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 $\alpha$-stable Lévy process ($1 < \alpha < 2$): an approximation scheme with the $\alpha$-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 $\eta^{\frac{\alpha}{2}-1}$, respectively, where $\epsilon \in (0, 1)$ is arbitrary and $\eta$ is the step size of the approximation schemes. For the Pareto-driven scheme, an explicit calculation for Ornstein–Uhlenbeck $\alpha$-stable process shows that the rate $\eta^{\frac{\alpha}{2}-1}$ cannot be improved.

1. Introduction

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

$$dX_t = b(X_t) \, dt + dZ_t, \quad X_0 = x,$$

where $x \in \mathbb{R}^d$ is the starting point, $(Z_t)_{t \geq 0}$ is a $d$-dimensional, rotationally invariant $\alpha$-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

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

see, e.g. [45, 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 $\alpha$-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 $\widetilde{Z}_1, \widetilde{Z}_2, \ldots$ be an iid sequence of $d$-dimensional random vectors, which are Pareto distributed, i.e.

$$\widetilde{Z}_1 \sim p(z) = \frac{\alpha}{\sigma d_{-1}|z|^\alpha} \mathbb{I}_{(1, \infty)}(|z|);$$

2020 Mathematics Subject Classification. 60H10; 37M25; 60G51; 60H07; 60H35; 60G52.

Key words and phrases. Euler-Maruyama method; invariant measure; convergence rate; Wasserstein distance.
where \( \alpha \) is the step size, \( \sigma^\alpha = \alpha / (\sigma_{d-1} C_{d,\alpha}) \), and

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

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, 42, 9, 35] for SDEs driven by a Brownian motion, and [19, 39, 17, 37, 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 discretization 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 \( \infty \) as \( T \to \infty \).

The recent use of Langevin samplings in machine learning, has caused a surge of interest in the invariant error bounds for the invariant measures of the solution of the SDE and of the EM discretization, see e.g. [30, 43, 51, 7, 18]. We refer the reader to [23, 40, 44] for discrete schemes for the invariant measure of SDEs driven by a Brownian motion. Panloup [36] 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 \( \alpha \)-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 \( (\tilde{Z}_k)_{k \geq 1} \) 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 \( \eta^{2-1} \) 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, 49]. This is due to the fact that \( \alpha \)-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 [50] 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 \( Y_t^y \) and \( Y_t^y \) 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.

\[
P_t f(x) = \mathbb{E} f(X_t^x), \quad Q_k f(x) = \mathbb{E} f(Y_t^x), \quad \text{and} \quad \tilde{Q}_k f(x) = \mathbb{E} f(\tilde{Y}_t^x).
\]

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, \( C(\mathbb{R}^d, \mathbb{R}) \) denote the continuous functions \( f : \mathbb{R}^d \to \mathbb{R} \), and \( C^2(\mathbb{R}^d, \mathbb{R}) [C^0(\mathbb{R}^d, \mathbb{R})] \) 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

\[
\nabla_v f(x) = \langle \nabla f(x), v \rangle = \lim_{\varepsilon \to 0} \frac{f(x + \varepsilon v) - f(x)}{\varepsilon},
\]

\[
\nabla_v \nabla_v f(x) = \langle \nabla^2 f(x), v_1 v_2^\top \rangle = \lim_{\varepsilon \to 0} \frac{\nabla_v f(x + \varepsilon v_2) - \nabla_v f(x)}{\varepsilon},
\]

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

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

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

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

The Wasserstein-1 distance between two probability measures \( \mu_1 \) and \( \mu_2 \) on \( \mathbb{R}^d \) is defined as

\[ W_1(\mu_1, \mu_2) = \inf_{(X,Y) \in \mathcal{C}(\mu_1, \mu_2)} \mathbb{E}|X - Y|, \]

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

\[
W_1(\mu_1, \mu_2) = \sup_{h \in \text{Lip}(1)} |\mu_1(h) - \mu_2(h)|,
\]

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

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 \( |x| \) 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

\[
\langle b(x) - b(y), x - y \rangle \leq -\theta_1 |x - y|^2 + K \quad \forall x, y \in \mathbb{R}^d
\]
and
\( |\nabla_v b(x)| \leq \theta_2 |v|, \quad |\nabla_{v_1} \nabla_{v_2} b(x)| \leq \theta_3 |v_1| |v_2| \quad \forall v, v_1, v_2, x \in \mathbb{R}^d. \)

Remark 1.1. Note that (1.8) immediately implies the following linear growth condition
\( |b(x) - b(0)| \leq \theta_2 |x|, \quad x \in \mathbb{R}^d. \)

Under Assumption A, we will show that both \((X_t)_{t \geq 0}\) and \((\tilde{Y}_k)_{k \geq 0}\) are ergodic; we write \(\mu\) 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 \(\lambda\) may depend on \(\theta_1, \theta_2, \theta_3, K, \alpha, d, |b(0)|\) and \(\beta\) 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 \((X_t)_{t \geq 0}\) and \((\tilde{Y}_k)_{k \geq 0}\) be defined by (1.1) and (1.4) (step size \(\eta\)), and denote by \(\mu\) 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 \(N \geq 2\) and step size \(\eta < \min \{1, \theta_1/(8\theta_3^2), 1/\theta_1\}\), one has
   \[ W_1(\text{law}(X_{\eta N}), \text{law}(\tilde{Y}_N)) \leq C(1 + |x|)\eta^{2/\alpha - 1}. \]
2. For every step size \(\eta < \min \{1, \theta_1/\theta_2, 1/\theta_1\}\), one has
   \[ W_1(\mu, \tilde{\mu}_\eta) \leq C\eta^{2/\alpha - 1}. \]

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

1. For every \(N \geq 2\) and step size \(\eta < \min \{1, \theta_1/(8\theta_3^2), 1/\theta_1\}\), one has
   \[ W_1(\text{law}(X_{\eta N}), \text{law}(Y_N)) \leq C(1 + |x|)^{\beta} \eta^{1 + \frac{1}{\alpha} - \frac{1}{\beta}}. \]
2. For every step size \(\eta < \min \{1, \theta_1/\theta_2, 1/\theta_1\}\), one has
   \[ W_1(\mu, \mu_\eta) \leq C\eta^{1 + \frac{1}{\alpha} - \frac{1}{\beta}}. \]

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 \((X_t)_{t \geq 0}\) and \((Y_k)_{k \geq 0}\). 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 \((X_t)_{t \geq 0}\) the solution to the SDE (1.1). Then, \((X_t)_{t \geq 0}\) admits a unique invariant probability measure \(\mu\) such that for \(1 \leq \beta < \alpha\)
\[ \sup_{|f| \leq V_\beta} |E[f(X_t^\beta)] - \mu(f)| \leq c_1 V_\beta(x) e^{-ct}, \quad t > 0, \]
for some constants \(c_1, c_2 > 0\). In particular, there exists a constant \(C > 0\) such that
\[ E|X_t^\beta| \leq C(1 + |x|^\beta), \quad t > 0. \]
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

\[
W_1(\text{law}(X^x_t), \text{law}(X^y_t)) \leq Ce^{-\lambda t}|x - y|.
\]

Proposition 1.7. Let Assumption A hold and denote by \((Y_k)_{k \geq 0} \) and \((\tilde{Y}_k)_{k \geq 0}\) 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 \((Y_k)_{k \geq 0}\) admits a unique invariant measure \( \mu_\eta \), such that for all \( x \in \mathbb{R}^d \) and \( k > 0 \),

\[
\sup_{|f| \leq V_k} |\mathbb{E}f(Y_k^x) - \mu_\eta(f)| \leq c_1 V_1(x)e^{-c_2k},
\]

for some constants \( c_1, c_2 > 0 \).

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

\[
\sup_{|f| \leq V_k} |\mathbb{E}f(\tilde{Y}_k^x) - \tilde{\mu}_\eta(f)| \leq c_3 V_1(x)e^{-c_4k},
\]

for some constants \( c_3, c_4 > 0 \).

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

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

\[
\mathbb{E}|\tilde{Y}_k^x| \leq C(1 + |x|),
\]

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 \((X_t)_{t \geq 0}\).

Lemma 2.1. Let \( h \in \text{Lip}(1) \) and \( X^x_t \) 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

\[
|\nabla_v P_t h(x)| \leq e^{\theta_2} |v|
\]

and

\[
|\nabla_{v_1} \nabla_{v_1} P_t h(x)| \leq C t^{-1/\alpha} |v_1| |v_2|,
\]

for some constant \( C > 0 \).

Remark 2.2. We will prove the above lemma using Malliavin’s calculus [33] in Section 3 further down. A very careful and knowledgeable referee suggested a shorter argument which avoids the use of Bismut’s formula and Norris’ formalism. The idea of her/his proof is to use the so called "variance of constants method" to establish the corresponding Bismut–Elworthy–Li formula. The alternative proof of Lemma 2.1 will be added in Appendix C below.

Using the inequalities (1.11) and (1.14), we can obtain the following estimates:
Lemma 2.3. Let \((X_t)_{t \geq 0}\) be the solution to the SDE \((1.1)\) and \((Y_k)_{k \geq 0}\) be the Markov chains defined by \((1.2)\). If the step size satisfies \(\eta < \min \left\{1, \frac{\theta}{\min\{\beta, d\}}\right\}\), then the following estimates hold for all \(t \in (0, 1], \beta \in [1, \alpha)\):

\[
(2.3) \quad E|Y^x_t - x|^\beta \leq C(1 + |x|^\beta)\eta^{\beta/\alpha},
\]

\[
(2.4) \quad E|X^x_t - x|^\beta \leq C(1 + |x|^\beta)\eta^{\beta/\alpha},
\]

\[
(2.5) \quad E|X^x_\eta - Y^x_t|^\beta \leq C(1 + |x|^\beta)\eta^{\beta + 2\alpha/\pi}.
\]

**Proof.** The first inequality follows immediately from

\[
E|Y^x_t - x|^\beta = E|\eta b(x) + Z_\eta|^\beta \leq 2 \left[|\eta^\beta|b(x)|^\beta + E|Z_\eta|^\beta\right] \leq C(1 + |x|^\beta)\eta^{\beta/\alpha}.
\]

From the Hölder inequality and \((1.11)\), we obtain

\[
E|X^x_t - x|^\beta = E \left| \int_0^t b(X^x_s) \, ds + Z_t \right|^\beta
\]

\[
\leq 2E \left| \int_0^t b(X^x_s) \, ds \right|^\beta + 2|Z_t|^\beta
\]

\[
\leq 2t^{\beta-1} \int_0^t E|b(X^x_s)|^\beta \, ds + 2E|Z_t|^\beta t^{\beta/\alpha}
\]

\[
\leq C(1 + |x|^\beta)t^{\beta/\alpha},
\]

which implies the second inequality.

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

\[
E|X^x_\eta - Y^x_t|^\beta = E \left| \int_0^\eta [b(X^x_s) - b(x)] \, ds \right|^\beta
\]

\[
\leq \eta^{\beta-1} \int_0^\eta E|b(X^x_s) - b(x)|^\beta \, ds
\]

\[
\leq \theta \eta^{\beta-1} \int_0^\eta E|X^x_s - x|^\beta \, ds
\]

\[
\leq C(1 + |x|^\beta)\eta^{\beta-1} \int_0^\eta s^{\frac{\beta}{\alpha}} \, ds
\]

\[
\leq C(1 + |x|^\beta)\eta^{\beta + \frac{2\alpha}{\pi}}. \quad \Box
\]

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 \(\alpha\)-stable Lévy process process \((Z_t)_{t \geq 0}\), which is defined as a principal value (p.v.) integral: for any \(f \in C^2(\mathbb{R}^d, \mathbb{R})\),

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

**Lemma 2.4.** Let \(\alpha \in (1, 2)\) and \(f : \mathbb{R}^d \to \mathbb{R}\) satisfying \(\|\nabla f\|_{\infty} < \infty\) and \(\|\nabla^2 f\|_{\text{HS}, \infty} < \infty\).

For all \(x, y \in \mathbb{R}^d\) one has

\[
(2.7) \quad \left| (-\Delta)^{\alpha/2} f(x) - (-\Delta)^{\alpha/2} f(y) \right| \leq C_{d, \alpha}\|\nabla^2 f\|_{\text{HS}, \infty} \sigma_{d-1} (2 - \alpha)(\alpha - 1)|x - y|^{2-\alpha}.
\]
Proof. From the definition of the fractional Laplacian (2.6) and the symmetry of the representing measure we have for any \( R > 0 \)

\[
(-\Delta)^{\alpha/2} f(x) = C_{d,\alpha} \int_{S^{d-1}} \int_0^\infty \frac{f(x + r\theta) - f(x) - r \langle \theta, \nabla f(x) \rangle I_{(0,R)}(r)}{r^{\alpha+1}} \, dr \, d\theta
\]

\[
= C_{d,\alpha} \int_{S^{d-1}} \int_0^R \int_0^r \frac{\langle \theta, \nabla f(x + rs) - \nabla f(x) \rangle}{r^{\alpha+1}} \, ds \, dr \, d\theta
+ C_{d,\alpha} \int_{S^{d-1}} \int_R^\infty \int_0^r \frac{\langle \theta, \nabla f(x + rs) \rangle}{r^{\alpha+1}} \, ds \, dr \, d\theta
\]

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

\[
\left| (-\Delta)^{\alpha/2} f(x) - (-\Delta)^{\alpha/2} f(y) \right|
\leq C_{d,\alpha} \int_{S^{d-1}} \int_0^R \int_0^r \frac{|\nabla f(x + rs) - \nabla f(x) - \nabla f(y + rs) + \nabla f(y)|}{r^{\alpha+1}} \, ds \, dr \, d\theta
\]

\[
+ C_{d,\alpha} \int_{S^{d-1}} \int_R^\infty \int_0^r \frac{|\nabla f(x + rs) - \nabla f(y + rs)|}{r^{\alpha+1}} \, ds \, dr \, d\theta.
\]

For the first integral we have

\[
C_{d,\alpha} \int_{S^{d-1}} \int_0^R \int_0^r \frac{|\nabla f(x + rs) - \nabla f(x) - \nabla f(y + rs) + \nabla f(y)|}{r^{\alpha+1}} \, ds \, dr \, d\theta
\leq C_{d,\alpha} \int_{S^{d-1}} \int_0^R \int_0^r \frac{|\nabla f(x + rs) - \nabla f(x)| + |\nabla f(y + rs) - \nabla f(y)|}{r^{\alpha+1}} \, ds \, dr \, d\theta
\leq 2C_{d,\alpha} \|\nabla^2 f\|_{HS,\infty} \int_{S^{d-1}} \int_0^r \frac{s}{r^{\alpha+1}} \, ds \, dr \, d\theta = \frac{C_{d,\alpha} \|\nabla^2 f\|_{HS,\infty} \sigma_{d-1}}{2 - \alpha} R^{2-\alpha},
\]

and for the second term we get

\[
C_{d,\alpha} \int_{S^{d-1}} \int_R^\infty \int_0^r \frac{|\nabla f(x + rs) - \nabla f(y + rs)|}{r^{\alpha+1}} \, ds \, dr \, d\theta
\leq C_{d,\alpha} \|\nabla^2 f\|_{HS,\infty} \int_{S^{d-1}} \int_R^\infty \int_0^r \frac{|x - y|}{r^{\alpha+1}} \, ds \, dr \, d\theta
= \frac{C_{d,\alpha} \|\nabla^2 f\|_{HS,\infty} \sigma_{d-1}}{\alpha - 1} |x - y| R^{1-\alpha}.
\]

Hence, the assertion follows upon taking \( R = |x - y| \). \( \square \)

Lemma 2.5. Let \((X_t)_{t\geq0}\) and \((\tilde{Y}_k)_{k\geq0}\) be defined by (1.1) and (1.4), respectively. There exists a constant \( C > 0 \) such that for all \( x \in \mathbb{R}^d \), \( \eta \in (0,1) \), \( f : \mathbb{R}^d \rightarrow \mathbb{R} \) satisfying \( \|\nabla f\|_{\infty} < \infty \) and \( \|\nabla^2 f\|_{HS,\infty} < \infty \),

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

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

\[
\mathbb{E}[f(X_{\eta}^x) - f(\tilde{Y}_1)] = \mathbb{E} \left[ f \left( x + \int_0^{\eta} b(X_r^x) \, dr + Z_\eta \right) - f \left( x + \eta \bar{b}(x) + \eta^{1/\alpha} \tilde{Z} \right) \right]
= J_1 + J_2,
\]

7
where

\[
J_1 := \mathbb{E} \left[ f \left( x + \int_0^\eta b(X_r^x) \, dr + Z_\eta \right) - f \left( x + \eta b(x) + Z_\eta \right) \right],
\]

\[
J_2 := \mathbb{E} \left[ f \left( x + \eta b(x) + Z_\eta \right) - f \left( x + \eta b(x) \right) \right] - \mathbb{E} \left[ f \left( x + \eta b(x) + \frac{\eta^{1/\alpha}}{\sigma} \tilde{Z} \right) - f \left( x + \eta b(x) \right) \right].
\]

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

\[
|J_1| \leq \|\nabla f\|_\infty \mathbb{E} \left[ \int_0^\eta b(X_r^x) \, dr - \eta b(x) \right]
\leq \|\nabla f\|_\infty \int_0^\eta \mathbb{E} |b(X_r^x) - b(x)| \, dr
\leq \theta_2 \|\nabla f\|_\infty \int_0^\eta |X_r^x - x| \, dr
\leq C\theta_2(1 + |x|) \|\nabla f\|_\infty \int_0^\eta r^{1/\alpha} \, dr
\leq C(1 + |x|) \|\nabla f\|_\infty \eta^{1+1/\alpha}.
\]

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

\[
\mathbb{E} \left[ f \left( x + \eta b(x) + Z_\eta \right) - f \left( x + \eta b(x) \right) \right] = \int_0^\eta \mathbb{E} \left[ (-\Delta)^{\alpha/2} f \left( x + \eta b(x) + Z_r \right) \right] \, dr.
\]

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

\[
\mathbb{E} \left[ f \left( x + \eta b(x) + \frac{\eta^{1/\alpha}}{\sigma} \tilde{Z} \right) - f \left( x + \eta b(x) \right) \right]
= \frac{\eta^{1/\alpha}}{\sigma} \mathbb{E} \left[ \int_0^1 \left\langle \nabla f \left( x + \eta b(x) + \frac{\eta^{1/\alpha}}{\sigma} t \tilde{Z} \right), \tilde{Z} \right\rangle \, dt \right]
= \frac{\eta^{1/\alpha}}{\sigma} \int_{|z| \geq 1} \int_0^1 \alpha \left\langle \nabla f \left( x + \eta b(x) + \frac{\eta^{1/\alpha}}{\sigma} z \right), z \right\rangle \frac{dt \, dz}{\sigma_{d-1} |z|^{\alpha+d}}
= \frac{\alpha \eta}{\sigma_{d-1} \sigma^\alpha} \int_{|z| \geq 1} \int_0^1 \left\langle \nabla f \left( x + \eta b(x) + t z \right), z \right\rangle \frac{dt \, dz}{z^{\alpha+d}}
= \eta (-\Delta)^{\alpha/2} f (x + \eta b(x)) - R,
\]

where

\[
R := \eta C_{d,\alpha} \int_{|z| < \sigma^{-1} \eta^{1/\alpha}} \int_0^1 \left\langle \nabla f \left( x + \eta b(x) + t z \right), z \right\rangle \frac{dt \, dz}{z^{\alpha+d}}.
\]

Together, the above estimates yield

\[
|J_2| \leq |R| + \left| \int_0^\eta \mathbb{E} \left[ (-\Delta)^{\alpha/2} f \left( x + \eta b(x) + Z_r \right) \right] \, dr - \eta (-\Delta)^{\alpha/2} f (x + \eta b(x)) \right|.
\]
Further, we have
\[ |R| = \eta C_{d,\alpha} \int_{|z|<\sigma^{-1}\eta^{1/\alpha}} \int_0^1 \langle \nabla f(x + \eta b(x) + tz) - \nabla f(x + \eta b(x)), z \rangle \frac{dt \, dz}{|z|^\alpha + d} \]
\[ \leq \eta C_{d,\alpha} \int_{|z|<\sigma^{-1}\eta^{1/\alpha}} \int_0^1 |\nabla f(x + \eta b(x) + tz) - \nabla f(x + \eta b(x))| \frac{dt \, dz}{|z|^\alpha + d - 1} \]
\[ \leq \frac{1}{2} \eta C_{d,\alpha} \|\nabla^2 f\|_{\text{HS},\infty} \int_{|z|<\sigma^{-1}\eta^{1/\alpha}} \frac{dz}{|z|^\alpha + d - 2} \leq C \|\nabla^2 f\|_{\text{HS},\infty} \eta^{2/\alpha}. \]

By Lemma 2.4, we also have
\[ \left| \int_0^n \mathbb{E} \left[ (-\Delta)^{\alpha/2} f(x + \eta b(x) + Z_r) \right] dr - \eta (-\Delta)^{\alpha/2} f(x + \eta b(x)) \right| \]
\[ \leq \int_0^n \mathbb{E} \left| (-\Delta)^{\alpha/2} f(x + \eta b(x) + Z_r) - (-\Delta)^{\alpha/2} f(x + \eta b(x)) \right| dr \]
\[ \leq C \|\nabla^2 f\|_{\text{HS},\infty} \int_0^\eta \mathbb{E} \left[ |Z_r|^{2-\alpha} \right] r^{2-\alpha} \, dr \]
\[ = C \|\nabla^2 f\|_{\text{HS},\infty} \int_0^\eta \mathbb{E} \left[ |Z_1|^{2-\alpha} \right] r^{2-\alpha} \, dr \]
\[ \leq C \mathbb{E} \left[ |Z_1|^{2-\alpha} \right] \|\nabla^2 f\|_{\text{HS},\infty} \eta^{2/\alpha}. \]

The proof follows if we combine all estimates. \(\square\)

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 2.6.** Assume that \( f \) satisfies \( \|\nabla f\|_{\infty} < \infty \) and \( \|\nabla^2 f\|_{\text{HS},\infty} < \infty \). For any \( \beta \in [1, 2] \) and \( x, y \in \mathbb{R}^d \), we have
\[ |\nabla f(x) - \nabla f(y)| \leq (2\|\nabla f\|_{\infty} + \|\nabla^2 f\|_{\text{HS},\infty}) |x - y|^{\beta-1}. \]

**Proof.** For \( |x - y| > 1 \) we have
\[ |\nabla f(x) - \nabla f(y)| \leq 2\|\nabla f\|_{\infty} \leq \|\nabla f\|_{\infty} |x - y|^{\beta-1}, \]
and for \( |x - y| \leq 1 \) we have
\[ |\nabla f(x) - \nabla f(y)| \leq \|\nabla^2 f\|_{\text{HS},\infty} |x - y| \leq \|\nabla^2 f\|_{\text{HS},\infty} |x - y|^{\beta-1}. \]
\(\square\)

**Lemma 2.7.** Let \( (X_t)_{t \geq 0} \) and \( (Y_t)_{t \geq 0} \) 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 \( \|\nabla^2 f\|_{\text{HS},\infty} < \infty \),
\[ |P_\eta f(x) - Q_\eta f(x)| \leq C(1 + |x|^{\beta}) (\|\nabla f\|_{\infty} + \|\nabla^2 f\|_{\text{HS},\infty}) \eta^{2+\frac{1}{\beta}-\frac{1}{\alpha}}. \]

**Proof.** We use a Taylor expansion to get
\[ \mathbb{E} f(X^\eta_t) - \mathbb{E} f(Y^\eta_t) \]
\[ = \mathbb{E} \langle \nabla f(Y^\eta_t), X^\eta_t - Y^\eta_t \rangle + \mathbb{E} \int_0^1 \langle \nabla f(Y^\eta_t + r(Y^\eta_t - Y^\eta_t)) - \nabla f(Y^\eta_t), X^\eta_t - Y^\eta_t \rangle \, dr \]
\[ = \mathbb{E} \langle \nabla f(x + \eta b(x) + Z_\eta) - \nabla f(x + \eta b(x)), X^\eta_t - Y^\eta_t \rangle + \mathbb{E} \left[ \langle \nabla f(x + \eta b(x)), X^\eta_t - Y^\eta_t \rangle \right] \]
\[ + \mathbb{E} \int_0^1 \langle \nabla f(Y^\eta_t + r(Y^\eta_t - Y^\eta_t)) - \nabla f(Y^\eta_t), X^\eta_t - Y^\eta_t \rangle \, dr \]
\[ =: I + II + III. \]
For the first term $I$ we have

$$I = E \left( \nabla f (x + \eta b(x) + Z_\eta) - \nabla f (x + \eta b(x)) , X_\eta^x - Y_1^x \right) \left( I_{(0,1)}(|Z_\eta|) + I_{(1,\infty)}(|Z_\eta|) \right)$$

$$=: I_1 + I_2.$$ 

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

$$P (|Z_\eta| > 1) \leq c\eta \quad \text{and} \quad E (|Z_\eta|^\lambda 1_{(0,1)}(|Z_\eta|)) \leq C\eta.$$

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

$$|I_1| \leq E \left( \nabla f (x + \eta b(x) + Z_\eta) - \nabla f (x + \eta b(x)) | 1_{(0,1)}(|Z_\eta|) | X_\eta^x - Y_1^x | \right)$$

$$\leq \|\nabla^2 f\|_{\text{HS},\infty} E \left( |Z_\eta| 1_{(0,1)}(|Z_\eta|) | X_\eta^x - Y_1^x | \right)$$

$$\leq \|\nabla^2 f\|_{\text{HS},\infty} (E \left( |Z_\eta| |X_\eta^x - Y_1^x|^\beta \right) \right)^{\frac{1}{\beta}} \right)^{\frac{\beta}{\beta - 1}}$$

$$\leq C (1 + |x|) \|\nabla^2 f\|_{\text{HS},\infty} \eta^{2+\frac{1}{\beta} - \frac{1}{\beta'}}$$

whereas by the Hölder inequality

$$|I_2| \leq E \left( \nabla f (x + \eta b(x) + Z_\eta) - \nabla f (x + \eta b(x)) | 1_{(1,\infty)}(|Z_\eta|) | X_\eta^x - Y_1^x | \right)$$

$$\leq 2 \|\nabla f\|_{\text{HS},\infty} (E \left( |Z_\eta| 1_{(1,\infty)}(|Z_\eta|) | X_\eta^x - Y_1^x | \right)$$

$$\leq 2 \|\nabla f\|_{\text{HS},\infty} (E \left( |Z_\eta| |X_\eta^x - Y_1^x|^\beta \right) \right)^{\frac{1}{\beta}} \right)^{\frac{\beta}{\beta - 1}}$$

$$\leq C (1 + |x|) \|\nabla f\|_{\text{HS},\infty} \eta^{2+\frac{1}{\beta} - \frac{1}{\beta'}}$$

Hence, we have

$$|I| \leq C (1 + |x|) \left( \|\nabla f\|_{\text{HS},\infty} + \|\nabla^2 f\|_{\text{HS},\infty} \right) \eta^{2+\frac{1}{\beta'} - \frac{1}{\beta}}.$$

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

$$II = E \left( \nabla f (x + \eta b(x)) , \int_0^\eta [b(X_\eta^x) - b(x)] \, ds \right)$$

$$= \left( \nabla f (x + \eta b(x)) , \int_0^\eta E \left[ b(X_\eta^x) - b(x) \right] \, ds \right)$$

$$= \left( \nabla f (x + \eta b(x)) , \int_0^\eta \int_0^s E \left[ (\nabla b(X_\eta^x), b(X_\eta^x)) + (-\Delta)^{\frac{\alpha}{2}} b(X_\eta^x) \right] \, dr \, ds \right)$$

$$\leq C \|\nabla f\|_{\text{infty}} (1 + |x|) \eta^2.$$ 

In the last inequality we use the estimate (1.9) (for $b(x_\eta^x)$) and Lemma 2.4 (for $(-\Delta)^{\alpha/2} b(X_\eta^x)$), combined with the moment estimate (2.4).
Finally, III is estimated by Lemma 2.6 and (2.5),
\[ III \leq C \left( \| \nabla f \|_\infty + \| \nabla^2 f \|_{\text{HS}, \infty} \right) \mathbb{E} |X_\eta^x - Y_1^x|^{\beta} \]
\[ \leq C (1 + |x|) \left( \| \nabla f \|_\infty + \| \nabla^2 f \|_{\text{HS}, \infty} \right) \eta^{\beta + \frac{\sigma}{2}} \]
\[ \leq C (1 + |x|) \left( \| \nabla f \|_\infty + \| \nabla^2 f \|_{\text{HS}, \infty} \right) \eta^2. \]
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 \( h \in \text{Lip}(1) \)
\[ P_{N\eta} h(x) - \tilde{Q}_N h(x) = \sum_{i=1}^{N} \tilde{Q}_{i-1} (P_{\eta} - \tilde{Q}_1) P_{(N-i)\eta} h(x). \]
Then we have
\[ W_1 \left( \text{law} (X_{\eta N}), \text{law} (\tilde{Y}_N) \right) = \sup_{h \in \text{Lip}(1)} \left| \sum_{i=1}^{N-1} \tilde{Q}_{i-1} (P_{\eta} - \tilde{Q}_1) P_{(N-i)\eta} h(x) \right| \]
\[ + \sup_{h \in \text{Lip}(1)} \left| \tilde{Q}_{N-1} (P_{\eta} - \tilde{Q}_1) h(x) \right|. \]
First, we bound the last term. By (2.4) with \( \beta = 1 \), (1.4) and (1.8), for \( h \in \text{Lip}(1) \) and \( \eta < 1 \),
\[ \left| (P_{\eta} - \tilde{Q}_1) h(x) \right| \leq \left| \mathbb{E} h(X_{\eta}^x) - \mathbb{E} \tilde{Y}_1 \right| \]
\[ \leq \left| \mathbb{E} h(X_{\eta}^x) - h(x) \right| + \left| \mathbb{E} \tilde{Y}_1 - h(x) \right| \]
\[ \leq \mathbb{E} |X_{\eta}^x - x| + \mathbb{E} |\tilde{Y}_1 - x| \]
\[ \leq C (1 + |x|) \eta^{1/\alpha} + \eta |b(x)| + \sigma^{-1} \eta^{1/\alpha} \mathbb{E} |\tilde{Z}_1| \]
\[ \leq C (1 + |x|) \eta^{1/\alpha} + \eta^{1/\alpha} (|b(0)| + \theta_2 |x|) + \sigma^{-1} \eta^{1/\alpha} \mathbb{E} |\tilde{Z}_1| \]
\[ \leq C (1 + |x|) \eta^{1/\alpha}. \]
Together with (1.15) we get
\[ \sup_{h \in \text{Lip}(1)} \left| \tilde{Q}_{N-1} (P_{\eta} - \tilde{Q}_1) h(x) \right| \leq C (1 + \mathbb{E} |\tilde{Y}_N^x|) \eta^{1/\alpha} \leq C (1 + |x|) \eta^{2/\alpha - 1}. \]
Next, we bound the first term in (2.9); we distinguish between two cases:
**Case 1:** \( N \leq \eta^{-1} + 1 \). By Lemmas 2.5 and 2.1,
\[ \left| (P_{\eta} - \tilde{Q}_1) P_{(N-i)\eta} h(x) \right| \leq C (1 + |x|) \left( \| \nabla P_{(N-i)\eta} h \|_\infty + \| \nabla^2 P_{(N-i)\eta} h \|_{\text{HS}, \infty} \right) \eta^{2/\alpha} \]
\[ \leq C (1 + |x|) \left( (N - i) \eta \right)^{-1/\alpha} \eta^{2/\alpha}. \]
Combining this with (1.15), we get
\[ \sup_{h \in \text{Lip}(1)} \left| \tilde{Q}_{i-1} (P_{\eta} - \tilde{Q}_1) P_{(N-i)\eta} h(x) \right| \leq C (1 + \mathbb{E} |\tilde{Y}_{N-1}^x|) \left( (N - i) \eta \right)^{-1/\alpha} \eta^{2/\alpha} \]
\[ \leq C (1 + |x|) \left( (N - i) \eta \right)^{-1/\alpha} \eta^{2/\alpha}. \]
Since $N - 1 \leq \eta^{-1}$,
\[ \sum_{i=1}^{N-1} (N - i) \eta^{-1/\alpha} = \eta^{-1/\alpha} \sum_{i=1}^{N-1} i^{-1/\alpha} \leq \eta^{-1/\alpha} \int_0^{N-1} r^{-1/\alpha} \, dr = \frac{\alpha}{\alpha - 1} \eta^{-1/\alpha} (N - 1)^{-1/\alpha+1} \leq \frac{\alpha}{\alpha - 1} \eta^{-1}. \]

This gives the upper bound
\[ \sum_{i=1}^{N-1} \sup_{h \in \text{Lip}(1)} \left| \tilde{Q}_{i-1} (P_\eta - \tilde{Q}_1) P_{(N-i)\eta} h(x) \right| \leq C (1 + |x|) \eta^{2/\alpha} \sum_{i=1}^{N-1} [(N - i) \eta]^{-1/\alpha} \]
\[ \leq C \frac{\alpha}{\alpha - 1} (1 + |x|) \eta^{2/\alpha-1}. \]

**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
\[ |P_i h(x) - P_i h(y)| \leq Ce^{-\lambda|x - y|}, \quad h \in \text{Lip}(1), \quad t \geq 0. \]

This implies that
\[ \sup_{h \in \text{Lip}(1)} \left| Q_{i-1} (P_\eta - \tilde{Q}_1) P_{(N-i)\eta} h(x) \right| = \sup_{h \in \text{Lip}(1)} \left| \tilde{Q}_{i-1} (P_\eta - \tilde{Q}_1) P_1 P_{(N-i)\eta-1} h(x) \right| \]
\[ \leq Ce^{-\lambda(N-i)\eta^{-1-1}} \sup_{g \in \text{Lip}(1)} \left| \tilde{Q}_{i-1} (P_\eta - \tilde{Q}_1) P_1 g(x) \right|, \]
where $i \leq [N - \eta^{-1}]$. By Lemmas 2.5 and 2.1,
\[ \left| (P_\eta - \tilde{Q}_1) P_1 g(x) \right| \leq C (1 + |x|) \left( \|
abla P_1 g\|_\infty + \|
abla^2 P_1 g\|_{\text{HS},\infty} \right) \eta^{2/\alpha} \leq C (1 + |x|) \eta^{2/\alpha}. \]

Combining this with (1.15), we get
\[ \sum_{i=1}^{[N-\eta^{-1}]} \sup_{h \in \text{Lip}(1)} \left| \tilde{Q}_{i-1} (P_\eta - \tilde{Q}_1) P_{(N-i)\eta} h(x) \right| \leq C \eta^{2/\alpha} \sum_{i=1}^{[N-\eta^{-1}]} e^{-\lambda(N-i)\eta^{-1-1}} \left( 1 + E|\tilde{y}_{i-1}^x| \right) \]
\[ \leq C (1 + |x|) \eta^{2/\alpha} \sum_{i=1}^{[N-\eta^{-1}]} e^{-\lambda(N-i)\eta^{-1-1}}. \]

Observe that
\[ \sum_{i=1}^{[N-\eta^{-1}]} e^{-\lambda(N-i)\eta^{-1-1}} = \sum_{i=|\eta^{-1}|}^{N-1} e^{-\lambda(n-1)} \leq e^\lambda \int_{|\eta^{-1}|}^{N-1} e^{-\lambda \eta r} \, dr \]
\[ \leq e^{\lambda \eta^{-1}} \int_0^\infty e^{-\lambda r} \, dr = \lambda^{-1} e^{\lambda \eta^{-1}}. \]

Thus, we get
\[ \sum_{i=1}^{[N-\eta^{-1}]} \sup_{h \in \text{Lip}(1)} \left| \tilde{Q}_{i-1} (P_\eta - \tilde{Q}_1) P_{(N-i)\eta} h(x) \right| \leq C \lambda^{-1} e^{\lambda (1 + |x|) \eta^{2/\alpha-1}}. \]

For $i \geq [N - \eta^{-1}] + 1$, by almost the same as the calculation in the first case, we find
\[ \sum_{i=[N-\eta^{-1}]+1}^{N-1} (N - i) \eta^{-1/\alpha} \leq \frac{\alpha}{\alpha - 1} \eta^{-1}. \]
Combining this with (2.11), we obtain
\[
\sum_{i=1}^{N-1} \sup_{h \in \text{Lip}(1)} |\tilde{Q}_{i-1}(P_\eta - \tilde{Q}_1) P_{(N-i)\eta} h(x)| \\
\leq C(1 + |x|) \eta^{2/\alpha} \sum_{i=1}^{N-1} [(N - i)\eta]^{-1/\alpha} \leq C \frac{\alpha}{\alpha - 1} (1 + |x|) \eta^{2/\alpha - 1}.
\]

We have just shown estimates for \(\sum_{i=1}^{[N-\eta^{-1}] - 1} \) and \(\sum_{i=1}^{N-1} \). Adding them up we arrive at
\[
\sum_{i=1}^{N-1} \sup_{h \in \text{Lip}(1)} |\tilde{Q}_{i-1}(P_\eta - \tilde{Q}_1) P_{(N-i)\eta} h(x)| \\
= \left( \sum_{i=1}^{[N-\eta^{-1}] - 1} + \sum_{i=1}^{N-1} \right) \sup_{h \in \text{Lip}(1)} |\tilde{Q}_{i-1}(P_\eta - \tilde{Q}_1) P_{(N-i)\eta} h(x)| \\
\leq C(1 + |x|) \eta^{2/\alpha - 1}.
\]

Both Case 1 and Case 2 lead to an estimate of the form
\[
\sum_{i=1}^{N-1} \sup_{h \in \text{Lip}(1)} |\tilde{Q}_{i-1}(P_\eta - \tilde{Q}_1) P_{(N-i)\eta} h(x)| \\
\leq C(1 + |x|) \eta^{2/\alpha - 1}.
\]

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
\[
\lim_{k \to \infty} W_1(\mu, (\text{law}(X_{\eta_k})) = \lim_{k \to \infty} W_1(\text{law}(\hat{Y}_k), \hat{\mu}_\eta) = 0.
\]

By the triangle inequality and Part (1) with \(x = 0\),
\[
W_1(\mu, \hat{\mu}_\eta) \leq W_1(\mu, (\text{law}(X_{\eta_k}))) + W_1(\text{law}(X_{\eta_k}), \text{law}(\hat{Y}_N)) + W_1(\text{law}(\hat{Y}_N), \hat{\mu}_\eta) \\
\leq W_1(\mu, (\text{law}(X_{\eta_k}))) + C \eta^{2/\alpha - 1} + W_1(\text{law}(\hat{Y}_N), \hat{\mu}_\eta).
\]

Letting \(N \to \infty\) finishes the proof. \(\square\)

**Proof of Theorem 1.3.** In the above proof, replacing Lemma 2.5 and (1.13) with Lemma 2.7 and (1.12), the proof of Theorem 1.3 is similar to that of Theorem 1.2. \(\square\)

3. **Malliavin calculus and the proof of Lemma 2.1**

3.1. **Jacobi flow associated with the SDE (1.1).** The *Jacobi flow* is the derivative of \(X^x_t\) with respect to the initial value \(x\); the Jacobi flow in direction \(v \in \mathbb{R}^d\) is defined by
\[
\nabla_v X^x_t := \lim_{\epsilon \to 0} \frac{X^{x+\epsilon v}_t - X^x_t}{\epsilon}, \quad t \geq 0.
\]

This limit exists and satisfies
\[
\frac{d}{dt} \nabla_v X^x_t = \nabla_{\nabla_v X^x_t} b(X^x_t), \quad \nabla_v X^x_0 = v.
\]

Similarly, for \(v_1, v_2 \in \mathbb{R}^d\), we can define \(\nabla_{v_2} \nabla_{v_1} X^x_t\), which satisfies
\[
\frac{d}{dt} \nabla_{v_2} \nabla_{v_1} X^x_t = \nabla_{\nabla_{v_2} \nabla_{v_1} X^x_t} b(X^x_t) + \nabla_{v_2} \nabla_{v_1} X^x_t \nabla_{\nabla_{v_2} \nabla_{v_1} X^x_t} b(X^x_t), \quad \nabla_{v_1} \nabla_{v_1} X^x_0 = 0.
\]

Then, we first have the following estimates of \(\nabla_{v_1} X^x_t\) and \(\nabla_{v_2} \nabla_{v_1} X^x_t\).
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:

\[
\begin{align*}
|\nabla_{v_1} X_t^x| &\leq e^{\|\theta\|_2} |v_1|, & t \in (0, 1], \\
|\nabla_{v_2} \nabla_{v_1} X_t^x| &\leq \frac{\theta_3}{2\sqrt{2\theta_2}} e^{4\theta_2} |v_1| |v_2|, & t \in (0, 1].
\end{align*}
\]

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

\[
\frac{d}{dt} |\nabla_{v_1} X_t^x|^2 = 2 \langle \nabla_{v_1} X_t^x, \nabla_{v_1} X_t^x b(X_t^x) \rangle \leq 2\theta_2 |\nabla_{v_1} X_t^x|^2,
\]

and Gronwall’s inequality yields for \( t \in (0, 1] \)

\[
|\nabla_{v_1} X_t^x|^2 \leq e^{2\theta_2 t} |v_1|^2 \leq e^{2\theta_2} |v_1|^2.
\]

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 \( 2\theta_2 \leq A^2 + B^2 \) that

\[
\frac{d}{dt} |\zeta(t)|^2 = 2 \langle \zeta(t), \nabla_{v_1} X_t^x b(X_t^x) \rangle \leq 2\theta_2 |\zeta(t)|^2 + 2\theta_3 e^{4\theta_2} |v_1| |v_2| |\zeta(t)| \\
\leq 4\theta_2 |\zeta(t)|^2 + \frac{\theta_3^2}{2\theta_2} e^{4\theta_2} |v_1|^2 |v_2|^2.
\]

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

\[
|\zeta(t)|^2 \leq \frac{\theta_3^2}{2\theta_2} e^{4\theta_2} |v_1|^2 |v_2|^2 \int_0^t e^{4\theta_2 (t-s)} \, ds \leq \frac{\theta_3^2}{8\theta_2^2} e^{8\theta_2} |v_1|^2 |v_2|^2.
\]

\[ \square \]

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

\[
U = \int_0^t u(s) \, ds.
\]

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

\[
D_U \langle F_t(W), G_t(W) \rangle = \langle D_U F_t(W), G_t(W) \rangle + \langle F_t(W), D_U G_t(W) \rangle,
\]

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

\[
\mathbb{E}[D_U F_t(W)] = \mathbb{E} \left[ F_t(W) \int_0^t \langle u(s), \, dW_s \rangle \right].
\]
3.3. **Time-change method for the SDE (1.1).** (See also [50]). We will now turn to the SDE driven by a rotationally invariant $\alpha$-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 $\alpha$-stable Lévy process, see e.g. [41]. This means that we can re-write (1.1) as

$$
\text{(3.7)} \quad dX_t = b(X_t) \, dt + dW_{S_t}, \quad X_0 = x.
$$

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

$$
W_t(w) = w_t
$$

are a standard $d$-dimensional Brownian motion. Let $S$ 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 $S$ with the Skorohod metric and the probability measure $\mu_S$ so that for any $l \in S$ the coordinate process

$$
S_l(l) := l_t
$$

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

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

we define

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

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

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

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

$$
\mathbb{E}f(X^x_t) = \int_S \int_W f(X^x_t(w_l)) \mu_W(dw) \mu_S(dl).
$$

For every fixed $l \in S$, we denote by $X^l_t$ the solution to the SDE

$$
\text{(3.8)} \quad dX^l_t = b(X^l_t) \, dt + dW_{l_t}, \quad X^l_0 = x.
$$

We will now fix a path $l \in S$, and consider the SDE (3.8). Unless otherwise mentioned, all expectations are taken with respect to the Wiener space $(W, \mathcal{B}(W), \mu_W)$. 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 $\mathcal{F}^W_t$-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 $\mathcal{F}^W_t$-adapted solution $X^{x,l}_t$, see e.g. [38, 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 $\phi \in C^2_b(\mathbb{R}^d, \mathbb{R})$, all directions $v_1, v_2 \in \mathbb{R}^d$ and $x \in \mathbb{R}^d$, $t \in (0, 1]$,

$$
\text{(3.9)} \quad |\nabla_{v_1} X^{x,l}_t| \leq e^{\theta_2} |v_1|
$$

and
\[
\left| E \left[ \nabla_{\nabla_{v_2} X_t^{x_l}} \nabla_{\nabla_{v_1} X_t^{x_l}} \phi (X_t^{x_l}) \right] \right| \\
\leq \left| E \left[ \frac{1}{t} \nabla_{\nabla_{v_2} X_t^{x_l}} \phi (X_t^{x_l}) \int_0^t \left\langle \nabla_{\nabla_{v_2} X_s^{x_l}}, dW_t \right\rangle \right] \right| + \| \nabla \phi \| \frac{\theta_3}{\sqrt{2\theta_2}} |v_1| |v_2|,
\]
where \( \nabla_{v_i} X_t^{x_l} (i = 1, 2) \) is determined by the following linear equation:
\[
(3.10) \quad \frac{d}{dt} \nabla_{v_i} X_t^{x_l} = \nabla_{\nabla_{v_i} X_t^{x_l}} b (X_t^{x_l}), \quad \nabla_{v_i} X_0^{x_l} = v_i.
\]

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
\[
l_t^\epsilon := \frac{1}{\epsilon} \int_t^{t+\epsilon} l_s \, ds + \epsilon t = \int_0^1 l_{t+t-s} \, ds + \epsilon t.
\]
Since \( t \mapsto l_t \) is increasing and right continuous, it follows that for each \( t \geq 0 \),
\[
l_t^\epsilon \downarrow l_t \quad \text{as} \quad \epsilon \downarrow 0.
\]
Moreover, \( t \mapsto l_t^\epsilon \) is absolutely continuous and strictly increasing. Let \( \gamma^\epsilon \) be the inverse function of \( l^\epsilon \), i.e.
\[
gamma_{t_0}^\epsilon = t, \quad t \geq l_0^\epsilon \quad \text{and} \quad \gamma_{l_t^\epsilon}^\epsilon = t, \quad t \geq 0.
\]
By definition, \( \gamma_{t}^\epsilon \) is absolutely continuous on \( [l_0^\epsilon, \infty) \). Let \( X_t^{x_l^\epsilon} \) be the solution to the SDE
\[
dX_t^{x_l^\epsilon} = b (X_t^{x_l^\epsilon}) \, dt + dW_{t-l_0^\epsilon}, \quad X_0^{x_l^\epsilon} = x.
\]
Let us now define
\[
Y_{t}^{x_l^\epsilon} := X_{\gamma_{t}^\epsilon}^{x_l^\epsilon}, \quad t \geq l_0^\epsilon.
\]
Changing variables in (3.11) we see that for \( t \geq l_0^\epsilon \),
\[
Y_{t}^{x_l^\epsilon} = x + \int_{l_0^\epsilon}^{\gamma_{t}^\epsilon} b \left( X_{s}^{x_l^\epsilon} \right) \, ds + W_t - l_0^\epsilon = x + \int_{l_0^\epsilon}^{t} b \left( Y_{s}^{x_l^\epsilon} \right) \hat{\gamma}_{s}^\epsilon \, ds + W_t - l_0^\epsilon
\]
(\( \hat{\gamma}_{s}^\epsilon \) denotes the derivative in \( s \)). Hence, for any vector \( v \in \mathbb{R}^d \), we have
\[
\nabla_{v} Y_t^{x_l^\epsilon} = v + \int_{l_0^\epsilon}^{t} \nabla_{v} Y_{s}^{x_l^\epsilon} b \left( Y_{s}^{x_l^\epsilon} \right) \hat{\gamma}_{s}^\epsilon \, ds,
\]
and the differential form can be written as
\[
\frac{d}{dt} \nabla_{v} Y_t^{x_l^\epsilon} = \nabla b \left( Y_t^{x_l^\epsilon} \right) \hat{\gamma}_{t}^\epsilon \nabla_{v} Y_t^{x_l^\epsilon}, \quad t \geq l_0^\epsilon,
\]
which has a solution of the form
\[
\nabla_{v} Y_t^{x_l^\epsilon} = J_{l_0^\epsilon, t}^{x_l^\epsilon} v,
\]
involving a matrix exponential
\[
J_{s, t}^{x_l^\epsilon} = \exp \left[ \int_{s}^{t} \nabla b \left( Y_{s}^{x_l^\epsilon} \right) \hat{\gamma}_{s}^\epsilon \, ds \right], \quad l_0^\epsilon \leq s \leq t < \infty.
\]
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 \( l_0^\epsilon \leq s \leq t < \infty \).

Now, we come back to the Malliavin calculus from Section 3.2. Fixing \( t \geq l_0^\epsilon \) and \( x \in \mathbb{R}^d \), the solution \( Y_{t}^{x_l^\epsilon} \) is a \( d \)-dimensional functional of Brownian motion \( \{W_s\}_{l_0^\epsilon \leq s \leq t} \).
Let $U$ be as in Section 3.2. The Malliavin derivative of $Y_t^{x; l_t}$ in direction $U$ exists in $L^2 \left((W, B(W), \mu_W); \mathbb{R}^d \right)$ and is given by

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

To simplify notation, we drop the $W$ in $D_U Y_t^{x; l_t}(W)$ and write $D_U Y_t^{x; l_t} = D_U Y_t^{x; l_t}(W)$. By (3.12), it satisfies the equation

$$D_U Y_t^{x; l_t} = \int_{t_0}^t \left( \nabla_{D_U Y_s^{x; l_t}} b \left( Y_s^{x; l_t} \right) \gamma_s + u(s) \right) ds,$$

the differential form of the above equation can be written as

$$\frac{d}{dt} D_U Y_t^{x; l_t} = \nabla b \left( Y_t^{x; l_t} \right) \gamma_t D_U Y_t^{x; l_t} + u(t), \quad t \geq t_0,$$

and this equation has a unique solution which is given via the matrix exponential (3.15):

$$D_U Y_t^{x; l_t} = \int_{t_0}^t J_{s, t}^{x; l_t} u(s) ds. $$

For a fixed $t > 0$, for any $v_1, v_2, x \in \mathbb{R}^d$, we define $u_i, U_i : [t_0, t] \to \mathbb{R}^d$ by

$$(3.17) \quad u_i(s) := \frac{1}{t - t_0} \nabla_{v_i} Y_s^{x; l_t}, \quad U_i : = \int_0^s u_i(r) dr$$

for $t_0 \leq s \leq t$ and $i = 1, 2$. Then

$$D_{U_i} Y_s^{x; l_t} = \frac{s - t_0}{t - t_0} \nabla_{v_i} Y_s^{x; l_t}, \quad t_0 \leq s \leq t.$$  

In addition, (3.13) implies that for $s \in [t_0, t]$ 

$$D_{U_2} \nabla_{v_1} Y_s^{x; l_t} = \int_{t_0}^s \left( \nabla_{D_{U_2} Y_r^{x; l_t}} \nabla_{v_1} Y_r^{x; l_t} b \left( Y_r^{x; l_t} \right) + \nabla_{D_{U_2} \nabla_{v_1} Y_r^{x; l_t}} b \left( Y_r^{x; l_t} \right) \right) \gamma_r dr.$$  

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,

$$|\nabla_{v_i} Y_s^{x; l_t}| \leq e^{\theta_2 \gamma_s} |v_i| $$

and

$$|D_{U_i} \nabla_{v_i} Y_s^{x; l_t}| \leq \frac{\theta_3}{\sqrt{2\theta_2}} e^{2\theta_2 \gamma_s} \sqrt{\gamma_s} |v_1||v_2|.$$  

for $t_0 \leq s \leq t$ and $i = 1, 2$.

**Proof.** Recall that $\theta_1 > 0$ and $\gamma_s \geq 0$. By (3.13) and (1.8), we have for any $l_0 \leq s \leq t$

$$\frac{d}{ds} |\nabla_{v_i} Y_s^{x; l_t}|^2 = 2 \gamma_s \left( \nabla_{v_i} Y_s^{x; l_t}, \nabla \nabla_{v_1} Y_s^{x; l_t} b \left( Y_s^{x; l_t} \right) \right) \leq 2\theta_2 \gamma_s |\nabla_{v_i} Y_s^{x; l_t}|^2,$$

and this implies, because of Gronwall’s lemma,

$$|\nabla_{v_i} Y_s^{x; l_t}|^2 \leq \exp \left[ 2\theta_2 \int_{l_0}^s \gamma_r dr \right] |v_i|^2 = e^{2\theta_2 (\gamma_s - \gamma_{l_0})} |v_i|^2 = e^{2\theta_2 \gamma_s} |v_i|^2.$$  

17
Using (3.19) and (1.8) we find for any \( l_0 \leq s \leq t \)
\[
\frac{d}{ds} |D_{U_s} \nabla_{v_s} Y_{s}^{x;i}|^2 = 2 \frac{d}{ds} \left< D_{U_s} \nabla_{v_s} Y_{s}^{x;i}, \nabla_{D_{U_s} \nabla_{v_s} Y_{s}^{x;i}} b (Y_{s}^{x;i}) \right> + 2 \frac{d}{ds} \left< D_{U_s} \nabla_{v_s} Y_{s}^{x;i}, \nabla_{D_{U_s} \nabla_{v_s} Y_{s}^{x;i}} b (Y_{s}^{x;i}) \right> \\
\leq 2 \theta_2 \frac{d}{ds} |D_{U_s} \nabla_{v_s} Y_{s}^{x;i}|^2 + 2 \theta_3 \frac{d}{ds} |D_{U_s} \nabla_{v_s} Y_{s}^{x;i}| |D_{U_s} Y_{s}^{x;i}| |\nabla_{v_s} Y_{s}^{x;i}|.
\]

Now (3.18) and (3.20) imply
\[
\frac{d}{ds} |D_{U_s} \nabla_{v_s} Y_{s}^{x;i}|^2 \leq 2 \theta_2 \frac{d}{ds} |D_{U_s} \nabla_{v_s} Y_{s}^{x;i}|^2 + 2 \theta_3 \frac{d}{ds} |D_{U_s} \nabla_{v_s} Y_{s}^{x;i}| |\nabla_{v_s} Y_{s}^{x;i}|^2 \\
\leq 4 \theta_2 \frac{d}{ds} |D_{U_s} \nabla_{v_s} Y_{s}^{x;i}|^2 + \frac{\theta_2^2}{2 \theta_2} \frac{d}{ds} |\nabla_{v_s} Y_{s}^{x;i}|^2 |\nabla_{v_s} Y_{s}^{x;i}|^2 \\
\leq 4 \theta_2 \frac{d}{ds} |D_{U_s} \nabla_{v_s} Y_{s}^{x;i}|^2 + \frac{\theta_2^2}{2 \theta_2} \frac{d}{ds} e^{4 \theta_2 \gamma_s} |v_1|^2 |v_2|^2.
\]

Since \( D_{U_s} \nabla_{v_s} Y_{s}^{x;i} = 0 \), we can use Gronwall’s inequality to see
\[
|D_{U_t} \nabla_{v_t} Y_{t}^{x;i}|^2 \leq \frac{\theta_2^2}{2 \theta_2} |v_1|^2 |v_2|^2 \int_{l_0}^{t} \frac{e^{4 \theta_2 \gamma_s} e^{4 \theta_2 (\gamma_s - \delta_s)}}{e^{4 \theta_2 \gamma_s}} du = \frac{\theta_2^2}{2 \theta_2} |v_1|^2 |v_2|^2 e^{4 \theta_2 \gamma_s} \gamma_s^2.
\]

**Proof of Lemma 3.2.** From (1.8) we see that
\[
\frac{d}{dt} |\nabla_{v_t} X_{t}^{x;i}|^2 = 2 \left< \nabla_{v_t} X_{t}^{x;i}, \nabla_{\nabla_{v_t} X_{t}^{x;i}} b (X_{t}^{x;i}) \right> \leq 2 \theta_2 |\nabla_{v_t} X_{t}^{x;i}|^2.
\]
This yields for all \( t \in (0, 1) \)
\[
|\nabla_{v_t} X_{t}^{x;i}|^2 \leq e^{2 \theta_2 t} |v_1|^2 \leq e^{2 \theta_2} |v_1|^2,
\]
i.e. (3.9) holds.

Using (3.18) with \( s = t \) and \( i = 2 \), the chain rule and Bismut’s formula (3.6), we see that
\[
E \left[ \nabla_{\nabla_{v_t} X_{t}^{x;i}} \nabla_{\nabla_{v_t} Y_{t}^{x;i}} \phi (Y_{t}^{x;i}) \right] \\
= E \left[ \nabla_{D_{U_t} Y_{t}^{x;i}} \nabla_{\nabla_{v_t} Y_{t}^{x;i}} \phi (Y_{t}^{x;i}) \right] \\
= E \left[ D_{U_t} \left( \nabla_{\nabla_{v_t} Y_{t}^{x;i}} \phi (Y_{t}^{x;i}) \right) \right] - E \left[ \nabla_{D_{U_t} \nabla_{v_t} Y_{t}^{x;i}} \phi (Y_{t}^{x;i}) \right] \\
= \frac{1}{t - l_0} E \left[ \nabla_{\nabla_{v_t} Y_{t}^{x;i}} \phi (Y_{t}^{x;i}) \int_{l_0}^{t} \left< \nabla_{v_t} Y_{s}^{x;i}, dW_s \right> \right] - E \left[ \nabla_{D_{U_t} \nabla_{v_t} Y_{t}^{x;i}} \phi (Y_{t}^{x;i}) \right].
\]

We can now use the fact that for each \( t \geq 0 \),
\[
Y_{t_t}^{x;i} = X_{t_t}^{x;i} \quad \text{and} \quad \nabla_{v_t} Y_{t_t}^{x;i} = \nabla_{v_t} X_{t_t}^{x;i}.
\]
Replacing \( t \) with \( l_t \) in (3.17), this yields
\[
E \left[ \nabla_{\nabla_{v_t} X_{t}^{x;i}} \nabla_{\nabla_{v_t} Y_{t}^{x;i}} \phi (X_{t}^{x;i}) \right] \\
= \frac{1}{l_t - l_0} E \left[ \nabla_{\nabla_{v_t} X_{t}^{x;i}} \phi (X_{t}^{x;i}) \int_{l_0}^{l_t} \left< \nabla_{v_t} Y_{s}^{x;i}, dW_s \right> \right] - E \left[ \nabla_{D_{U_t} \nabla_{v_t} Y_{t}^{x;i}} \phi (X_{t}^{x;i}) \right] \\
= \frac{1}{l_t - l_0} E \left[ \nabla_{\nabla_{v_t} X_{t}^{x;i}} \phi (X_{t}^{x;i}) \int_{0}^{t} \left< \nabla_{v_t} X_{s}^{x;i}, dW_s \right> \right] - E \left[ \nabla_{D_{U_t} \nabla_{v_t} Y_{t}^{x;i}} \phi (X_{t}^{x;i}) \right].
\]
Since $\gamma_1^{t} = t$, (3.21) implies that for every $t \in (0, 1]$
\[
\begin{align*}
\left| E \left[ \nabla_{D_{t}^{2}} \nabla_{v_{1}^{t}} \phi \left( X_{t}^{x,t} \right) \right] \right| &= \left| E \left[ \langle \nabla \phi \left( X_{t}^{x,t} \right), D_{t}^{2} \nabla_{v_{1}^{t}} \phi \left( X_{t}^{x,t} \right) \rangle \right] \right| \\
&\leq \| \nabla \phi \|_{\infty} \frac{\theta_{3}}{\sqrt{2} \theta_{2}} e^{2 \theta_{2}^{\gamma_{1}^{t}} t} |v_{1}| |v_{2}| \\
&= \| \nabla \phi \|_{\infty} \frac{\theta_{3}}{\sqrt{2} \theta_{2}} e^{2 \theta_{2}^{t}} |v_{1}| |v_{2}| \\
&\leq \| \nabla \phi \|_{\infty} \frac{\theta_{3}}{\sqrt{2} \theta_{2}} e^{2 \theta_{2}^{t}} |v_{1}| |v_{2}| ,
\end{align*}
\]
and so
(3.22)
\[
\begin{align*}
\left| E \left[ \nabla_{\nabla_{v_{2}^{t}}^{x,t}} \nabla_{v_{1}^{t}} \phi \left( X_{t}^{x,t} \right) \right] \right| \\
&\leq \left| \frac{1}{l_{t} - l_{0}} E \left[ \nabla_{\nabla_{v_{1}^{t}}^{x,t}} \phi \left( X_{t}^{x,t} \right) \int_{0}^{t} \langle \nabla_{v_{2}^{t}}^{x,s} X_{s}^{x,t}, dW_{t_{s}} \rangle \right] \right| + \| \nabla \phi \|_{\infty} \frac{\theta_{3}}{\sqrt{2} \theta_{2}} e^{2 \theta_{2}^{t}} |v_{1}| |v_{2}| .
\end{align*}
\]
By the same argument as in the proof of [50, Lemma 2.5], we obtain
(3.23)
\[
\begin{align*}
\lim_{\epsilon \to 0} \frac{1}{l_{t} - l_{0}} E \left[ \nabla_{\nabla_{v_{1}^{t}}^{x,t}} \phi \left( X_{t}^{x,t} \right) \int_{0}^{t} \langle \nabla_{v_{2}^{t}}^{x,s} X_{s}^{x,t}, dW_{t_{s}} \rangle \right] \\
= \frac{1}{l_{t}} E \left[ \nabla_{\nabla_{v_{1}^{t}}^{x,t}} \phi \left( X_{t}^{x,t} \right) \int_{0}^{t} \langle \nabla_{v_{2}^{t}}^{x,s} X_{s}^{x,t}, dW_{t_{s}} \rangle \right] .
\end{align*}
\]
On the other hand, from [50, Lemma 2.2], we know that
(3.24)
\[
\begin{align*}
\lim_{\epsilon \to 0} E \left[ \nabla_{\nabla_{v_{2}^{t}}^{x,t}} \nabla_{v_{1}^{t}} \phi \left( X_{t}^{x,t} \right) \right] = E \left[ \nabla_{\nabla_{v_{2}^{t}}^{x,t}} \nabla_{v_{1}^{t}} \phi \left( X_{t}^{x,t} \right) \right] .
\end{align*}
\]
Letting in (3.22) $\epsilon \to 0$ and using (3.23) and (3.24), completes the proof. \hfill \square

3.4. **Proof of Lemma 2.1.** Because of (3.3), we can use the differentiation theorem for parameter dependent integrals to get
\[
|\nabla_{\epsilon} P_{h}(x)| = |\nabla_{\epsilon} E[h(X_{t}^{x})]| = |E \left[ \nabla_{\epsilon} X_{t}^{x} h(X_{t}^{x}) \right] | \\
= |E \left[ \langle \nabla_{\epsilon} h(X_{t}^{x}), \nabla_{\epsilon} X_{t}^{x} \rangle \right] | \leq \| \nabla h \| e^{\theta_{2}} |v| \leq e^{\theta_{2}} |v| .
\]
In order to see the second inequality, we define for every $\epsilon > 0$
(3.25)
\[
\begin{align*}
\epsilon_{x}(x) := \int_{\mathbb{R}^{d}} g_{\epsilon}(y) h(x - y) \, dy ,
\end{align*}
\]
where $g_{\epsilon}$ is the density of the normal distribution $N(0, \epsilon^{2} I_{d})$. It is easy to see that $h_{\epsilon}$ is smooth, $\lim_{\epsilon \to 0} h_{\epsilon}(x) = h(x)$, $\lim_{\epsilon \to 0} \nabla h_{\epsilon}(x) = \nabla h(x)$ and $|h_{\epsilon}(x)| \leq C(1 + |x|)$ for all $x \in \mathbb{R}^{d}$ and some $C > 0$. Moreover, $\| \nabla h_{\epsilon} \| \leq \| \nabla h \| \leq 1$. Using the differentiability theorem for parameter dependent integrals we get
\[
\begin{align*}
\nabla_{\epsilon} X_{t}^{x} h_{\epsilon}(X_{t}^{x}) = E \left[ \nabla_{\epsilon} X_{t}^{x} h_{\epsilon}(X_{t}^{x}) \right] + E \left[ \nabla_{\epsilon} X_{t}^{x} \nabla_{\epsilon} X_{t}^{x} h_{\epsilon}(X_{t}^{x}) \right] .
\end{align*}
\]
From (3.4) we get
\[
\begin{align*}
\left| E \left[ \nabla_{\epsilon} X_{t}^{x} h_{\epsilon}(X_{t}^{x}) \right] \right| = \left| E \left[ \langle \nabla_{\epsilon} h_{\epsilon}(X_{t}^{x}), \nabla_{\epsilon} X_{t}^{x} \rangle \right] \right| \leq \frac{\theta_{3}}{2 \sqrt{2 \theta_{2}}} e^{\theta_{2}} |v_{1}| |v_{2}| .
\end{align*}
\]
It follows from Lemma 3.2, [50, (3.3)] and (3.3) that
\[
\E[\nabla v_2 X_t \nabla v_1 X_t h_e(X_t^x)] \\
\leq \left[ \E \left[ \frac{1}{S_t} \nabla v_2 X_t \nabla v_1 X_t h_e(X_t^x) \int_0^t \langle \nabla v_2 X_s^x, dW_s \rangle \right] \right] + \|\nabla h_e\|_\infty \frac{\theta_3}{\sqrt{2\theta_2}} e^{\theta_2} |v_1| |v_2| \\
\leq e^{\theta_2} |v_1| \E \left[ \frac{1}{S_t} \int_0^t \langle \nabla v_2 X_s^x, dW_s \rangle \right] + \frac{\theta_3}{\sqrt{2\theta_2}} e^{\theta_2} |v_1| |v_2|.
\]

The Cauchy–Schwarz inequality, Itô’s isometry and (3.9) give
\[
\E \left[ \frac{1}{S_t} \int_0^t \langle \nabla v_2 X_s^x, dW_s \rangle \right] = \left( \int_0^t \E \left[ \frac{1}{S_t} |\nabla v_2 X_s^x| \right] \right)^{1/2} \mu_S(\text{d}l) \\
\leq \left( \int_0^t \E \left[ \frac{1}{S_t} |\nabla v_2 X_s^x|^2 \right] \right)^{1/2} \mu_S(\text{d}l) \\
\leq e^{\theta_2} |v_2| \int_0^t \frac{1}{\sqrt{l}} \mu_S(\text{d}l) \\
= e^{\theta_2} |v_2| \E \left[ S_t^{-1/2} \right] \\
\leq C e^{\theta_2} |v_2| t^{-1/\alpha},
\]
where the last inequality is taken from [10, Theorem 2.1 (ii) (c)]. Thus, we have for all \( t \in (0, 1], \)
\[
|\nabla v_2 \nabla v_1 \E[h_e(X_t^x)]| \leq |\E \left[ \nabla v_2 \nabla v_1 X_t^x h_e(X_t^x) \right]| + |\E \left[ \nabla v_2 X_t^x \nabla v_1 X_t^x h_e(X_t^x) \right]| \\
\leq \frac{\theta_3}{2\sqrt{2\theta_2}} e^{\theta_2} |v_1| |v_2| + e^{\theta_2} |v_1| C e^{\theta_2} |v_2| t^{-1/\alpha} + \frac{\theta_3}{\sqrt{2\theta_2}} e^{\theta_2} |v_1| |v_2| \\
\leq e^{\theta_2} \left( \frac{\theta_3}{2\sqrt{2\theta_2}} + C t^{-1/\alpha} \right) |v_1| |v_2| \\
\leq C e^{\theta_2} t^{-1/\alpha} |v_2|.
\]

Finally, we can let \( \epsilon \to 0 \) using dominated convergence,
\[
\lim_{\epsilon \to 0} \nabla v_2 \nabla v_1 \E[h_e(X_t^x)] = \nabla v_2 \nabla v_1 \E[h(X_t^x)],
\]
completing the proof of Lemma 2.1.

\[\square\]

**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
\[
\mathcal{L}^\alpha f(x) = \langle b(x), \nabla f(x) \rangle + (-\Delta)^{\alpha/2} f(x), \quad f \in C^2(\mathbb{R}^d, \mathbb{R})
\]
where \((-\Delta)^{\alpha/2}\) is the fractional Laplace operator, which is the generator of the rotationally symmetric \( \alpha \)-stable Lévy process \( Z_t \); it is defined as a principal value integral
\[
(-\Delta)^{\alpha/2} f(x) = C_{d,\alpha} \cdot \text{p.v.} \int_{\mathbb{R}^d} (f(x + y) - f(x)) \frac{dy}{|y|^{d+\alpha}}.
\]

It is not difficult to see that for all functions from the set
\[
\mathcal{D} := \left\{ f \in C^2(\mathbb{R}^d, \mathbb{R}) ; \int_{|z| \geq 1} |f(x + z) - f(x)| \frac{dz}{|z|^{\alpha+d}} < \infty \right\},
\]

\( \mathcal{L}^\alpha f \) is well-defined; moreover \( \mathcal{D} \times \mathcal{L}^\alpha(\mathcal{D}) \) can be embedded into the full generator \( \mathcal{L}^\alpha \), i.e. the set of all pairs of (bounded) Borel functions \((f, g)\) such that \( f(X_t) - f(X_0) - \int_0^t g(X_s) \, ds \) 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

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

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

\[
|\nabla V_\beta(x)| \leq \beta|x|^{\beta-1}, \quad \|\nabla^2 V_\beta(x)\|_{\text{HS}} \leq \beta(3 - \beta)\sqrt{d}.
\]

Thus, (1.7) and Young’s inequality \((AB \leq \frac{1}{p}A^p + \frac{1}{q}B^q \text{ with } p = \beta \text{ and } q = \beta/(\beta - 1))\) imply

\[
\langle b(x), \nabla V_\beta(x) \rangle = \frac{\beta}{(1 + |x|^2)^{\frac{\beta}{2}}} \langle b(x) - b(0), x \rangle + \frac{\beta}{(1 + |x|^2)^{\frac{\beta}{2}}} \langle b(0), x \rangle
\]

\[
\leq -\theta_1 \frac{\beta|x|^2}{(1 + |x|^2)^{\frac{\beta}{2}}} + \frac{\beta K}{(1 + |x|^2)^{\frac{\beta}{2}}} + \frac{\beta |b(0)||x|}{(1 + |x|^2)^{\frac{\beta}{2}}}
\]

\[
\leq -\theta_1 V_\beta(x) + \theta_1 \beta + \beta K + \beta |b(0)||x|^{\beta-1}
\]

\[
\leq -\theta_1 V_\beta(x) + \theta_1 \beta + \beta K + \theta_1^{1-\beta} |b(0)|^\beta.
\]

Therefore, we see from (A.1) that

(A.2)

\[
(-\Delta)^{\alpha/2} V_\beta(x) = C_{d, \alpha} \int (V_\beta(x + y) - V_\beta(x) - \langle \nabla V_\beta(x), y \rangle \mathbb{1}_{(0,1)}(|y|)) \frac{dy}{|y|^{\alpha+d}}
\]

\[
= C_{d, \alpha} \int_{|y|<1} \int_0^1 \int_0^r \langle \nabla^2 V_\beta(x + sy), yy^T \rangle_{\text{HS}} \, ds \, dy \frac{dr}{|y|^{\alpha+d}}
\]

\[
+ C_{d, \alpha} \int_{|y|\geq1} \int_0^1 \langle \nabla V_\beta(x + ry), y \rangle \, dr \frac{dy}{|y|^{\alpha+d}}
\]

\[
\leq C_{d, \alpha} \frac{\beta(3 - \beta)}{2} \sqrt{d} \int_{|y|<1} \frac{|y|^2}{|y|^{\alpha+d}} \, dy + C_{d, \alpha} \beta \int_{|y|\geq1} \frac{|x|^{\beta-1} dy}{|y|^{\alpha+d}} + \frac{C_{d, \alpha} \beta}{2(2 - \alpha)} + C_{d, \alpha} \beta \sigma_{d-1} \left( \frac{|x|^{\beta-1}}{\alpha - 1} + \frac{1}{\alpha - \beta} \right).
\]

Again by Young’s inequality we get

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

Hence, we have

(A.3)

\[
\mathcal{L}^\alpha V_\beta(x) \leq -\lambda_1 V_\beta(x) + q_1 \mathbb{1}_{A_1}(x),
\]

with \( \lambda_1 = \frac{1}{\alpha} \theta_1 \),

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

and the compact set \( A_1 = \left\{ x \in \mathbb{R}^d : |x| \leq (4\theta_1^{-1} q_1)^{1/\beta} \right\} \).
Thus, [27, Theorem 5.1] yields that the process \( (X_t^x)_{t \geq 0} \) is ergodic, i.e. there exists a unique invariant probability measure \( \mu \) such that for all \( x \in \mathbb{R}^d \) and \( t > 0 \),

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

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

\[
\sup_{|f| \leq V_0} \left| \mathbb{E}[f(X_t^x)] - \mu(f) \right| \leq c_1 V_0(x) e^{-c_2 t}
\]

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

\[
\langle b(x) - b(y), x - y \rangle \leq \theta_2 |x - y|^2,
\]

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

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

Hence, we can use [48, Theorem 1.2] with \( K_1 = \theta_2, K_2 = \theta_2/\theta_1 \) and \( L_0 = 2K/\theta_1 \) 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 \( P(x, dy) = \mathbb{P}(Y_1 \in dy \mid Y_0 = x) \). Since \( V_1(y) \leq |y| + 1 \) and

\[
Y_1 = x + \eta b(x) + Z_\eta,
\]

we have

\[
\int_{\mathbb{R}^d} V_1(y) P(x, dy) \leq \int_{\mathbb{R}^d} (|y| + 1)p(\eta, y - x - \eta b(x)) \, dy
= \int_{\mathbb{R}^d} (|z + x + \eta b(x)| + 1) p(\eta, z) \, dz
\leq \int_{\mathbb{R}^d} (|z| + |x + \eta b(x)| + 1) p(\eta, z) \, dz
\leq \mathbb{E}[Z_\eta] + |x + \eta b(x) - b(0)| + \eta|b(0)| + 1.
\]

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

\[
|x + \eta b(x) - b(0)|^2 = |x|^2 + 2\eta \langle b(x) - b(0), x \rangle + \eta^2 |b(x) - b(0)|^2
\leq \left( 1 - 2\theta_1 \eta + \theta_2^2 \eta^2 \right) |x|^2 + 2K \eta,
\]

which implies

\[
\int_{\mathbb{R}^d} V_1(y) P(x, dy) \leq \eta^2 \mathbb{E}[Z_1] + (1 - 2\theta_1 \eta + \theta_2^2 \eta^2)^{1/2} |x| + \sqrt{2K \eta} + \eta|b(0)| + 1
\leq (1 - \theta_1 \eta) |x| + \eta^2 \mathbb{E}[Z_1] + \sqrt{2K \eta} + \eta|b(0)| + 1,
\]

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

\[
\int_{\mathbb{R}^d} V_1(y) P(x, dy) \leq \lambda_2 V_1(x) + q_2 1_{A_2}(x)
\]

\[\text{22}\]
with
\[ \lambda_2 = 1 - \frac{\theta_2}{2} \eta < 1, \quad q_2 = 1 + \frac{\theta_2}{2} \eta + \eta^{\frac{1}{\theta_2}} E|Z_1| + \sqrt{2K} \eta + \eta |b(0)|, \]
and the compact set \( A_2 = \left\{ x \in \mathbb{R}^d : |x| \leq 1 + \frac{2}{\theta_1} E|Z_1| \eta^{\frac{1}{\theta_1}} + \frac{2 \sqrt{2K}}{\theta_1} \eta^{-\frac{1}{\theta_1}} + \frac{2 |b(0)|}{\theta_1} \right\} \).

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 \( (Y_k^x)_{k \geq 0} \) is exponentially ergodic, i.e. there exists a unique invariant probability \( \mu_\eta \) such that for all \( x \in \mathbb{R}^d \) and \( t > 0 \),
\[ \sup_{|f| \leq V_1} |E f(Y_k) - \mu_\eta(f)| \leq c_1 V_1(x) e^{-c_2 k}, \]
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 \( |y|^\beta \leq V_\beta(y) \) and
\[
V_\beta(Y_{k+1}) = V_\beta(Y_k + \eta b(Y_k) + Z_\eta) \\
= V_\beta(Y_k + \eta b(Y_k)) + V_\beta(Y_k + \eta b(Y_k) + Z_\eta) - V_\beta(Y_k + \eta b(Y_k)) \\
= V_\beta(Y_k) + \int_0^\eta \langle \nabla V_\beta(Y_k + rb(Y_k)), b(Y_k) \rangle \, dr \\
+ \int_0^\eta \langle \nabla V_\beta(Y_k + \eta b(Y_k) + Z_\eta) - V_\beta(Y_k + \eta b(Y_k)), b(Y_k) \rangle \, dr,
\]
where \( Z_\eta \) is independent of \( Y_k \). Since \( \nabla V_\beta(x) = \beta x (1 + |x|^2)^{-\frac{\beta}{2-\beta}} \), (1.7) implies that
\[
\int_0^\eta \langle \nabla V_\beta(Y_k + rb(Y_k)), b(Y_k) \rangle \, dr \\
\leq \int_0^\eta \frac{\beta (Y_k, b(Y_k)) + \beta r |b(Y_k)|^2}{(1 + |Y_k + rb(Y_k)|^2)^{\frac{\beta}{2-\beta}}} \, dr \\
\leq \frac{\int_0^\eta (Y_k, b(Y_k) - b(0)) + \beta |b(0)||Y_k| + \beta r |b(Y_k)|^2}{(1 + |Y_k + rb(Y_k)|^2)^{\frac{\beta}{2-\beta}}} \, dr \\
\leq \frac{\int_0^\eta - \theta_1 \beta |Y_k|^2 + \beta K + \beta |b(0)||Y_k| + \beta r |b(Y_k)|^2}{(1 + |Y_k + rb(Y_k)|^2)^{\frac{\beta}{2-\beta}}} \, dr.
\]

One can write by (1.9) and the fact \( r \leq \eta \leq \min \left\{ 1, \frac{\theta_1}{8\theta_2}, \frac{1}{\theta_1} \right\} \)
\[
- \theta_1 \beta |Y_k|^2 + \beta K + \beta |b(0)||Y_k| + \beta r |b(Y_k)|^2 \\
\leq - \frac{3}{4} \theta_1 \beta |Y_k|^2 + 2 \eta \theta_2 \beta |Y_k|^2 + \frac{\beta |b(0)|^2}{\theta_1} + 2 \beta r |b(0)|^2 + \beta K \\
\leq - \frac{\theta_1 \beta}{2} |Y_k|^2 + \frac{\beta |b(0)|^2}{\theta_1} + 2 \beta r |b(0)|^2 + \beta K,
\]
\[ 23 \]
whereas
\[ |Y_k + rb(Y_k)|^2 + 1 \]
\[ = |Y_k|^2 + 2r(Y_k, b(Y_k)) + r^2b(Y_k)^2 + 1 \]
\[ \leq |Y_k|^2 + 2r(Y_k, b(Y_k) - b(0)) + 2r|b(0)||Y_k| + 2r^2\theta_2^2|Y_k|^2 + 2r^2|b(0)|^2 + 1 \]
\[ \leq (1 - 2r\theta_1)|Y_k|^2 + 2r|b(0)||Y_k| + 2r^2\theta_2^2|Y_k|^2 + 2r^2|b(0)|^2 + 1 + 2\eta K \]
\[ \leq |Y_k|^2 + \frac{2r|b(0)|^2}{\theta_1} + 2\eta^2|b(0)|^2 + 1 + 2\eta K. \]

Noting that \((1 + |Y_k + rb(Y_k)|^2)^{\frac{2-\alpha}{2}} \geq 1\), we get
\[
-\theta_1 \beta |Y_k|^2 + \beta K + \beta|b(0)||Y_k| + \beta r|b(Y_k)|^2 \]
\[
\leq \frac{-\theta_1 \beta}{2} \left( |Y_k|^2 \frac{2r|b(0)|^2}{\theta_1} + 2\eta^2|b(0)|^2 + 1 + 2\eta K \right)^{\frac{2-\alpha}{2}} + \frac{\beta|b(0)|^2}{\theta_1} + 2\beta r|b(0)|^2 + \beta K \]
\[
\leq \frac{-\theta_1 \beta}{2} \left( |Y_k|^2 \frac{2r|b(0)|^2}{\theta_1} + 2\eta^2|b(0)|^2 + 1 + 2\eta K \right)^{\frac{2-\alpha}{2}} + \frac{\beta|b(0)|^2}{\theta_1} + 2\beta r|b(0)|^2 + \beta K. \]

Combining this with (A.4), we arrive at
\[
\int_0^\eta \langle \nabla V_\beta (Y_k + rb(Y_k)), b(Y_k) \rangle \, dr \leq -\frac{\theta_1 \beta}{2} \eta V_\beta(Y_k) + C_1 \eta. \]

In addition, for any \( y \in \mathbb{R}^d\), Itô’s formula and the inequality (A.2) imply that
\[
|E[V_\beta(y + \eta b(y) + Z_\eta) - V_\beta(y + \eta b(y))]| \]
\[
= \left| \int_0^\eta E[(-\Delta)^{\alpha/2} V_\beta(y + \eta b(y) + Z_r)] \, dr \right| \]
\[
\leq \int_0^\eta \left[ C_{d,\alpha}\beta(3-\beta)\sqrt{\sigma_{d-1}} + C_{d,\alpha}\beta\sigma_{d-1} \left( \frac{E|y + \eta b(y) + Z_r|^{\beta-1}}{\alpha - 1} + \frac{1}{\alpha - \beta} \right) \right] \, dr \]
\[
\leq C_{d,\alpha}\beta\sigma_{d-1} \int_0^\eta \left[ \frac{(3-\beta)\sqrt{d}}{2(2-\alpha)} + \frac{|y|^{\beta-1} + \eta^{\beta-1}|b(y)|^{\beta-1} + E|Z_r|^{\beta-1}}{\alpha - 1} + \frac{1}{\alpha - \beta} \right] \, dr. \]

This yields, in turn,
\[
|E[V_\beta(y + \eta b(y) + Z_\eta) - V_\beta(y + \eta b(y))]| \]
\[
\leq C_{d,\alpha}\beta\sigma_{d-1} \left( \frac{(3-\beta)\sqrt{d}\eta}{2(2-\alpha)} + \frac{\eta}{\alpha - \beta} + \frac{1 + \theta_2^{\beta-1}}{\alpha - 1} \eta|y|^{\beta-1} + \eta|b(0)|^{\beta-1} + \int_0^\eta \frac{E|Z_r|^{\beta-1}}{\alpha - 1} \, dr \right) \]
\[
\leq C_{d,\alpha}\beta\sigma_{d-1} \left( \frac{(3-\beta)\sqrt{d}\eta}{2(2-\alpha)} + \frac{\eta}{\alpha - \beta} + \frac{1 + \theta_2^{\beta-1}}{\alpha - 1} \eta|y|^{\beta-1} + \eta|b(0)|^{\beta-1} + \frac{E|Z_1|^{\beta-1}}{\alpha - 1} \eta \right), \]

24
where the first inequality uses (1.9) and the fact that $0 < \eta < 1$. Since $Z_\eta$ is independent of $Y_k$, we can derive that
\[
|V_\beta (Y_k + \eta b(Y_k) + Z_\eta) - V_\beta (Y_k + \eta b(Y_k))| \\
\leq C_{d,\alpha \beta \sigma_{d-1}} \left( \frac{(3 - \beta) \sqrt{d \eta}}{2(2 - \alpha)} + \frac{\eta}{\alpha - \beta} + \frac{1 + \theta_2^{\beta-1}}{\alpha - 1} \eta E|Y_k|^\beta - 1 + \frac{E|Z_1|^\beta - 1}{\alpha - 1} \right) \\
\leq \frac{\theta_1 (\beta - 1)}{2} \eta V_\beta (Y_k) + C_2 \eta,
\]
where the last inequality follows from Young’s inequality and
\[
C_2 = C_{d,\alpha \beta \sigma_{d-1}} \left( \frac{(3 - \beta) \sqrt{d}}{2(2 - \alpha)} + \frac{1}{\alpha - \beta} + |b(0)|^{\beta - 1} + \frac{E|Z_1|^\beta - 1}{\alpha - 1} \right) \\
+ \left( \frac{C_{d,\alpha \sigma_{d-1}}(1 + \theta_2^{\beta-1})}{\alpha - 1} \right)^\beta \left( \frac{2}{\theta_1} \right)^{\beta - 1}.
\]
Therefore,
\[
E[V_\beta (Y_{k+1})] \leq \left( 1 - \frac{\theta_1}{2} \eta \right) E[V_\beta (Y_k)] + (C_1 + C_2) \eta,
\]
which we can iterate this to get
\[
E[V_\beta (Y_{k+1})] \leq \left( 1 - \frac{\theta_1}{2} \eta \right)^{k+1} V_\beta (x) + (C_1 + C_2) \eta \sum_{j=0}^{k} \left( 1 - \frac{\theta_1}{2} \eta \right)^j \\
\leq V_\beta (x) + \frac{2(C_1 + C_2)}{\theta_1}.
\]
Using that $V_\beta (x) \leq 1 + |x|^\beta$, we finally get
\[
E|Y_\beta^x|^{\beta} \leq E[V_\beta (Y_\beta^x)] \leq C(1 + |x|^\beta),
\]
for some constant $C$ which is independent of $\eta$. □

**APPENDIX B. EXACT RATE FOR THE ORNSTEIN–UHLENBECK PROCESS**

In this section, we assume that $\mu$ is the invariant measure of the Ornstein–Uhlenbeck process on $\mathbb{R}$:
\[
(B.1) \quad dX_t = -X_t \, dt + dZ_t, \quad X_0 = x,
\]
where $Z_t$ is a rotationally symmetric $\alpha$-stable Lévy process ($1 < \alpha < 2$), and $\tilde{\mu}_\eta$ is the invariant measure of
\[
\tilde{Y}_{k+1} = \tilde{Y}_k - \eta \tilde{Y}_k + \frac{\eta^{1/\alpha}}{\sigma} \tilde{Z}_{k+1}, \quad k = 0, 1, 2, \ldots,
\]
where $\eta \in (0, 1)$, $\tilde{Y}_0 = x$, $\sigma = (\frac{\alpha}{2 \alpha})^{1/\alpha}$ with $d_\alpha = C_{1,\alpha} = \left( 2 \int_0^\infty \frac{1 - \cos y}{y^{\alpha+1}} \, dy \right)^{-1}$, and $\tilde{Z}_j$ are i.i.d. random variables with density
\[
(B.2) \quad p(z) = \frac{\alpha}{2 |z|^{\alpha+1}} \mathbb{I}_{(1,\infty)}(|z|).
\]

**Proposition B.1.** For every $x \in \mathbb{R}$ and $\alpha \in (1, 2)$,
\[
0 < \liminf_{\eta \downarrow 0} \frac{W_1(\mu, \tilde{\mu}_\eta)}{\eta^{2/\alpha - 1}} \leq \limsup_{\eta \downarrow 0} \frac{W_1(\mu, \tilde{\mu}_\eta)}{\eta^{2/\alpha - 1}} < \infty.
\]
Proof. Since \( X_t = e^{-t}x + e^{-t} \int_0^t e^s \, dZ_s \), we get

\[
E \left[ e^{i \xi X_t} \right] = e^{i \xi e^{-t}x} E \left[ e^{i \int_0^t e^{-s} \, ds} \right] = e^{i \xi e^{-t}e^{-t}x \alpha \, ds} = e^{i \xi e^{-t}e^{-t}x \alpha \, ds} \xrightarrow{t \to \infty} e^{-\alpha^{-1} |\xi|^\alpha} = E \left[ e^{i \xi e^{-1/\alpha} Z_1} \right].
\]

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

It is easy to see that

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

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

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

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

(B.3)

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

For \( \xi > 0 \), we have

\[
1 - \varphi(\xi) = 2 \int_1^\infty \left[ 1 - \cos(\xi z) \right] p(z) \, dz = \alpha \int_1^\infty \left[ 1 - \cos(\xi z) \right] \frac{dz}{z^{(\alpha+1)}}
\]

\[
= \alpha \xi^{\alpha} \int_0^\infty \left[ 1 - \cos u \right] \frac{du}{u^{(\alpha+1)}} - \alpha \xi^{\alpha} \frac{1}{\alpha}. \int_0^\xi \left[ 1 - \cos u \right] \frac{du}{u^{(\alpha+1)}}
\]

\[
= \sigma^\alpha \xi^{\alpha} - \alpha \xi^{\alpha} \frac{1}{\alpha}. \int_0^\xi \left[ 1 - \cos u \right] \frac{du}{u^{(\alpha+1)}}.
\]

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

\[
\varphi(\xi) = 1 - \sigma^\alpha |\xi|^\alpha + \alpha |\xi|^\alpha \int_{0}^{1} \frac{du}{u^{(\alpha+1)}}.
\]

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

\[
\varphi(\xi) \geq 1 - \sigma^\alpha |\xi|^\alpha + \alpha |\xi|^\alpha \int_{0}^{1} \frac{du}{u^{(\alpha+1)}} = 1 - \sigma^\alpha |\xi|^\alpha + \frac{c\alpha}{2 - \alpha} |\xi|^2.
\]

Thus, for \( |\xi| \leq 1 \) and \( 0 < \eta < 1 \wedge \sigma^\alpha \),

\[
\log \varphi \left( \frac{\eta^{1/\alpha}}{\sigma} (1 - \eta)^i \xi \right) \geq \log \left( 1 - \sigma^\alpha \left| \frac{\eta^{1/\alpha}}{\sigma} (1 - \eta)^i \xi \right|^\alpha + \frac{c\alpha}{2 - \alpha} \left| \frac{\eta^{1/\alpha}}{\sigma} (1 - \eta)^i \xi \right|^2 \right)
\]

(B.4)

\[
= \log \left( 1 - \eta (1 - \eta)^{\alpha i} |\xi|^\alpha + \frac{c\alpha}{(2 - \alpha)\sigma^2} \eta^{2\alpha i} (1 - \eta)^{2i} |\xi|^2 \right).
\]

Observe that

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

26
Therefore, there is some constant $C = C(\alpha, \sigma) > 0$ such that for small enough $x > 0$
\[\log \left( 1 - x + \frac{c_1 \alpha}{(2 - \alpha)\sigma^2} x^{2/\alpha} \right) \geq -x + C x^{2/\alpha}.\]
If we use this in (B.4), we obtain for all $|\xi| \leq 1$ and small enough $\eta > 0$,
\[\log \varphi \left( \frac{\eta^{1/\alpha}}{\sigma} (1 - \eta)^i \xi \right) \geq -\eta |(1 - \eta)^{ai}| \xi|^{\alpha} + C \eta^{2/\alpha} (1 - \eta)^{2i}|\xi|^2.\]
Inserting this into (B.3), we see for all $|\xi| \leq 1$ and small enough $\eta > 0$
\[\log E \left[ e^{i\xi Y} \right] = \sum_{i=0}^{\infty} \log \varphi \left( \frac{\eta^{1/\alpha}}{\sigma} (1 - \eta)^i \xi \right) \geq -\eta |(1 - \eta)^a| \xi|^{\alpha} + C \eta^2 (1 - \eta)^2|\xi|^2 \sum_{i=0}^{\infty} (1 - \eta)^{2i} \]
\[= -\eta |\xi|^{\alpha} \frac{\eta}{1 - (1 - \eta)^a} + C |\xi|^2 \frac{\eta^{2/\alpha}}{1 - (1 - \eta)^2} \]
\[= -\frac{1}{\alpha} |\xi|^{\alpha} |\eta|^{\alpha} + |\xi|^2 \frac{\eta^{2/\alpha}}{1 - (1 - \eta)^2} \Omega(\eta^{2/\alpha - 1}).\]
In the last equality we use that
\[\lim_{\eta \to 0} \frac{\eta}{1 - (1 - \eta)^a} = \frac{\alpha - 1}{2\alpha} \quad \text{and} \quad \lim_{\eta \to 0} \frac{\eta^{2/\alpha}}{1 - (1 - \eta)^2} = \frac{1}{2}.\]
Here and in the following, the notation $f(\eta) = \Omega(g(\eta))$ as $\eta \downarrow 0$ means that $\lim_{\eta \to 0} \frac{f(\eta)}{g(\eta)}$ is a positive (finite) constant, where $f$ and $g$ are some positive functions. With the elementary inequality $e^x \geq 1 + x$ for $x \in \mathbb{R}$ we see for all $|\xi| \leq 1$ and sufficiently small $\eta > 0$ that
\[E \left[ e^{i\xi Y} \right] \geq \exp \left[ -\frac{1}{\alpha} |\xi|^{\alpha} - |\xi|^2 \frac{\eta^{2/\alpha}}{1 - (1 - \eta)^2} \right] \Omega(\eta^{2/\alpha - 1}) \]
\[\geq e^{-|\xi|^{\alpha}/\alpha} \left[ 1 - |\xi|^2 \frac{\eta^{2/\alpha}}{1 - (1 - \eta)^2} \right] \Omega(\eta^{2/\alpha - 1}),\]
which yields
\[(B.5) \quad \int_{-1}^{1} \left( E \left[ e^{i\xi Y} \right] - E \left[ e^{i\xi^{2/\alpha - 1} Z_1} \right] \right) d\xi \geq \int_{-1}^{1} e^{-|\xi|^{\alpha}/\alpha} \left[ -|\xi|^2 \frac{\eta^{2/\alpha}}{1 - (1 - \eta)^2} \right] d\xi \]
\[= -\Omega(\eta) + \Omega(\eta^{2/\alpha - 1}) = \Omega(\eta^{2/\alpha - 1}).\]
Define
\[h(x) := \frac{1}{M} \left( \frac{\sin x}{x} \mathbf{1}_{\{x \neq 0\}} + \mathbf{1}_{\{x = 0\}} \right), \quad x \in \mathbb{R},\]
where
\[M := \sup_{x \in \mathbb{R}\setminus\{0\}} \left| \frac{x \cos x - \sin x}{x^2} \right| \in (0, \infty).\]
Since $h \in \text{Lip}(1)$ and
\[h(x) = \frac{1}{2M} \int_{-1}^{1} e^{ix} d\xi, \quad x \in \mathbb{R},\]
it follows from Fubini’s theorem and (B.5) that
\[
W_1(\mu, \tilde{\mu}_\eta) \geq \left| \mathbb{E} \left[ h(\tilde{Y}_\eta) \right] - \mathbb{E} \left[ h(\alpha^{-1/\alpha} Z_1) \right] \right| \\
= \left| \int_{\mathbb{R}} \left( \frac{1}{2M} \int_{-1}^1 e^{i\xi x} \, dx \right) \mathbb{P}(\tilde{Y}_\eta \in dx) - \int_{\mathbb{R}} \left( \frac{1}{2M} \int_{-1}^1 e^{i\xi x} \, dx \right) \mathbb{P}(\alpha^{-1/\alpha} Z_1 \in dx) \right| \\
= \left| \frac{1}{2M} \int_{-1}^1 \mathbb{E} \left[ e^{i\xi \tilde{Y}_\eta} \right] \, d\xi - \frac{1}{2M} \int_{-1}^1 \mathbb{E} \left[ e^{i\xi \alpha^{-1/\alpha} Z_1} \right] \, d\xi \right| \\
= \frac{1}{2M} \left| \int_{-1}^1 \left( \mathbb{E} \left[ e^{i\xi \tilde{Y}_\eta} \right] - \mathbb{E} \left[ e^{i\xi \alpha^{-1/\alpha} Z_1} \right] \right) \, d\xi \right| \\
\geq \Omega(\eta^{2/\alpha-1}).
\]

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

**APPENDIX C. ALTERNATIVE PROOF OF LEMMA 2.1 BY MALLIAVIN CALCULUS [34]**

In this section, we will prove Lemma 2.1 by Malliavin calculus established in [34]. Keep the same notations as in Subsection 3.3 and we consider the integral form of SDE (3.11), that is,

\[
(C.1) \quad X_t^{x;\mu} = x + \int_0^t b(X_r^{x;\mu}) \, dr + W_{t_t - t_0}.
\]

Then \(X_t^{x;\mu}\) is Malliavin differentiable. In the following we will use the standard notations from [34].

Recall that \(\gamma^c\) is the inverse function of \(l^c\). The Malliavin derivative \(D_sX_t^{x;\mu} = 0\) for \(s > l_t^c\), which amounts to \(\gamma_s^c > t\). Thus,

\[
D_sX_t^{x;\mu} = I_{\{t \geq \gamma_s^c\}} D_sX_t^{x;\mu}.
\]

For \(t \geq \gamma_s^c\), (C.1) implies

\[
D_sX_t^{x;\mu} = I_d + \int_{\gamma_s^c}^t \nabla_b(X_r^{x;\mu}) D_sX_r^{x;\mu} \, dr.
\]

Now, we consider the tangent flow \(\nabla_x X_t^{x;\mu}\), which verifies the equation

\[
\nabla_x X_t^{x;\mu} = I_d + \int_0^t \nabla_b(X_r^{x;\mu}) \nabla_x X_r^{x;\mu} \, dr \\
= \nabla_x X_{\gamma_s^c}^t + \int_{\gamma_s^c}^t \nabla_b(X_r^{x;\mu}) \nabla_x X_r^{x;\mu} \, dr.
\]

It is clear that \(\nabla_x X_t^{x;\mu}\) is invertible, so it holds that

\[
D_sX_t^{x;\mu} = \nabla_x X_t^{x;\mu} \left( \nabla_x X_{\gamma_s^c}^t \right)^{-1}.
\]

Using the chain rule, for \(\phi \in C^1_0(\mathbb{R}^d, \mathbb{R})\),

\[
D_s\phi(X_t^{x;\mu}) = \nabla \phi(X_t^{x;\mu})D_sX_t^{x;\mu} = \nabla \phi(X_t^{x;\mu}) \nabla_x X_t^{x;\mu} \left( \nabla_x X_{\gamma_s^c}^t \right)^{-1}.
\]

Hence, we have

\[
\nabla \phi(X_t^{x;\mu}) \nabla_x X_t^{x;\mu} = D_s\phi(X_t^{x;\mu}) \nabla_x X_{\gamma_s^c}^t.
\]

Based on the above framework, one can derive the following Bismut–Elworthy–Li formula.
Lemma C.1. Let $G$ be a $d$-dimensional vector which is Malliavin differentiable. Then for any $f \in C^1_b(\mathbb{R}^d, \mathbb{R})$,

$$
\mathbb{E}[\nabla f(X_t^{x:l'}) \nabla_x X_t^{x:l'} G] = \frac{1}{l'_t - l'_0} \mathbb{E} \left[ f(X_t^{x:l'}) \left( G \int_{l'_0}^{l'_t} \nabla_x X_{\gamma^{-1}_s}^{x:l'} \, dW_{s-l'_0} + \int_{l'_0}^{l'_t} \nabla_x X_{\gamma^{-1}_s}^{x:l'} D_s G \, ds \right) \right].
$$

Proof. For any $s \in [l'_0, l'_t]$ (this amounts to $t \geq \gamma^{-1}_s$), we have

$$
\mathbb{E} \left[ \nabla f(X_t^{x:l'}) \nabla_x X_t^{x:l'} G \right] = \mathbb{E} \left[ D_s f(X_t^{x:l'}) \nabla_x X_t^{x:l'} G \right].
$$

Integrating over $s \in [l'_0, l'_t]$, one can derive from the duality relation [34, (1.42)] that

$$
\mathbb{E} \left[ \nabla f(X_t^{x:l'}) \nabla_x X_t^{x:l'} G \right] = \frac{1}{l'_t - l'_0} \int_{l'_0}^{l'_t} \mathbb{E} \left[ D_s f(X_t^{x:l'}) \nabla_x X_t^{x:l'} G \right] \, ds
$$

$$
= \frac{1}{l'_t - l'_0} \mathbb{E} \left[ f(X_t^{x:l'}) \int_{l'_0}^{l'_t} \nabla_x X_t^{x:l'} G \, dW_{s-l'_0} \right],
$$

where $\tilde{d}W_{s-l'_0}$ denotes the Skorohod integral (see [34, Subsection 1.3.2]). Recall the following well-known formula (cf. [34, (1.56)])

$$
\int_{l'_0}^{l'_t} \nabla_x X_{\gamma^{-1}_s}^{x,l'} G \, dW_{s-l'_0} = G \int_{l'_0}^{l'_t} \nabla_x X^{x,l'}_{\gamma^{-1}_s} \, dW_{s-l'_0} + \int_{l'_0}^{l'_t} \nabla_x X_{\gamma^{-1}_s}^{x,l'} D_s G \, ds.
$$

Noting that $X^{x,l'}_{\gamma^{-1}_s}$ depends on $W_r$ with $r \leq s-l'_0$, it follows that $s \mapsto \nabla_x X^{x,l'}_{\gamma^{-1}_s}$ is $F^{W}_{s-l'_0}$-adapted and consequently the Skorohod integral coincides with the Itô integral (see [34, Subsection 1.3.3]). Then the desired result follows. \hfill \Box

Now, we can use the above Bismut-Elworthy-Li formula to prove the following lemma.

Lemma C.2. Under Assumption A, for all $\phi \in C^2_b(\mathbb{R}^d, \mathbb{R})$, $x \in \mathbb{R}^d$ and $t \in (0, 1]$,

$$
\left\| \mathbb{E} \left[ \nabla^2 \phi(X_t^{x,l'}) \right] \right\|_{HS} \leq C \| \nabla \phi \|_{\infty} \left( 1 + \frac{1}{\sqrt{t_t}} \right).
$$

Proof. By the chain rule, for $i, j = 1, \ldots, d$,

$$
\partial_i \partial_j (X_t^{x,l'}) = \sum_{k=1}^{d} \partial_k \phi(X_t^{x,l'}) \partial_k \partial_i X_{tk}^{x,l'} + \sum_{p=1}^{d} \sum_{k=1}^{d} \partial_p \partial_k \phi(X_t^{x,l'}) \partial_j X_{tkp}^{x,l'} \partial_k X_{tk}^{x,l'},
$$

where $X_{tk}^{x,l'}$ is the $k$-th component of the $d$-dimensional vector $X_{tk}^{x,l'}$, $k = 1, 2, \ldots, d$. Using Lemma C.1 with $d = 1, G = \partial_t X_{tk}^{x,l'}$ and $f = \partial_k \phi$, we have

$$
\mathbb{E} \left[ \partial_t \partial_k \phi(X_t^{x,l'}) \partial_j X_{tkp}^{x,l'} \partial_k X_{tk}^{x,l'} \right]
$$

$$
= \frac{1}{l'_t - l'_0} \mathbb{E} \left[ \partial_k \phi(X_t^{x,l'}) \left( \partial_t X_{tk}^{x,l'} \int_{l'_0}^{l'_t} \partial_j X_{\gamma^{-1}_s}^{x,l'} \, dW_{s-l'_0} + \int_{l'_0}^{l'_t} \partial_j X_{\gamma^{-1}_s}^{x,l'} D_s \partial_t X_{tk}^{x,l'} \, ds \right) \right],
$$

where $W_{s-l'_0}$ is the $p$-th component of the $d$-dimensional Brownian motion $W_{s-l'_0}$, $p = 1, \ldots, d$. Then we obtain that

$$
\left| \mathbb{E} \left[ \partial_j \partial_t \phi(X_t^{x,l'}) \right] \right| \leq \| \nabla \phi \|_{\infty} \sum_{k=1}^{d} \mathbb{E} \left| \partial_j \partial_k X_{tk}^{x,l'} \right|
$$

+ \left| \nabla \phi \right|_{\infty} \frac{1}{l'_t - l'_0} \sum_{p=1}^{d} \sum_{k=1}^{d} \mathbb{E} \left[ \partial_t X_{tk}^{x,l'} \int_{l'_0}^{l'_t} \partial_j X_{\gamma^{-1}_s}^{x,l'} \, dW_{s-l'_0} + \int_{l'_0}^{l'_t} \partial_j X_{\gamma^{-1}_s}^{x,l'} D_s \partial_t X_{tk}^{x,l'} \, ds \right].
In addition, by a similar argument as in the proof of Lemma 3.3, it is not hard to verify that for any \( t \in (0, 1] \), \( s \in [t_0, t] \) and \( i, j, k = 1, \ldots, d \), the derivatives \( \partial_j \partial_i \varphi(X_t^{s,t}) \), \( \partial_i X_t^{s,t} \) and \( D_s \partial_i X_t^{s,t} \) are all bounded. Then it follows from the Cauchy–Schwarz inequality and Itô’s isometry that

\[
\left| \mathbb{E} \left[ \partial_j \partial_i \varphi(X_t^{s,t}) \right] \right| \leq C \| \nabla \varphi \|_{\infty} + C \| \nabla \varphi \|_{\infty} \left( \sqrt{t_t - t_0} + t_t - t_0 \right)
\]

\[
\leq C \| \nabla \varphi \|_{\infty} \left( 1 + \frac{1}{\sqrt{t_t - t_0}} \right),
\]

which further implies

\[
\left\| \mathbb{E} \left[ \nabla_x^2 \varphi(X_t^{s,t}) \right] \right\|_{\text{HS}} \leq C \| \nabla \varphi \|_{\infty} \left( 1 + \frac{1}{\sqrt{t_t - t_0}} \right).
\]

By the same argument as in the proof of [50, Lemma 2.2], one can prove that

\[(C.2) \quad \lim_{\epsilon \downarrow 0} \mathbb{E} \left[ \nabla_x^2 \varphi(X_t^{s,t}) \right] = \mathbb{E} \left[ \nabla_x^2 \varphi(X_t^{s,t}) \right].\]

It remains to let \( \epsilon \downarrow 0 \) to get the desired estimate. \( \square \)

With the help of Lemma C.2, we can prove Lemma 2.1 as in the proof presented in Subsection 3.4.

**Acknowledgement.** We would like to gratefully thank the associate editor and the two anonymous referees for their very helpful and constructive comments. 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). C.-S. Deng is supported by Natural Science Foundation of Hubei Province of China (2022CFB129). 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).

**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. Chet 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. J. Jacod (2004): The Euler scheme for Lévy driven stochastic differential equations: limit theorems. *The Annals of Probability* 32, 1830–1872.

17. 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.

18. A. Janicki, Z. Michna and A. Weron (1996): Approximation of stochastic differential equations driven by $\alpha$-stable Lévy motion. *Applicationes Mathematicae* 24, 149–168.

19. 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.

20. 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.

21. 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.

22. 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.

23. J. Norris (1986): Simplified Malliavin calculus. In: *Séminaire de Probabilités XX 1984/85*. Springer, Berlin, pp. 101–130.

24. D. Nualart (2006): *The Malliavin Calculus and Related Topics* (2nd ed). Springer, Berlin.

25. J.-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.

26. J.P. Nolan (2008): An overview of multivariate stable distributions. Online: http://hdl.handle.net/1961/auislandora:68717 (accessed January 12, 2023).

27. J.P. Nolan (2020): *Univariate stable distributions: models for heavy tailed data*. Springer Series in Operations Research and Financial Engineering, Springer, Cham.

28. 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.

29. C. Thomas (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.
39. P. Protter and D. Talay (1997): The Euler scheme for Lévy driven stochastic differential equations. *The Annals of Probability* 25, 393–423.

40. 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.

41. K. Sato (1999): *Lévy processes and infinitely divisible distributions*. Cambridge University Press, Cambridge.

42. J. Shao (2018): Weak convergence of Euler-Maruyama's approximation for SDEs under integrability condition. Preprint arXiv:1808.07250.

43. 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.

44. 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.

45. B. Tarami and M. Avaji (2018): Convergence of Euler-Maruyama Method for Stochastic Differential Equations Driven by $\alpha$-stable Lévy Motion. *Journal of Mathematical Extension* 12, 31–53.

46. F. Y. Wang (2020): Exponential contraction in Wasserstein distances for diffusion semigroups with negative curvature. *Potential Analysis* 53, 1123–1144.

47. J. Wang (2008): Criteria for ergodicity of Lévy type operators in dimension one. *Stochastic processes and their applications* 118, 1909–1928.

48. J. Wang (2016): $L^p$-Wasserstein distance for stochastic differential equations driven by Lévy processes. *Bernoulli* 22, 1598–1616.

49. L. Xu (2019): Approximation of stable law in Wasserstein-1 distance by Stein’s method. *Annals of Applied Probability* 29, 458–504.

50. X. Zhang (2013): Derivative formulas and gradient estimates for SDEs driven by $\alpha$-stable processes. *Stochastic Processes and their Applications* 123, 1213–1228.

51. 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

Email address: chenpengmath@nuaa.edu.cn

(C.-S. Deng) SCHOOL OF MATHEMATICS AND STATISTICS, WUHAN UNIVERSITY, WUHAN 430072, CHINA

Email address: dengcs@whu.edu.cn

(R.L. Schilling) TU DRESDEN, FAKULTÄT MATHEMATIK, INSTITUT FÜR MATHEMATISCHE STOCHASTIK, 01062 DRESDEN, GERMANY

Email 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

Email address: lihuxu@um.edu.mo