8). Unless otherwise mentioned, all expectations are taken with respect to the Wiener space $(\mathbb{W}, \mathcal{B}(\mathbb{W}), \mu_{\mathbb{W}})$. First of all, notice that $t \to W_{l_t}$ is a centered Gaussian process with independent increments. In particular, W_{l_t} is a càdlàg $\mathcal{F}_{l_t}^{\mathbb{W}}$-martingale. Thus, under **Assumption A**, it is well known that for each $x \in \mathbb{R}^d$, the SDE (3.8) admits a unique càdlàg $\mathcal{F}_{l_t}^{\mathbb{W}}$-adapted solution $X_t^{x;l}$, see e.g. [37, p.249, Theorem 6].

The main aim of this section is to establish the following result:

Lemma 3.2. *Under **Assumption A** one has for all functions $\phi \in C_b^2(\mathbb{R}^d, \mathbb{R})$, all directions $v_1, v_2 \in \mathbb{R}^d$ and $x \in \mathbb{R}^d$, $t \in (0,1]$*

$$|\nabla_{v_1} X_t^{x;l}| \leq \mathrm{e}^{\theta_2}\, |v_1| \tag{3.9}$$

and

$$\left|\mathbb{E}\left[\nabla_{\nabla_{v_2}X_t^{x;l}}\nabla_{\nabla_{v_1}X_t^{x;l}}\phi(X_t^{x;l})\right]\right|$$
$$\le \left|\mathbb{E}\left[\frac{1}{l_t}\nabla_{\nabla_{v_1}X_t^{x;l}}\phi(X_t^{x;l})\int_0^t \langle \nabla_{v_2}X_s^{x;l}, \mathrm{d}W_{l_s}\rangle\right]\right| + \|\nabla\phi\|_\infty \frac{\theta_3}{\sqrt{2\theta_2}}\,\mathrm{e}^{2\theta_2}\,|v_1|\,|v_2|\,,$$

where $\nabla_{v_i}X_t^{x;l}$ $(i=1,2)$ *is determined by the following linear equation:*

$$\frac{\mathrm{d}}{\mathrm{d}t}\nabla_{v_i}X_t^{x;l} = \nabla_{\nabla_{v_i}X_t^{x;l}}b(X_t^{x;l}), \quad \nabla_{v_i}X_0^{x;l} = v_i. \tag{3.10}$$

In order to prove Lemma 3.2, we use a time-change argument to transform the SDE (3.8) into an SDE driven by a standard Brownian motion; this allows us to use Bismut's formula (3.6). For every $\epsilon \in (0,1)$ we define

$$l_t^\epsilon := \frac{1}{\epsilon}\int_t^{t+\epsilon} l_s\,\mathrm{d}s + \epsilon t = \int_0^1 l_{\epsilon s+t}\,\mathrm{d}s + \epsilon t.$$

Since $t \mapsto l_t$ is increasing and right continuous, it follows that for each $t \ge 0$,

$$l_t^\epsilon \downarrow l_t \quad \text{as} \quad \epsilon \downarrow 0.$$

Moreover, $t \mapsto l_t^\epsilon$ is absolutely continuous and strictly increasing. Let γ^ϵ be the inverse function of l^ϵ, i.e.

$$l^\epsilon_{\gamma^\epsilon_t} = t, \quad t \ge l_0^\epsilon \quad \text{and} \quad \gamma^\epsilon_{l^\epsilon_t} = t, \quad t \ge 0.$$

By definition, γ_t^ϵ is absolutely continuous on $[l_0^\epsilon, \infty)$. Let $X_t^{x;\ell^\epsilon}$ be the solution to the SDE

$$\mathrm{d}X_t^{x;\ell^\epsilon} = b(X_t^{x;\ell^\epsilon})\,\mathrm{d}t + \mathrm{d}W_{l_t^\epsilon - l_0^\epsilon}, \quad X_0^{x;\ell^\epsilon} = x.$$

Let us now define

$$Y_t^{x;l^\epsilon} := X^{x;l^\epsilon}_{\gamma^\epsilon_t}, \quad t \ge l_0^\epsilon.$$

Changing variables in (3.8) we see that for $t \ge l_0^\epsilon$,

$$Y_t^{x;l^\epsilon} = x + \int_0^{\gamma_t^\epsilon} b\left(X_s^{x;l^\epsilon}\right)\,\mathrm{d}s + W_t = x + \int_{l_0^\epsilon}^t b\left(Y_s^{x;l^\epsilon}\right)\dot\gamma_s^\epsilon\,\mathrm{d}s + W_t \tag{3.11}$$

($\dot\gamma_s^\epsilon$ denotes the derivative in s). Hence, for any vector $v \in \mathbb{R}^d$, we have

$$\nabla_v Y_t^{x;l^\epsilon} = v + \int_{l_0^\epsilon}^t \nabla_{\nabla_v Y_s^{x;l^\epsilon}} b\left(Y_s^{x;l^\epsilon}\right)\,\dot\gamma_s^\epsilon\,\mathrm{d}s, \tag{3.12}$$

and the differential form can be written as

$$\frac{\mathrm{d}}{\mathrm{d}t}\nabla_v Y_t^{x;l^\epsilon} = \nabla b\left(Y_t^{x;l^\epsilon}\right)\,\dot\gamma_t^\epsilon \nabla_v Y_t^{x;l^\epsilon}, \quad t \ge l_0^\epsilon,$$

which has a solution of the form

$$\nabla_v Y_t^{x;l^\epsilon} = J^{x;l^\epsilon}_{l_0^\epsilon,t}v, \tag{3.13}$$

involving a matrix exponential

$$J_{s,t}^{x;l^\epsilon} = \exp\left[\int_s^t \nabla b\left(Y_s^{x;l^\epsilon}\right)\dot\gamma_s^\epsilon\,\mathrm{d}s\right], \quad l_0^\epsilon \le s \le t < \infty. \tag{3.14}$$

It is easy to see that $J_{s,t}^{x;l^\epsilon} J_{l_0^\epsilon,s}^{x;l^\epsilon} = J_{l_0^\epsilon,t}^{x;l^\epsilon}$ for all $l_0^\epsilon \le s \le t < \infty$.

Now, we come back to the Malliavin calculus from Section 3.2. Fixing $t \ge l_0^\epsilon$ and $x \in \mathbb{R}^d$, the solution $Y_t^{x;l^\epsilon}$ is a d-dimensional functional of Brownian motion $\{W_s\}_{l_0^\epsilon \le s \le t}$.

Let U be as in Section 3.2. The Malliavin derivative of $Y_t^{x;l^\epsilon}$ in direction U exists in $L^2\left((\mathbb{W},\mathcal{B}(\mathbb{W}),\mu_{\mathbb{W}});\mathbb{R}^d\right)$ and is given by

$$D_U Y_t^{x;l^\epsilon}(W) = \lim_{\delta\to 0} \frac{Y_t^{x;l^\epsilon}(W+\delta U) - Y_t^{x;l^\epsilon}(W)}{\delta}.$$

To simplify notation, we drop the W in $D_U Y_t^{x;l^\epsilon}(W)$ and write $D_U Y_t^{x;l^\epsilon} = D_U Y_t^{x;l^\epsilon}(W)$. By (3.11), it satisfies the equation

$$D_U Y_t^{x;l^\epsilon} = \int_{l_0^\epsilon}^t \left(\nabla_{D_U Y_s^{x;l^\epsilon}} b\left(Y_s^{x;l^\epsilon}\right) \dot{\gamma}_s^\epsilon + u(s)\right) \mathrm{d}s,$$

the differential form of the above equation can be written as

$$\frac{\mathrm{d}}{\mathrm{d}t} D_U Y_t^{x;l^\epsilon} = \nabla b\left(Y_t^{x;l^\epsilon}\right) \dot{\gamma}_t^\epsilon D_U Y_t^{x;l^\epsilon} + u(t), \quad t \geq l_0^\epsilon,$$

and this equation has a unique solution which is given via the matrix exponential (3.14):

$$D_U Y_t^{x;l^\epsilon} = \int_{l_0^\epsilon}^t J_{s,t}^{x;l^\epsilon} u(s)\,\mathrm{d}s. \tag{3.15}$$

For a fixed $t>0$, for any $v_1, v_2, x \in \mathbb{R}^d$, we define $u_i, U_i : [l_0^\epsilon, t] \to \mathbb{R}^d$ by

$$u_i(s) := \frac{1}{t - l_0^\epsilon} \nabla_{v_i} Y_s^{x;l^\epsilon}, \quad U_{i;s} := \int_0^s u_i(r)\,\mathrm{d}r \tag{3.16}$$

for $l_0^\epsilon \leq s \leq t$ and $i = 1,2$. Then

$$D_{U_i} Y_s^{x;l^\epsilon} = \frac{s - l_0^\epsilon}{t - l_0^\epsilon} \nabla_{v_i} Y_s^{x;l^\epsilon}, \quad l_0^\epsilon \leq s \leq t. \tag{3.17}$$

In addition, (3.12) implies that for $s \in [l_0^\epsilon, t]$

$$D_{U_2} \nabla_{v_1} Y_s^{x;l^\epsilon} = \int_{l_0^\epsilon}^s \left(\nabla_{D_{U_2} Y_r^{x;l^\epsilon}} \nabla_{\nabla_{v_1} Y_r^{x;l^\epsilon}} b\left(Y_r^{x;l^\epsilon}\right) + \nabla_{D_{U_2}\nabla_{v_1} Y_r^{x;l^\epsilon}} b\left(Y_r^{x;l^\epsilon}\right)\right) \dot{\gamma}_r^\epsilon \,\mathrm{d}r. \tag{3.18}$$

The following lemma contains the upper bounds on the derivatives.

Lemma 3.3. *Let $v_1, v_2, x \in \mathbb{R}^d$ and $t \in (0,1]$. Then,*

$$|\nabla_{v_i} Y_s^{x;l^\epsilon}| \leq \mathrm{e}^{\theta_2 \gamma_s^\epsilon} |v_i| \tag{3.19}$$

and

$$|D_{U_2} \nabla_{v_1} Y_s^{x;l^\epsilon}| \leq \frac{\theta_3}{\sqrt{2\theta_2}} \mathrm{e}^{2\theta_2 \gamma_s^\epsilon} \sqrt{\gamma_s^\epsilon}\, |v_1|\,|v_2|\,. \tag{3.20}$$

for $l_0^\epsilon \leq s \leq t$ and $i = 1,2$.

Proof. Recall that $\theta_1 > 0$ and $\dot{\gamma}_s^\epsilon \geq 0$. By (3.12) and (1.8), we have for any $l_0^\epsilon \leq s \leq t$

$$\frac{\mathrm{d}}{\mathrm{d}s} |\nabla_{v_i} Y_s^{x;l^\epsilon}|^2 = 2\dot{\gamma}_s^\epsilon \left\langle \nabla_{v_i} Y_s^{x;l^\epsilon},\ \nabla_{\nabla_{v_i} Y_s^{x;l^\epsilon}} b\left(Y_s^{x;l^\epsilon}\right)\right\rangle \leq 2\theta_2 \dot{\gamma}_s^\epsilon |\nabla_{v_i} Y_s^{x;l^\epsilon}|^2,$$

and this implies, because of Gronwall's lemma,

$$|\nabla_{v_i} Y_s^{x;l^\epsilon}|^2 \leq \exp\left[2\theta_2 \int_{l_0^\epsilon}^s \dot{\gamma}_r^\epsilon \,\mathrm{d}r\right] |v_i|^2 = \mathrm{e}^{2\theta_2(\gamma_s^\epsilon - \gamma_{l_0^\epsilon}^\epsilon)} |v_i|^2 = \mathrm{e}^{2\theta_2 \gamma_s^\epsilon} |v_i|^2.$$

Using (3.18) and (1.8) we find for any $l_0^\epsilon \le s \le t$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}|D_{U_2}\nabla_{v_1}Y_s^{x;l^\epsilon}|^2 &= 2\dot\gamma_s^\epsilon\left\langle D_{U_2}\nabla_{v_1}Y_s^{x;l^\epsilon},\ \nabla_{D_{U_2}\nabla_{v_1}Y_s^{x;l^\epsilon}}b\left(Y_s^{x;l^\epsilon}\right)\right\rangle\\ &\quad+2\dot\gamma_s^\epsilon\left\langle D_{U_2}\nabla_{v_1}Y_s^{x;l^\epsilon},\ \nabla_{D_{U_2}Y_s^{x;l^\epsilon}}\nabla_{\nabla_{v_1}Y_s^{x;l^\epsilon}}b\left(Y_s^{x;l^\epsilon}\right)\right\rangle\\ &\le 2\theta_2\dot\gamma_s^\epsilon\left|D_{U_2}\nabla_{v_1}Y_s^{x;l^\epsilon}\right|^2+2\theta_3\dot\gamma_s^\epsilon\left|D_{U_2}\nabla_{v_1}Y_s^{x;l^\epsilon}\right|\left|D_{U_2}Y_s^{x;l^\epsilon}\right|\left|\nabla_{v_1}Y_s^{x;l^\epsilon}\right|.\end{aligned}$$

Now (3.17) and (3.19) imply

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\left|D_{U_2}\nabla_{v_1}Y_s^{x;l^\epsilon}\right|^2 &\le 2\theta_2\dot\gamma_s^\epsilon\left|D_{U_2}\nabla_{v_1}Y_s^{x;l^\epsilon}\right|^2+2\theta_3\dot\gamma_s^\epsilon\left|D_{U_2}\nabla_{v_1}Y_s^{x;l^\epsilon}\right|\left|\nabla_{v_2}Y_s^{x;l^\epsilon}\right|\left|\nabla_{v_1}Y_s^{x;l^\epsilon}\right|\\ &\le 4\theta_2\dot\gamma_s^\epsilon\left|D_{U_2}\nabla_{v_1}Y_s^{x;l^\epsilon}\right|^2+\frac{\theta_3^2}{2\theta_2}\dot\gamma_s^\epsilon\left|\nabla_{v_2}Y_s^{x;l^\epsilon}\right|^2\left|\nabla_{v_1}Y_s^{x;l^\epsilon}\right|^2\\ &\le 4\theta_2\dot\gamma_s^\epsilon\left|D_{U_2}\nabla_{v_1}Y_s^{x;l^\epsilon}\right|^2+\frac{\theta_3^2}{2\theta_2}\dot\gamma_s^\epsilon \mathrm{e}^{4\theta_2\gamma_s^\epsilon}\left|v_1\right|^2\left|v_2\right|^2.\end{aligned}$$

Since $D_{U_2}\nabla_{v_1}Y_{l_0^\epsilon}^{x;l^\epsilon}=0$, we can use Gronwall's inequality to see

$$\left|D_{U_2}\nabla_{v_1}Y_s^{x;l^\epsilon}\right|^2\le\frac{\theta_3^2}{2\theta_2}\left|v_1\right|^2\left|v_2\right|^2\int_{l_0^\epsilon}^s\dot\gamma_u^\epsilon\mathrm{e}^{4\theta_2\gamma_u^\epsilon}\mathrm{e}^{4\theta_2(\gamma_s^\epsilon-\gamma_u^\epsilon)}\,\mathrm{d}u=\frac{\theta_3^2}{2\theta_2}|v_1|^2|v_2|^2\mathrm{e}^{4\theta_2\gamma_s^\epsilon}\gamma_s^\epsilon.\qquad\square$$

Proof of Lemma 3.2. From (1.8) we see that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left|\nabla_{v_1}X_t^{x;l}\right|^2=2\left\langle\nabla_{v_1}X_t^{x;l},\ \nabla_{\nabla_{v_1}X_t^{x;l}}b(X_t^{x;l})\right\rangle\le 2\theta_2\left|\nabla_{v_1}X_t^{x;l}\right|^2.$$

This yields for all $t\in(0,1]$

$$\left|\nabla_{v_1}X_t^{x;l}\right|^2\le\mathrm{e}^{2\theta_2 t}\left|v_1\right|^2\le\mathrm{e}^{2\theta_2}\left|v_1\right|^2,$$

i.e. (3.9) holds.

Using (3.17) with $s=t$ and $i=2$, the chain rule and Bismut's formula (3.6), we see that

$$\begin{aligned}&\mathbb{E}\left[\nabla_{\nabla_{v_2}Y_t^{x;l^\epsilon}}\nabla_{\nabla_{v_1}Y_t^{x;l^\epsilon}}\phi\big(Y_t^{x;l^\epsilon}\big)\right]\\ &=\mathbb{E}\left[\nabla_{D_{U_2}Y_t^{x;l^\epsilon}}\nabla_{\nabla_{v_1}Y_t^{x;l^\epsilon}}\phi\big(Y_t^{x;l^\epsilon}\big)\right]\\ &=\mathbb{E}\left[D_{U_2}\left(\nabla_{\nabla_{v_1}Y_t^{x;l^\epsilon}}\phi\big(Y_t^{x;l^\epsilon}\big)\right)\right]-\mathbb{E}\left[\nabla_{D_{U_2}\nabla_{v_1}Y_t^{x;l^\epsilon}}\phi\big(Y_t^{x;l^\epsilon}\big)\right]\\ &=\frac{1}{t-l_0^\epsilon}\mathbb{E}\left[\nabla_{\nabla_{v_1}Y_t^{x;l^\epsilon}}\phi\big(Y_t^{x;l^\epsilon}\big)\int_0^t\left\langle\nabla_{v_2}Y_s^{x;l^\epsilon},\mathrm{d}W_s\right\rangle\right]-\mathbb{E}\left[\nabla_{D_{U_2}\nabla_{v_1}Y_t^{x;l^\epsilon}}\phi\big(Y_t^{x;l^\epsilon}\big)\right].\end{aligned}$$

We can now use the fact that for each $t\ge0$,

$$Y_{l_t^\epsilon}^{x;l^\epsilon}=X_t^{x;l^\epsilon}\quad\text{and}\quad\nabla_v Y_{l_t^\epsilon}^{x;l^\epsilon}=\nabla_v X_t^{x;l^\epsilon}.$$

Replacing t with l_t^ϵ in (3.16), this yields

$$\begin{aligned}&\mathbb{E}\left[\nabla_{\nabla_{v_2}X_t^{x;l^\epsilon}}\nabla_{\nabla_{v_1}X_t^{x;l^\epsilon}}\phi\big(X_t^{x;l^\epsilon}\big)\right]\\ &=\frac{1}{l_t^\epsilon-l_0^\epsilon}\,\mathbb{E}\left[\nabla_{\nabla_{v_1}X_t^{x;l^\epsilon}}\phi\big(X_t^{x;l^\epsilon}\big)\int_0^{l_t^\epsilon}\left\langle\nabla_{v_2}Y_s^{x;l^\epsilon},\mathrm{d}W_s\right\rangle\right]-\mathbb{E}\left[\nabla_{D_{U_2}\nabla_{v_1}Y_{l_t^\epsilon}^{x;l^\epsilon}}\phi\big(X_t^{x;l^\epsilon}\big)\right]\\ &=\frac{1}{l_t^\epsilon-l_0^\epsilon}\,\mathbb{E}\left[\nabla_{\nabla_{v_1}X_t^{x;l^\epsilon}}\phi\big(X_t^{x;l^\epsilon}\big)\int_0^t\left\langle\nabla_{v_2}X_s^{x;l^\epsilon},\mathrm{d}W_{l_s^\epsilon}\right\rangle\right]-\mathbb{E}\left[\nabla_{D_{U_2}\nabla_{v_1}Y_{l_t^\epsilon}^{x;l^\epsilon}}\phi\left(X_t^{x;l^\epsilon}\right)\right].\end{aligned}$$

Since $\gamma^\epsilon_{l^\epsilon_t} = t$, (3.20) implies that for every $t \in (0,1]$

$$\begin{aligned}\left|\mathbb{E}\left[\nabla_{D_{U_2}\nabla_{v_1}Y^{x;l^\epsilon}_{l^\epsilon_t}}\phi\big(X^{x;l^\epsilon}_t\big)\right]\right| &= \left|\mathbb{E}\left[\left\langle\nabla\phi\big(X^{x;l^\epsilon}_t\big), D_{U_2}\nabla_{v_1}Y^{x;l^\epsilon}_{l^\epsilon_t}\right\rangle\right]\right| \\ &\le \|\nabla\phi\|_\infty \frac{\theta_3}{\sqrt{2\theta_2}}\,\mathrm{e}^{2\theta_2\gamma^\epsilon_{l^\epsilon_t}}\sqrt{\gamma^\epsilon_{l^\epsilon_t}}\,|v_1|\,|v_2| \\ &= \|\nabla\phi\|_\infty \frac{\theta_3}{\sqrt{2\theta_2}}\,\mathrm{e}^{2\theta_2 t}\sqrt{t}\,|v_1|\,|v_2| \\ &\le \|\nabla\phi\|_\infty \frac{\theta_3}{\sqrt{2\theta_2}}\,\mathrm{e}^{2\theta_2}\,|v_1|\,|v_2|\,,\end{aligned}$$

and so

(3.21)
$$\begin{aligned}&\left|\mathbb{E}\left[\nabla_{\nabla_{v_2}X^{x;l^\epsilon}_t}\nabla_{\nabla_{v_1}X^{x;l^\epsilon}_t}\phi\big(X^{x;l^\epsilon}_t\big)\right]\right| \\ &\le \left|\frac{1}{l^\epsilon_t - l^\epsilon_0}\,\mathbb{E}\left[\nabla_{\nabla_{v_1}X^{x;l^\epsilon}_t}\phi\big(X^{x;l^\epsilon}_t\big)\int_0^t\left\langle\nabla_{v_2}X^{x;l^\epsilon}_s, \mathrm{d}W_{l^\epsilon_s}\right\rangle\right]\right| + \|\nabla\phi\|_\infty\frac{\theta_3}{\sqrt{2\theta_2}}\,\mathrm{e}^{2\theta_2}\,|v_1|\,|v_2|\,.\end{aligned}$$

By the same argument as in the proof of [49, Lemma 2.5], we obtain

$$\begin{aligned}&\lim_{\epsilon\to 0}\frac{1}{l^\epsilon_t - l^\epsilon_0}\mathbb{E}\left[\nabla_{\nabla_{v_1}X^{x;l^\epsilon}_t}\phi\big(X^{x;l^\epsilon}_t\big)\int_0^t\left\langle\nabla_{v_2}X^{x;l^\epsilon}_s, \mathrm{d}W_{l^\epsilon_s}\right\rangle\right] \\ &\quad= \frac{1}{l_t}\,\mathbb{E}\left[\nabla_{\nabla_{v_1}X^{x;l}_t}\phi\left(X^{x;l}_t\right)\int_0^t\left\langle\nabla_{v_2}X^{x;l}_s, \mathrm{d}W_{l_s}\right\rangle\right].\end{aligned}\tag{3.22}$$

On the other hand, from [49, Lemma 2.2], we know that

$$\lim_{\epsilon\to 0}\mathbb{E}\left[\nabla_{\nabla_{v_2}X^{x;l^\epsilon}_t}\nabla_{\nabla_{v_1}X^{x;l^\epsilon}_t}\phi\big(X^{x;l^\epsilon}_t\big)\right] = \mathbb{E}\left[\nabla_{\nabla_{v_2}X^{x;l}_t}\nabla_{\nabla_{v_1}X^{x;l}_t}\phi\left(X^{x;l}_t\right)\right].\tag{3.23}$$

Letting in (3.21) $\epsilon \to 0$ and using (3.22) and (3.23), completes the proof. □

3.4. **Proof of Lemma 2.1.** Because of (3.3), we can use the differentiation theorem for parameter dependent integrals to get

$$\begin{aligned}|\nabla_v P_t h(x)| = |\nabla_v \mathbb{E}[h(X^x_t)]| &= \left|\mathbb{E}\left[\nabla_{\nabla_v X^x_t}h(X^x_t)\right]\right| \\ &= |\mathbb{E}\left[\langle\nabla h(X^x_t), \nabla_v X^x_t\rangle\right]| \le \|\nabla h\|\,\mathrm{e}^{\theta_2}\,|v| \le \mathrm{e}^{\theta_2}\,|v|\,.\end{aligned}$$

In order to see the second inequality, we define for every $\epsilon > 0$

$$h_\epsilon(x) := \int_{\mathbb{R}^d} g_\epsilon(y)h(x-y)\,\mathrm{d}y,\tag{3.24}$$

where g_ϵ is the density of the normal distribution $N(0,\epsilon^2 I_d)$. It is easy to see that h_ϵ is smooth, $\lim_{\epsilon\to 0} h_\epsilon(x) = h(x)$, $\lim_{\epsilon\to 0}\nabla h_\epsilon(x) = \nabla h(x)$ and $|h_\epsilon(x)| \le C(1+|x|)$ for all $x \in \mathbb{R}^d$ and some $C > 0$. Moreover, $\|\nabla h_\epsilon\| \le \|\nabla h\| \le 1$. Using the differentiability theorem for parameter dependent integrals we get

$$\nabla_{v_2}\nabla_{v_1}\mathbb{E}\left[h_\epsilon(X^x_t)\right] = \mathbb{E}\left[\nabla_{\nabla_{v_2}\nabla_{v_1}X^x_t}h_\epsilon(X^x_t)\right] + \mathbb{E}\left[\nabla_{\nabla_{v_2}X^x_t}\nabla_{\nabla_{v_1}X^x_t}h_\epsilon(X^x_t)\right].$$

From (3.4) we get

$$\left|\mathbb{E}\left[\nabla_{\nabla_{v_2}\nabla_{v_1}X^x_t}h_\epsilon(X^x_t)\right]\right| = |\mathbb{E}\left[\langle\nabla h_\epsilon(X^x_t), \nabla_{v_2}\nabla_{v_1}X^x_t\rangle\right]| \le \frac{\theta_3}{2\sqrt{2\theta_2}}\,\mathrm{e}^{4\theta_2}\,|v_1|\,|v_2|\,.$$

It follows from Lemma 3.2, [49, (3.3)] and (3.3) that

$$\begin{aligned}
&\left|\mathbb{E}\left[\nabla_{\nabla_{v_2}X_t^x}\nabla_{\nabla_{v_1}X_t^x}h_\epsilon(X_t^x)\right]\right| \\
&\le \left|\mathbb{E}\left[\frac{1}{S_t}\nabla_{\nabla_{v_1}X_t^x}h_\epsilon\left(X_t^x\right)\int_0^t\langle\nabla_{v_2}X_s^x,\mathrm{d}W_{S_s}\rangle\right]\right| + \|\nabla h_\epsilon\|_\infty\frac{\theta_3}{\sqrt{2\theta_2}}\,\mathrm{e}^{2\theta_2}\,|v_1|\,|v_2| \\
&\le \mathrm{e}^{\theta_2}\,|v_1|\,\mathbb{E}\left[\frac{1}{S_t}\left|\int_0^t\langle\nabla_{v_2}X_s^x,\mathrm{d}W_{S_s}\rangle\right|\right] + \frac{\theta_3}{\sqrt{2\theta_2}}\,\mathrm{e}^{2\theta_2}\,|v_1|\,|v_2|.
\end{aligned}$$

The Cauchy–Schwarz inequality, Itô's isometry and (3.9) give

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{S_t}\left|\int_0^t\langle\nabla_{v_2}X_s^x,\mathrm{d}W_{S_s}\rangle\right|\right] &= \int_{\mathbb{S}}\frac{1}{l_t}\,\mathbb{E}\left|\int_0^t\left\langle\nabla_{v_2}X_s^{x;l},\mathrm{d}W_{l_s}\right\rangle\right|\mu_{\mathbb{S}}(\mathrm{d}l) \\
&\le \int_{\mathbb{S}}\frac{1}{l_t}\left(\mathbb{E}\int_0^t|\nabla_{v_2}X_s^{x;l}|^2\,\mathrm{d}l_s\right)^{1/2}\mu_{\mathbb{S}}(\mathrm{d}l) \\
&\le \mathrm{e}^{\theta_2}\,|v_2|\int_{\mathbb{S}}\frac{1}{\sqrt{l_t}}\,\mu_{\mathbb{S}}(\mathrm{d}l) \\
&= \mathrm{e}^{\theta_2}\,|v_2|\,\mathbb{E}\left[S_t^{-1/2}\right] \\
&\le C\mathrm{e}^{\theta_2}\,|v_2|\,t^{-1/\alpha},
\end{aligned}$$

where the last inequality is taken from [10, Theorem 2.1 (ii) (c)]. Thus, we have for all $t \in (0,1]$,

$$\begin{aligned}
|\nabla_{v_2}\nabla_{v_1}\mathbb{E}\left[h_\epsilon(X_t^x)\right]| &\le \left|\mathbb{E}\left[\nabla_{\nabla_{v_2}\nabla_{v_1}X_t^x}h_\epsilon(X_t^x)\right]\right| + \left|\mathbb{E}\left[\nabla_{\nabla_{v_2}X_t^x}\nabla_{\nabla_{v_1}X_t^x}h_\epsilon(X_t^x)\right]\right| \\
&\le \frac{\theta_3}{2\sqrt{2\theta_2}}\,\mathrm{e}^{4\theta_2}\,|v_1|\,|v_2| + \mathrm{e}^{\theta_2}\,|v_1|\,C\mathrm{e}^{\theta_2}\,|v_2|\,t^{-1/\alpha} + \frac{\theta_3}{\sqrt{2\theta_2}}\,\mathrm{e}^{2\theta_2}\,|v_1|\,|v_2| \\
&\le \mathrm{e}^{4\theta_2}\left(\frac{\theta_3}{2\sqrt{2\theta_2}} + Ct^{-1/\alpha} + \frac{\theta_3}{\sqrt{2\theta_2}}\right)|v_1|\,|v_2| \\
&\le C\mathrm{e}^{4\theta_2}t^{-1/\alpha}\,|v_1|\,|v_2|.
\end{aligned}$$

Finally, we can let $\epsilon \to 0$ using dominated convergence,

$$\lim_{\epsilon\to 0}\nabla_{v_2}\nabla_{v_1}\mathbb{E}\left[h_\epsilon(X_t^x)\right] = \nabla_{v_2}\nabla_{v_1}\mathbb{E}\left[h(X_t^x)\right],$$

completing the proof of Lemma 2.1. □

APPENDIX A. PROOFS OF PROPOSITIONS 1.5, 1.6 AND 1.7, AND LEMMA 1.8

Proof of Proposition 1.5. The generator $\mathcal{L}^\alpha$ of the process X_t is given

$$\mathcal{L}^\alpha f(x) = \langle b(x), \nabla f(x)\rangle + (-\Delta)^{\alpha/2}f(x), \quad f \in \mathcal{C}^2(\mathbb{R}^d,\mathbb{R})$$

where $(-\Delta)^{\alpha/2}$ is the fractional Laplace operator, which is the generator of the rotationally symmetric α-stable Lévy process Z_t; it is defined as a principal value integral

$$(-\Delta)^{\alpha/2}f(x) = C_{d,\alpha}\cdot\text{p.v.}\int_{\mathbb{R}^d}\left(f(x+y)-f(x)\right)\frac{\mathrm{d}y}{|y|^{\alpha+d}}. \tag{A.1}$$

It is not difficult to see that for all functions from the set

$$\mathcal{D} := \left\{f \in \mathcal{C}^2(\mathbb{R}^d,\mathbb{R})\,;\ \int_{|z|\ge 1}|f(x+z)-f(x)|\,\frac{\mathrm{d}z}{|z|^{\alpha+d}} < \infty\right\}.$$

$\mathcal{L}^\alpha f$ is well-defined; moreover $\mathcal{D}\times\mathcal{L}^\alpha(\mathcal{D})$ can be embedded into the *full generator* $\widehat{\mathcal{L}^\alpha}$, i.e. the set of all pairs of (bounded) Borel functions (f,g) such that $f(X_t)-f(X_0)-\int_0^t g(X_s)\,\mathrm{d}s$ is a (local) martingale, see the discussion in [4, pp. 24–26].

Recall that $V_\beta(x)=(1+|x|^2)^{\beta/2}$. It is easy to check that $V_\beta\in\mathcal{D}(\mathcal{L}^\alpha)$. Since

$$\nabla V_\beta(x)=\frac{\beta x}{(1+|x|^2)^{\frac{2-\beta}{2}}},\quad \nabla^2 V_\beta(x)=\frac{\beta I_d}{(1+|x|^2)^{1-\frac{\beta}{2}}}+\frac{\beta(\beta-2)xx^\top}{(1+|x|^2)^{2-\frac{\beta}{2}}},$$

(I_d denotes the $d\times d$ identity matrix) we see that for any $x\in\mathbb{R}^d$

$$|\nabla V_\beta(x)|\le\beta|x|^{\beta-1},\quad \|\nabla^2 V_\beta(x)\|_{\mathrm{HS}}\le\beta(3-\beta)\sqrt{d}.$$

Thus, (1.7) and Young's inequality ($AB\le\frac1p A^p+\frac1q B^q$ with $p=\beta$ and $q=\beta/(\beta-1)$) imply

$$\begin{aligned}\langle b(x),\nabla V_\beta(x)\rangle&=\frac{\beta}{(1+|x|^2)^{\frac{2-\beta}{2}}}\langle b(x)-b(0),x\rangle+\frac{\beta}{(1+|x|^2)^{\frac{2-\beta}{2}}}\langle b(0),x\rangle\\&\le-\theta_1\frac{\beta|x|^2}{(1+|x|^2)^{\frac{2-\beta}{2}}}+\frac{\beta K}{(1+|x|^2)^{\frac{2-\beta}{2}}}+\frac{\beta|b(0)||x|}{(1+|x|^2)^{\frac{2-\beta}{2}}}\\&\le-\theta_1\beta V_\beta(x)+\theta_1\beta+\beta K+\beta|b(0)||x|^{\beta-1}\\&\le-\theta_1 V_\beta(x)+\theta_1\beta+\beta K+\theta_1^{1-\beta}|b(0)|^\beta.\end{aligned}$$

Therefore, we see from (A.1) that

(A.2)

$$\begin{aligned}(-\Delta)^{\alpha/2}V_\beta(x)&=C_{d,\alpha}\int\left(V_\beta(x+y)-V_\beta(x)-\langle\nabla V_\beta(x),y\rangle\,\mathbb{1}_{(0,1)}(|y|)\right)\frac{\mathrm{d}y}{|y|^{\alpha+d}}\\&=C_{d,\alpha}\int_{|y|<1}\int_0^1\int_0^r\left\langle\nabla^2 V_\beta(x+sy),yy^\top\right\rangle_{\mathrm{HS}}\,\mathrm{d}s\,\mathrm{d}r\,\frac{\mathrm{d}y}{|y|^{\alpha+d}}\\&\quad+C_{d,\alpha}\int_{|y|\ge1}\int_0^1\langle\nabla V_\beta(x+ry),y\rangle\,\mathrm{d}r\,\frac{\mathrm{d}y}{|y|^{\alpha+d}}\\&\le C_{d,\alpha}\frac{\beta(3-\beta)}{2}\sqrt{d}\int_{|y|<1}\frac{|y|^2}{|y|^{\alpha+d}}\,\mathrm{d}y+C_{d,\alpha}\beta\int_{|y|\ge1}\frac{|x|^{\beta-1}|y|+|y|^\beta}{|y|^{\alpha+d}}\,\mathrm{d}y\\&=\frac{C_{d,\alpha}\beta(3-\beta)\sqrt{d}\sigma_{d-1}}{2(2-\alpha)}+C_{d,\alpha}\beta\sigma_{d-1}\left(\frac{|x|^{\beta-1}}{\alpha-1}+\frac{1}{\alpha-\beta}\right).\end{aligned}$$

Again by Young's inequality we get

$$\left|(-\Delta)^{\alpha/2}V_\beta(x)\right|\le\frac{C_{d,\alpha}\beta(3-\beta)\sqrt{d}\sigma_{d-1}}{2(2-\alpha)}+\frac{C_{d,\alpha}\beta\sigma_{d-1}}{\alpha-\beta}+\left(\frac{\theta_1}{4}\right)^{1-\beta}\left(\frac{C_{d,\alpha}\sigma_{d-1}}{\alpha-1}\right)^\beta+\frac{\theta_1}{4}V_\beta(x).$$

Hence, we have

(A.3)
$$\mathcal{L}^\alpha V_\beta(x)\le-\lambda_1 V_\beta(x)+q_1\mathbb{1}_{A_1}(x),$$

with $\lambda_1=\frac12\theta_1$,

$$q_1=\theta_1\beta+\beta K+\theta_1^{1-\beta}|b(0)|^\beta+\frac{C_{d,\alpha}\beta(3-\beta)\sqrt{d}\sigma_{d-1}}{2(2-\alpha)}+\frac{C_{d,\alpha}\beta\sigma_{d-1}}{\alpha-\beta}+\left(\frac{\theta_1}{4}\right)^{1-\beta}\left(\frac{C_{d,\alpha}\sigma_{d-1}}{\alpha-1}\right)^\beta,$$

and the compact set $A_1=\left\{x\in\mathbb{R}^d\,:\,|x|\le\left(4\theta_1^{-1}q_1\right)^{1/\beta}\right\}$.

Thus, [27, Theorem 5.1] yields that the process $(X_t^x)_{t\ge 0}$ is ergodic, i.e. there exists a unique invariant probability measure μ such that for all $x\in\mathbb{R}^d$ and $t>0$,

$$\lim_{t\to\infty}\|P_t(x,\cdot)-\mu\|_{\mathrm{TV}}=0,$$

where $P_t(x,dz)$ is the transition function of the process $(X_t^x)_{t\ge 0}$ and $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation norm on the space of signed measures. Furthermore, because of the inequality above and [27, Theorem 6.1], we have

$$\sup_{|f|\le V_\beta}\left|\mathbb{E}[f(X_t^x)]-\mu(f)\right|\le c_1 V_\beta(x)\mathrm{e}^{-c_2 t}$$

for suitable constants $c_1,c_2>0$. In addition, by Itô's formula, the integrability of X_t^x can be derived directly from the Lyapunov condition (A.3). □

Proof of Proposition 1.6. For any $x,y\in\mathbb{R}^d$, (1.8) implies that

$$\langle b(x)-b(y),x-y\rangle\le\theta_2|x-y|^2,$$

and (1.7) shows for all $|x-y|^2>2K/\theta_1$ that

$$\langle b(x)-b(y),x-y\rangle\le-\theta_1|x-y|^2+K\le-\frac{\theta_1}{2}|x-y|^2.$$

Hence, we can use [47, Theorem 1.2] with $K_1=\theta_2$, $K_2=\frac{\theta_1}{2}$ and $L_0=\frac{2K}{\theta_1}$ to get the desired estimate. □

Proof of Proposition 1.7. We show only (1.12), as (1.13) can be proved in the same way.

Denote by $P(x,\mathrm{d}y)=\mathbb{P}(Y_1\in\mathrm{d}y\mid Y_0=x)$. Since $V_1(y)\le|y|+1$ and

$$Y_1=x+\eta b(x)+Z_\eta,$$

we have

$$\begin{aligned}\int_{\mathbb{R}^d}V_1(y)\,P(x,\mathrm{d}y)&\le\int_{\mathbb{R}^d}(|y|+1)p\left(\eta,y-x-\eta b(x)\right)\mathrm{d}y\\&=\int_{\mathbb{R}^d}\left(|z+x+\eta b(x)|+1\right)p(\eta,z)\,\mathrm{d}z\\&\le\int_{\mathbb{R}^d}\left(|z|+|x+\eta b(x)|+1\right)p(\eta,z)\,\mathrm{d}z\\&\le\mathbb{E}|Z_\eta|+\left|x+\eta\big(b(x)-b(0)\big)\right|+\eta|b(0)|+1.\end{aligned}$$

By (1.7) and (1.9), we further have

$$\begin{aligned}\left|x+\eta\big(b(x)-b(0)\big)\right|^2&=|x|^2+2\eta\left\langle b(x)-b(0),x\right\rangle+\eta^2|b(x)-b(0)|^2\\&\le\left(1-2\theta_1\eta+\theta_2^2\eta^2\right)|x|^2+2K\eta,\end{aligned}$$

which implies

$$\begin{aligned}\int_{\mathbb{R}^d}V_1(y)\,P(x,\mathrm{d}y)&\le\eta^{\frac{1}{\alpha}}\mathbb{E}|Z_1|+(1-2\theta_1\eta+\theta_2^2\eta^2)^{1/2}|x|+\sqrt{2K\eta}+\eta|b(0)|+1\\&\le(1-\theta_1\eta)\,|x|+\eta^{\frac{1}{\alpha}}\mathbb{E}|Z_1|+\sqrt{2K\eta}+\eta|b(0)|+1,\end{aligned}$$

where the last two inequalities hold because of $\eta<\min\left\{1,\,\theta_1\theta_2^{-2},\,\theta_1^{-1}\right\}$. Hence, we have

$$\int_{\mathbb{R}^d}V_1(y)\,P(x,\mathrm{d}y)\le\lambda_2V_1(x)+q_2\mathbb{1}_{A_2}(x)$$

with

$$\lambda_2 = 1 - \frac{\theta_1}{2}\eta < 1, \quad q_2 = 1 + \frac{\theta_1}{2}\eta + \eta^{\frac{1}{\alpha}}\mathbb{E}|Z_1| + \sqrt{2K\eta} + \eta|b(0)|,$$

and the compact set $A_2 = \left\{ x \in \mathbb{R}^d \,:\, |x| \le 1 + \frac{2}{\theta_1}\mathbb{E}|Z_1|\eta^{\frac{1}{\alpha}} + \frac{2\sqrt{2K}}{\theta_1}\eta^{-\frac{1}{2}} + \frac{2|b(0)|}{\theta_1} \right\}$.

The proof of irreducibility is standard, see e.g. [24, Appendix A].

We can now use this and [26, Theorem 6.3] to see that the process $(Y_k^x)_{k\ge 0}$ is exponentially ergodic, i.e. there exists a unique invariant probability μ_η such that for all $x \in \mathbb{R}^d$ and $t > 0$,

$$\sup_{|f|\le V_1} |\mathbb{E}f(Y_k) - \mu_\eta(f)| \le c_1 V_1(x) \mathrm{e}^{-c_2 k},$$

for suitable constants $c_1, c_2 > 0$. □

Proof of Lemma 1.8. We show only (1.14) as the inequality (1.15) can be proved in the same way.

Notice that $|y|^\beta \le V_\beta(y)$ and

$$\begin{aligned} V_\beta(Y_{k+1}) &= V_\beta\left(Y_k + \eta b(Y_k) + Z_\eta\right) \\ &= V_\beta\left(Y_k + \eta b(Y_k)\right) + V_\beta\left(Y_k + \eta b(Y_k) + Z_\eta\right) - V_\beta\left(Y_k + \eta b(Y_k)\right) \\ &= V_\beta(Y_k) + \int_0^\eta \left\langle \nabla V_\beta\left(Y_k + r b(Y_k)\right), b(Y_k)\right\rangle \mathrm{d}r \\ &\quad + V_\beta\left(Y_k + \eta b(Y_k) + Z_\eta\right) - V_\beta\left(Y_k + \eta b(Y_k)\right), \end{aligned}$$

where Z_η is independent of Y_k. Since $\nabla V_\beta(x) = \beta x(1+|x|^2)^{-\frac{2-\beta}{2}}$, (1.7) implies that

$$\begin{aligned} &\int_0^\eta \left\langle \nabla V_\beta\left(Y_k + r b(Y_k)\right), b(Y_k)\right\rangle \mathrm{d}r \\ &\quad\le \int_0^\eta \frac{\beta\left\langle Y_k, b(Y_k)\right\rangle + \beta r|b(Y_k)|^2}{(1+|Y_k + rb(Y_k)|^2)^{\frac{2-\beta}{2}}}\mathrm{d}r \\ &\quad\le \int_0^\eta \frac{\beta\left\langle Y_k, b(Y_k) - b(0)\right\rangle + \beta|b(0)||Y_k| + \beta r|b(Y_k)|^2}{(1+|Y_k + rb(Y_k)|^2)^{\frac{2-\beta}{2}}}\mathrm{d}r \\ &\quad\le \int_0^\eta \frac{-\theta_1\beta|Y_k|^2 + \beta K + \beta|b(0)||Y_k| + \beta r|b(Y_k)|^2}{(1+|Y_k + rb(Y_k)|^2)^{\frac{2-\beta}{2}}}\mathrm{d}r. \end{aligned} \tag{A.4}$$

One can write by (1.9) and the fact $r \le \eta \le \min\left\{1, \frac{\theta_1}{8\theta_2^2}, \frac{1}{\theta_1}\right\}$

$$\begin{aligned} &-\theta_1\beta|Y_k|^2 + \beta K + \beta|b(0)||Y_k| + \beta r|b(Y_k)|^2 \\ &\quad\le -\frac{3}{4}\theta_1\beta|Y_k|^2 + 2\eta\theta_2^2\beta|Y_k|^2 + \frac{\beta|b(0)|^2}{\theta_1} + 2\beta r|b(0)|^2 + \beta K \\ &\quad\le -\frac{\theta_1\beta}{2}|Y_k|^2 + \frac{\beta|b(0)|^2}{\theta_1} + 2\beta r|b(0)|^2 + \beta K, \end{aligned}$$

whereas

$$
\begin{aligned}
&|Y_k + rb(Y_k)|^2 + 1 \\
&\quad = |Y_k|^2 + 2r\langle Y_k, b(Y_k)\rangle + r^2 b(Y_k)^2 + 1 \\
&\quad \le |Y_k|^2 + 2r\langle Y_k, b(Y_k) - b(0)\rangle + 2r|b(0)||Y_k| + 2r^2\theta_2^2|Y_k|^2 + 2r^2|b(0)|^2 + 1 \\
&\quad \le (1 - 2r\theta_1)|Y_k|^2 + 2r|b(0)||Y_k| + 2r^2\theta_2^2|Y_k|^2 + 2\eta^2|b(0)|^2 + 1 + 2\eta K \\
&\quad \le |Y_k|^2 + \frac{2r|b(0)|^2}{\theta_1} + 2\eta^2|b(0)|^2 + 1 + 2\eta K.
\end{aligned}
$$

Noting that $(1 + |Y_k + rb(Y_k)|^2)^{\frac{2-\beta}{2}} \ge 1$, we get

$$
\begin{aligned}
&\frac{-\theta_1\beta|Y_k|^2 + \beta K + \beta|b(0)||Y_k| + \beta r|b(Y_k)|^2}{(1 + |Y_k + rb(Y_k)|^2)^{\frac{2-\beta}{2}}} \\
&\le -\frac{\theta_1\beta}{2}\frac{|Y_k|^2}{\left(|Y_k|^2 + \frac{2r|b(0)|^2}{\theta_1} + 2\eta^2|b(0)|^2 + 1 + 2\eta K\right)^{\frac{2-\beta}{2}}} + \frac{\beta|b(0)|^2}{\theta_1} + 2\beta r|b(0)|^2 + \beta K \\
&\le -\frac{\theta_1\beta}{2}\left(|Y_k|^2 + \frac{2r|b(0)|^2}{\theta_1} + 2\eta^2|b(0)|^2 + 1 + 2\eta K\right)^{\frac{\beta}{2}} + C_1 \\
&\le -\frac{\theta_1\beta}{2}V_\beta(Y_k) + C_1,
\end{aligned}
$$

where

$$
C_1 := \frac{\theta_1\beta}{2}\left(\frac{2r|b(0)|^2}{\theta_1} + 2\eta^2|b(0)|^2 + 1 + 2\eta K\right) + \frac{\beta|b(0)|^2}{\theta_1} + 2\beta r|b(0)|^2 + \beta K.
$$

Combining this with (A.4), we arrive at

$$
\int_0^\eta \langle \nabla V_\beta(Y_k + rb(Y_k)), b(Y_k)\rangle \,\mathrm{d}r \le -\frac{\theta_1\beta}{2}\eta V_\beta(Y_k) + C_1\eta.
$$

In addition, for any $y \in \mathbb{R}^d$, Itô's formula and the inequality (A.2) imply that

$$
\begin{aligned}
&|\mathbb{E}[V_\beta(y + \eta b(y) + Z_\eta) - V_\beta(y + \eta b(y))]| \\
&= \left|\int_0^\eta \mathbb{E}\left[(-\Delta)^{\alpha/2}V_\beta(y + \eta b(y) + Z_r)\right]\mathrm{d}r\right| \\
&\le \int_0^\eta \left[\frac{C_{d,\alpha}\beta(3-\beta)\sqrt{d}\sigma_{d-1}}{2(2-\alpha)} + C_{d,\alpha}\beta\sigma_{d-1}\left(\frac{\mathbb{E}|y + \eta b(y) + Z_r|^{\beta-1}}{\alpha - 1} + \frac{1}{\alpha-\beta}\right)\right]\mathrm{d}r \\
&\le C_{d,\alpha}\beta\sigma_{d-1}\int_0^\eta \left[\frac{(3-\beta)\sqrt{d}}{2(2-\alpha)} + \frac{|y|^{\beta-1} + \eta^{\beta-1}|b(y)|^{\beta-1} + \mathbb{E}|Z_r|^{\beta-1}}{\alpha-1} + \frac{1}{\alpha-\beta}\right]\mathrm{d}r.
\end{aligned}
$$

This yields, in turn,

$$
\begin{aligned}
&|\mathbb{E}[V_\beta(y + \eta b(y) + Z_\eta) - V_\beta(y + \eta b(y))]| \\
&\le C_{d,\alpha}\beta\sigma_{d-1}\left(\frac{(3-\beta)\sqrt{d}\eta}{2(2-\alpha)} + \frac{\eta}{\alpha-\beta} + \frac{1+\theta_2^{\beta-1}}{\alpha-1}\eta|y|^{\beta-1} + \eta|b(0)|^{\beta-1} + \int_0^\eta \frac{\mathbb{E}|Z_r|^{\beta-1}}{\alpha-1}\mathrm{d}r\right) \\
&\le C_{d,\alpha}\beta\sigma_{d-1}\left(\frac{(3-\beta)\sqrt{d}\eta}{2(2-\alpha)} + \frac{\eta}{\alpha-\beta} + \frac{1+\theta_2^{\beta-1}}{\alpha-1}\eta|y|^{\beta-1} + \eta|b(0)|^{\beta-1} + \frac{\mathbb{E}|Z_1|^{\beta-1}}{\alpha-1}\eta\right),
\end{aligned}
$$

where the first inequality uses (1.9) and the fact that $0 < \eta < 1$. Since Z_η is independent of Y_k, we can derive that

$$
\begin{aligned}
&|V_\beta(Y_k + \eta b(Y_k) + Z_\eta) - V_\beta(Y_k + \eta b(Y_k))| \\
&\le C_{d,\alpha}\beta\sigma_{d-1}\left(\frac{(3-\beta)\sqrt{d}\eta}{2(2-\alpha)} + \frac{\eta}{\alpha-\beta} + \frac{1+\theta_2^{\beta-1}}{\alpha-1}\eta\mathbb{E}|Y_k|^{\beta-1} + \eta|b(0)|^{\beta-1} + \frac{\mathbb{E}|Z_1|^{\beta-1}}{\alpha-1}\eta\right) \\
&\le \frac{\theta_1(\beta-1)}{2}\eta\, V_\beta(Y_k) + C_2\eta,
\end{aligned}
$$

where the last inequality follows from Young's inequality and

$$
\begin{aligned}
C_2 = C_{d,\alpha}\beta\sigma_{d-1}&\left(\frac{(3-\beta)\sqrt{d}}{2(2-\alpha)} + \frac{1}{\alpha-\beta} + |b(0)|^{\beta-1} + \frac{\mathbb{E}|Z_1|^{\beta-1}}{\alpha-1}\right) \\
&+ \left(\frac{C_{d,\alpha}\sigma_{d-1}(1+\theta_2^{\beta-1})}{\alpha-1}\right)^\beta \left(\frac{2}{\theta_1}\right)^{\beta-1}.
\end{aligned}
$$

Therefore,

$$
\mathbb{E}[V_\beta(Y_{k+1})] \le \left(1 - \frac{\theta_1}{2}\eta\right)\mathbb{E}[V_\beta(Y_k)] + (C_1 + C_2)\eta,
$$

which we can iterate this to get

$$
\begin{aligned}
\mathbb{E}[V_\beta(Y_{k+1})] &\le \left(1 - \frac{\theta_1}{2}\eta\right)^{k+1} V_\beta(x) + (C_1 + C_2)\eta\sum_{j=0}^{k}\left(1 - \frac{\theta_1}{2}\eta\right)^j \\
&\le V_\beta(x) + \frac{2(C_1 + C_2)}{\theta_1}.
\end{aligned}
$$

Using that $V_\beta(x) \le 1 + |x|^\beta$, we finally get

$$
\mathbb{E}|Y_k^x|^\beta \le \mathbb{E}[V_\beta(Y_k^x)] \le C(1 + |x|^\beta),
$$

for some constant C which is independent of η. □

APPENDIX B. EXACT RATE FOR THE ORNSTEIN–UHLENBECK PROCESS

In this section, we assume that μ is the invariant measure of the Ornstein–Uhlenbeck process on $\mathbb{R}$:

$$
\mathrm{d}X_t = -X_t\,\mathrm{d}t + \mathrm{d}Z_t, \quad X_0 = x, \tag{B.1}
$$

where Z_t is a rotationally symmetric α-stable Lévy process ($1 < \alpha < 2$), and $\tilde{\mu}_\eta$ is the invariant measure of

$$
\tilde{Y}_{k+1} = \tilde{Y}_k - \eta\tilde{Y}_k + \frac{\eta^{1/\alpha}}{\sigma}\widetilde{Z}_{k+1}, \quad k = 0, 1, 2, \ldots,
$$

where $\eta \in (0,1)$, $\tilde{Y}_0 = x$, $\sigma = \left(\frac{\alpha}{2d_\alpha}\right)^{1/\alpha}$ with $d_\alpha = C_{1,\alpha} = \left(2\int_0^\infty \frac{1-\cos y}{y^{\alpha+1}}\,\mathrm{d}y\right)^{-1}$, and $\widetilde{Z}_j$ are i.i.d. random variables with density

$$
p(z) = \frac{\alpha}{2|z|^{\alpha+1}}\mathbb{1}_{(1,\infty)}(|z|). \tag{B.2}
$$

Proposition B.1. *For every $x \in \mathbb{R}$ and $\alpha \in (1,2)$,*

$$
0 < \liminf_{\eta\downarrow 0}\frac{W_1(\mu, \tilde{\mu}_\eta)}{\eta^{2/\alpha-1}} \le \limsup_{\eta\downarrow 0}\frac{W_1(\mu, \tilde{\mu}_\eta)}{\eta^{2/\alpha-1}} < \infty.
$$

Proof. Since $X_t = \mathrm{e}^{-t}x + \mathrm{e}^{-t}\int_0^t \mathrm{e}^s \, \mathrm{d}Z_s$, we get

$$\begin{aligned}\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\xi X_t}\right] = \mathrm{e}^{\mathrm{i}\xi \mathrm{e}^{-t}x}\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\int_0^t \xi \mathrm{e}^{-t}\mathrm{e}^s \, \mathrm{d}Z_s}\right] &= \mathrm{e}^{\mathrm{i}\xi \mathrm{e}^{-t}x}\mathbb{E}\left[\mathrm{e}^{\int_0^t |\xi \mathrm{e}^{-t}\mathrm{e}^s|^\alpha \, \mathrm{d}s}\right] \\ &= \mathrm{e}^{\mathrm{i}\xi \mathrm{e}^{-t}x}\mathrm{e}^{-\alpha^{-1}|\xi|^\alpha(1-\mathrm{e}^{-\alpha t})} \\ &\xrightarrow[t\to\infty]{} \mathrm{e}^{-\alpha^{-1}|\xi|^\alpha} = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\xi\alpha^{-1/\alpha}Z_1}\right].\end{aligned}$$

Hence, the invariant measure μ is given by the law of $\alpha^{-1/\alpha}Z_1$.

It is easy to see that

$$\tilde{Y}_{k+1} = (1-\eta)^{k+1}x + \frac{\eta^{1/\alpha}}{\sigma}\sum_{i=0}^{k}(1-\eta)^i \widetilde{Z}_{k+1-i}.$$

Denote by $\varphi(\xi) = \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\xi\widetilde{Z}_j}\right]$ the characteristic function of the Pareto distribution. Then we have

$$\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\xi\tilde{Y}_{k+1}}\right] = \mathrm{e}^{\mathrm{i}\xi(1-\eta)^{k+1}x}\prod_{i=0}^{k}\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\xi\frac{\eta^{1/\alpha}}{\sigma}(1-\eta)^i\widetilde{Z}_{k+1-i}}\right] = \mathrm{e}^{\mathrm{i}\xi(1-\eta)^{k+1}x}\prod_{i=0}^{k}\varphi\left(\frac{\eta^{1/\alpha}}{\sigma}(1-\eta)^i\xi\right).$$

Letting $k \to \infty$ and denoting by $\tilde{Y}_\eta$ a random variable with distribution $\tilde{\mu}_\eta$, we get

$$\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\xi\tilde{Y}_\eta}\right] = \prod_{i=0}^{\infty}\varphi\left(\frac{\eta^{1/\alpha}}{\sigma}(1-\eta)^i\xi\right). \tag{B.3}$$

For $\xi > 0$, we have

$$\begin{aligned}1-\varphi(\xi) = 2\int_1^\infty [1-\cos(\xi z)]\, p(z)\, \mathrm{d}z &= \alpha\int_1^\infty [1-\cos(\xi z)]\frac{\mathrm{d}z}{z^{\alpha+1}} \\ &= \alpha\xi^\alpha\left(\int_0^\infty [1-\cos u]\frac{\mathrm{d}u}{u^{\alpha+1}} - \int_0^\xi [1-\cos u]\frac{\mathrm{d}u}{u^{\alpha+1}}\right) \\ &= \sigma^\alpha\xi^\alpha - \alpha\xi^\alpha\int_0^\xi [1-\cos u]\frac{\mathrm{d}u}{u^{\alpha+1}}.\end{aligned}$$

Since $p(z)$ is symmetric, cf. (B.2), we have $\varphi(\xi) = \varphi(-\xi)$, and so

$$\varphi(\xi) = 1 - \sigma^\alpha|\xi|^\alpha + \alpha|\xi|^\alpha\int_0^{|\xi|}[1-\cos u]\frac{\mathrm{d}u}{u^{\alpha+1}}.$$

for all $\xi \in \mathbb{R}$. Since $c := \inf_{0<u\le 1}(1-\cos u)/u^2 > 0$, we get for all $|\xi| \le 1$,

$$\varphi(\xi) \ge 1 - \sigma^\alpha|\xi|^\alpha + \alpha|\xi|^\alpha\int_0^{|\xi|} cu^2\frac{\mathrm{d}u}{u^{\alpha+1}} = 1 - \sigma^\alpha|\xi|^\alpha + \frac{c\alpha}{2-\alpha}|\xi|^2.$$

Thus, for $|\xi| \le 1$ and $0 < \eta < 1 \wedge \sigma^\alpha$,

$$\begin{aligned}\log\varphi\left(\frac{\eta^{1/\alpha}}{\sigma}(1-\eta)^i\xi\right) &\ge \log\left(1 - \sigma^\alpha\left|\frac{\eta^{1/\alpha}}{\sigma}(1-\eta)^i\xi\right|^\alpha + \frac{c\alpha}{2-\alpha}\left|\frac{\eta^{1/\alpha}}{\sigma}(1-\eta)^i\xi\right|^2\right) \\ &= \log\left(1 - \eta(1-\eta)^{\alpha i}|\xi|^\alpha + \frac{c\alpha}{(2-\alpha)\sigma^2}\eta^{2/\alpha}(1-\eta)^{2i}|\xi|^2\right).\end{aligned} \tag{B.4}$$

Observe that

$$\lim_{x\downarrow 0}\frac{\log\left(1-x+\frac{c\alpha}{(2-\alpha)\sigma^2}x^{2/\alpha}\right)+x}{x^{2/\alpha}} = \frac{c\alpha}{(2-\alpha)\sigma^2}.$$

Therefore, there is some constant $C = C(\alpha,\sigma) > 0$ such that for small enough $x > 0$

$$\log\left(1 - x + \frac{c_1\alpha}{(2-\alpha)\sigma^2}\,x^{2/\alpha}\right) \geq -x + Cx^{2/\alpha}.$$

If we use this in (B.4), we obtain for all $|\xi| \leq 1$ and small enough $\eta > 0$,

$$\log\varphi\left(\frac{\eta^{1/\alpha}}{\sigma}(1-\eta)^i\xi\right) \geq -\eta(1-\eta)^{\alpha i}|\xi|^\alpha + C\eta^{2/\alpha}(1-\eta)^{2i}|\xi|^2.$$

Inserting this into (B.3), we see for all $|\xi| \leq 1$ and small enough $\eta > 0$

$$\begin{aligned}\log\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\xi\tilde{Y}_\eta}\right] &= \sum_{i=0}^{\infty}\log\varphi\left(\frac{\eta^{1/\alpha}}{\sigma}(1-\eta)^i\xi\right)\\ &\geq -\eta|\xi|^\alpha\sum_{i=0}^{\infty}(1-\eta)^{\alpha i} + C\eta^{2/\alpha}|\xi|^2\sum_{i=0}^{\infty}(1-\eta)^{2i}\\ &= -|\xi|^\alpha\frac{\eta}{1-(1-\eta)^\alpha} + C|\xi|^2\frac{\eta^{2/\alpha}}{1-(1-\eta)^2}\\ &= -\frac{1}{\alpha}\,|\xi|^\alpha - |\xi|^\alpha\Omega(\eta) + |\xi|^2\Omega(\eta^{2/\alpha-1}).\end{aligned}$$

In the last equality we use that

$$\lim_{\eta\downarrow 0}\left[\frac{\eta}{1-(1-\eta)^\alpha} - \frac{1}{\alpha}\right]\eta^{-1} = \frac{\alpha-1}{2\alpha} \quad\text{and}\quad \lim_{\eta\downarrow 0}\frac{\eta^{2/\alpha}}{1-(1-\eta)^2}\,\eta^{-2/\alpha+1} = \frac{1}{2}.$$

Here and in the following, the notation $f(\eta) = \Omega(g(\eta))$ as $\eta \downarrow 0$ means that $\lim_{\eta\downarrow 0}\frac{f(\eta)}{g(\eta)}$ is a positive (finite) constant, where f and g are some positive functions. With the elementary inequality $\mathrm{e}^x \geq 1 + x$ for $x \in \mathbb{R}$ we see for all $|\xi| \leq 1$ and sufficiently small $\eta > 0$ that

$$\begin{aligned}\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\xi\tilde{Y}_\eta}\right] &\geq \exp\left[-\frac{1}{\alpha}\,|\xi|^\alpha - |\xi|^\alpha\Omega(\eta) + |\xi|^2\Omega(\eta^{2/\alpha-1})\right]\\ &\geq \mathrm{e}^{-|\xi|^\alpha/\alpha}\left[1 - |\xi|^\alpha\Omega(\eta) + |\xi|^2\Omega(\eta^{2/\alpha-1})\right],\end{aligned}$$

which yields

$$\begin{aligned}\int_{-1}^{1}\left(\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\xi\tilde{Y}_\eta}\right] - \mathbb{E}\left[\mathrm{e}^{\mathrm{i}\xi\alpha^{-1/\alpha}Z_1}\right]\right)\mathrm{d}\xi &\geq \int_{-1}^{1}\mathrm{e}^{-\alpha^{-1}|\xi|^\alpha}\left[-|\xi|^\alpha\Omega(\eta) + |\xi|^2\Omega(\eta^{2/\alpha-1})\right]\mathrm{d}\xi\\ &= -\Omega(\eta) + \Omega(\eta^{2/\alpha-1}) = \Omega(\eta^{2/\alpha-1}).\end{aligned} \tag{B.5}$$

Define

$$h(x) := \frac{1}{M}\left(\frac{\sin x}{x}\mathbb{1}_{\{x\neq 0\}} + \mathbb{1}_{\{x=0\}}\right), \quad x \in \mathbb{R},$$

where

$$M := \sup_{x\in\mathbb{R}\setminus\{0\}}\left|\frac{x\cos x - \sin x}{x^2}\right| \in (0,\infty).$$

Since $h \in \mathrm{Lip}(1)$ and

$$h(x) = \frac{1}{2M}\int_{-1}^{1}\mathrm{e}^{\mathrm{i}\xi x}\,\mathrm{d}\xi, \quad x \in \mathbb{R},$$

it follows from Fubini's theorem and (B.5) that

$$
\begin{aligned}
W_1(\mu,\tilde\mu_\eta) &\ge \left|\mathbb{E}\left[h(\tilde Y_\eta)\right]-\mathbb{E}\left[h(\alpha^{-1/\alpha}Z_1)\right]\right| \\
&= \left|\int_{\mathbb{R}}\left(\frac{1}{2M}\int_{-1}^{1}\mathrm{e}^{\mathrm{i}\xi x}\,\mathrm{d}\xi\right)\mathbb{P}(\tilde Y_\eta\in\mathrm{d}x)-\int_{\mathbb{R}}\left(\frac{1}{2M}\int_{-1}^{1}\mathrm{e}^{\mathrm{i}\xi x}\,\mathrm{d}\xi\right)\mathbb{P}(\alpha^{-1/\alpha}Z_1\in\mathrm{d}x)\right| \\
&= \left|\frac{1}{2M}\int_{-1}^{1}\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\xi\tilde Y_\eta}\right]\mathrm{d}\xi-\frac{1}{2M}\int_{-1}^{1}\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\xi\alpha^{-1/\alpha}Z_1}\right]\mathrm{d}\xi\right| \\
&= \frac{1}{2M}\left|\int_{-1}^{1}\left(\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\xi\tilde Y_\eta}\right]-\mathbb{E}\left[\mathrm{e}^{\mathrm{i}\xi\alpha^{-1/\alpha}Z_1}\right]\right)\mathrm{d}\xi\right| \\
&\ge \Omega(\eta^{2/\alpha-1}).
\end{aligned}
$$

Combining this with the upper bound in Theorem 1.2 (2), finishes the proof. □

Acknowledgement. We are grateful to the associate editor and the anonymous referees for their helpful comments and professional handling. The research of L. Xu is supported in part by NSFC No. 12071499, Macao S.A.R. grant FDCT 0090/2019/A2 and University of Macau grant MYRG2020-00039-FST. R.L. Schilling was supported through the joint Polish–German NCN–DFG 'Beethoven 3' grant (NCN 2018/31/G/ST1/02252; DFG SCHI 419/11-1). P. Chen is supported by the NSF of Jiangsu Province grant BK20220867 and the Initial Scientific Research Fund of Young Teachers in Nanjing University of Aeronautics and Astronautics (1008-YAH21111). C.-S. Deng is supported by Natural Science Foundation of Hubei Province of China (2022CFB129).

References

1. V. Bally and D. Talay (1996): The law of the Euler scheme for stochastic differential equations. *Probability Theory and Related Fields* **104**, 43–60.
2. C. Berg and G. Forst (1975): *Potential Theory on Locally Compact Abelian Groups*. Springer, Ergebnisse der Mathematik und ihrer Grenzgebiete. II. Ser. Bd. **87**, Berlin.
3. R.M. Blumenthal and R.K. Getoor (1960): Some theorems on stable processes. *Transactions of the American Mathematical Society* **95**, 263–273.
4. B. Böttcher, R. L. Schilling and J. Wang (2013): *Lévy-Type Processes: Construction, Approximation and Sample Path Properties*. Lecture Notes in Mathematics **2099**, Lévy Matters III, Springer, Cham.
5. J.M. Chambers, C.L. Mallows and B. Stuck (1976): A method for simulating stable random variables. *Journal of the American Statistical Association* **71**, 340–344.
6. P. Chen, I. Nourdin, L. Xu and X. Yang (2019): Multivariate stable approximation in Wasserstein distance by Stein's method. Preprint *arXiv*: 1911.12917.
7. P. Chen, J. Lu and L. Xu (2022): Approximation to stochastic variance reduced gradient Langevin dynamics by stochastic delay differential equations. *Applied Mathematics & Optimization* **85**, 1–40.
8. P. Chen and L. Xu (2019): Approximation to stable law by the Lindeberg principle. *Journal of Mathematical Analysis and Applications* **480**, 123338.
9. K. Dareiotis and M. Gerencs (2020): On the regularisation of the noise for the Euler-Maruyama scheme with irregular drift. *Electronic Journal of Probability* **25**, 1–18.
10. C.-S. Deng, R.L. Schilling and Y.-H. Song (2017): Subgeometric rates of convergence for Markov processes under subordination. *Advances in Applied Probability* **49**, 162–181.
11. A. Eberle and A. Guillin (2019): Couplings and quantitative contraction rates for Langevin dynamics. *The Annals of Probability* **47**, 1982–2010.
12. A. Eberle, A. Guillin and R. Zimmer (2019): Quantitative Harris-type theorems for diffusions and McKean-Vlasov processes. *Transactions of the American Mathematical Society* **371**, 7135–7173.
13. X. Fang, Q.-M. Shao and L. Xu (2019): Multivariate approximations in Wasserstein distance by Stein's method and Bismut's formula. *Probability Theory and Related Fields* **174**, 945–979.
14. W. Fang and M.B. Giles (2016): Adaptive Euler-Maruyama method for SDEs with non-globally Lipschitz drift. In: *International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*. Springer, Cham, pp. 217–234.

15. Y. Guan and J. Wu (2017): Exponential ergodicity for non-Lipschitz multivalued stochastic differential equations with Lévy jumps. *Infinite Dimensional Analysis, Quantum Probability and Related Topics* **20**, 1750002.
16. P. Hall (1981): Two-sided bounds on the rate of convergence to a stable law. *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete* **57**, 349–364.
17. J. Jacod (2004): The Euler scheme for Lévy driven stochastic differential equations: limit theorems. *The Annals of Probability* **32**, 1830–1872.
18. X. Jin, G. Pang, L. Xu and X. Xu (2021): An approximation to steady-state of $M/Ph/n+M$ queue. Preprint *arXiv*:2109.03623.
19. A. Janicki, Z. Michna and A. Weron (1996): Approximation of stochastic differential equations driven by α-stable Lévy motion. *Applicationes Mathematicae* **24**, 149–168.
20. F. Kühn and R.L. Schilling (2019): Strong convergence of the Euler-Maruyama approximation for a class of Lévy-driven SDEs. *Stochastic Processes and their Applications* **129**, 2654–2680.
21. M. Liang and J. Wang (2020): Gradient estimates and ergodicity for SDEs driven by multiplicative Lévy noises via coupling. *Stochastic Processes and their Applications* **130**, 3053–3094.
22. V. Lemaire (2007): An adaptive scheme for the approximation of dissipative systems. *Stochastic Processes and their Applications* **117**, 1491–1518.
23. X. Li, Q. Ma, H. Yang and C. Yuan (2018): The numerical invariant measure of stochastic differential equations with Markovian switching. *SIAM Journal on Numerical Analysis* **56**, 1435–1455.
24. J. Lu, Y. Tan and L. Xu (2020): Central limit theorem and self-normalized Cramér-type moderate deviation for Euler-Maruyama scheme. *Bernoulli* **28**, 937–964.
25. D. Luo and J. Wang: (2019): Refined basic couplings and Wasserstein-type distances for SDEs with Lévy noises. *Stochastic Processes and their Applications* **129**, 3129–3173.
26. S.P. Meyn and R.L. Tweedie (1992): Stability of Markovian processes I: Criteria for discrete-time chains. *Advances in Applied Probability* **24**, 542–574.
27. S.P. Meyn and R.L. Tweedie (1993): Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes. *Advances in Applied Probability* **25**, 518–548.
28. R. Mikulevičius and F. Xu (2018): On the rate of convergence of strong Euler approximation for SDEs driven by Lévy processes. *Stochastics* **90**, 569–604.
29. R. Modarres and J.P. Nolan (1994): A method for simulating stable random vectors. *Computational Statistics* **9**, 11–19.
30. T. H. Nguyen, U. Simsekli and G. Richard (2019): Non-asymptotic analysis of Fractional Langevin Monte Carlo for non-convex optimization. In: *International Conference on Machine Learning*. Proceedings of Machine Learning Research, pp. 4810–4819.
31. J.P. Nolan (2008): An overview of multivariate stable distributions. Online: http://hdl.handle.net/1961/auislandora:68717 (accessed January 12, 2023).
32. J.P. Nolan (2020): *Univariate stable distributions: models for heavy tailed data*. Springer Series in Operations Research and Financial Engineering, Springer, Cham.
33. J. Norris (1986): Simplified Malliavin calculus. In: *Séminaire de Probabilités XX 1984/85*. Springer, Berlin, pp. 101–130.
34. G. Pagès and F. Panloup (2020): Unadjusted Langevin algorithm with multiplicative noise: Total variation and Wasserstein bounds. Preprint *arXiv*:2012.14310.
35. F. Panloup (2008): Recursive computation of the invariant measure of a stochastic differential equation driven by a Lévy process. *The Annals of Applied Probability* **18**, 379–426.
36. O.M. Pamen and D. Taguchi (2017): Strong rate of convergence for the Euler-Maruyama approximation of SDEs with Hölder continuous drift coefficient. *Stochastic Processes and their Applications* **127**, 2542–2559.
37. P.E. Protter (2004): *Stochastic Integration and Differential Equations* 2nd ed. Springer, Berlin.
38. P. Protter and D. Talay (1997): The Euler scheme for Lévy driven stochastic differential equations. *The Annals of Probability* **25**, 393–423.
39. J.M. Sanz-Serna and K.C. Zygalakis (2021): Wasserstein distance estimates for the distributions of numerical approximations to ergodic stochastic differential equations. *Journal of Machine Learning Research* **22**, 1–37.
40. K. Sato (1999): *Lévy processes and infinitely divisible distributions*. Cambridge University Press, Cambridge.
41. J. Shao (2018): Weak convergence of Euler-Maruyama's approximation for SDEs under integrability condition. Preprint *arXiv*:1808.07250.
42. U. Simsekli, L. Sagun and M. Gurbuzbalaban (2019): A tail-index analysis of stochastic gradient noise in deep neural networks. In: *International Conference on Machine Learning*. Proceedings of Machine Learning Research, pp. 5827–5837.

43. D. Talay (1990): Second-order discretization schemes of stochastic differential systems for the computation of the invariant law. *Stochastics: An International Journal of Probability and Stochastic Processes* **29**, 13–36.
44. B. Tarami and M. Avaji (2018): Convergence of Euler-Maruyama Method for Stochastic Differential Equations Driven by α-stable Lévy Motion. *Journal of Mathematical Extension* **12**, 31–53.
45. F. Y. Wang (2020): Exponential contraction in Wasserstein distances for diffusion semigroups with negative curvature. *Potential Analysis* **53**, 1123–1144.
46. J. Wang (2008): Criteria for ergodicity of Lévy type operators in dimension one. *Stochastic processes and their applications* **118**, 1909-1928.
47. J. Wang (2016): L^p-Wasserstein distance for stochastic differential equations driven by Lévy processes. *Bernoulli* **22**, 1598–1616.
48. L. Xu (2019): Approximation of stable law in Wasserstein-1 distance by Stein's method. *Annals of Applied Probability* **29**, 458–504.
49. X. Zhang (2013): Derivative formulas and gradient estimates for SDEs driven by α-stable processes. *Stochastic Processes and their Applications* **123**, 1213–1228.
50. P. Zhou, J. Feng, C. Ma, C. Xiong, S.C.H. Hoi (2020): Towards theoretically understanding why sgd generalizes better than adam in deep learning. *Advances in Neural Information Processing Systems* **33**, 21285–21296.

(P. Chen) School of Mathematics, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China
E-mail address: chenpengmath@nuaa.edu.cn

(C.-S. Deng) School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China
E-mail address: dengcs@whu.edu.cn

(R.L. Schilling) TU Dresden, Fakultät Mathematik, Institut für Mathematische Stochastik, 01062 Dresden, Germany
E-mail address: rene.schilling@tu-dresden.de

(L. Xu) 1. Department of Mathematics, Faculty of Science and Technology, University of Macau, Macau S.A.R., China. 2. Zhuhai UM Science & Technology Research Institute, Zhuhai, China
E-mail address: lihuxu@um.edu.mo