Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2021, Vol. 57, No. 1, 250–271
https://doi.org/10.1214/20-AIHP1077
© Association des Publications de l'Institut Henri Poincaré, 2021

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES

# Existence of densities for stochastic differential equations driven by Lévy processes with anisotropic jumps

Martin Friesen[a,*], Peng Jin[b] and Barbara Rüdiger[a,†]

[a] *School of Mathematics and Natural Sciences, University of Wuppertal, Germany.*
*E-mail: [*]friesen@math.uni-wuppertal.de; [†]ruediger@uni-wuppertal.de*
[b] *Department of Mathematics, Shantou University, Shantou, Guangdong 515063, China. E-mail: pjin@stu.edu.cn*

Received 17 October 2018; revised 1 October 2019; accepted 15 June 2020

**Abstract.** We study existence and Besov regularity of densities for solutions to stochastic differential equations with Hölder continuous coefficients driven by a $d$-dimensional Lévy process $Z = (Z(t))_{t\geq 0}$, where, for $t > 0$, the density function $f_t$ of $Z(t)$ exists and satisfies, for some $(\alpha_i)_{i=1,\dots,d} \subset (0,2)$ and $C > 0$,

$$\limsup_{t\to 0} t^{1/\alpha_i} \int_{\mathbb{R}^d} \left| f_t(z + e_i h) - f_t(z) \right| dz \leq C|h|, \quad h \in \mathbb{R}, i = 1, \dots, d.$$

Here $e_1, \dots, e_d$ denote the canonical basis vectors in $\mathbb{R}^d$. The latter condition covers anisotropic $(\alpha_1, \dots, \alpha_d)$-stable laws but also particular cases of subordinate Brownian motion. To prove our result we use some ideas taken from (J. Funct. Anal. **264** (2013), 1757–1778).

**Résumé.** Nous étudions le problème de l'existence et de l'appartenance à un espace de Besov pour les densités de solutions d'équations différentielles stochastiques à coefficients hölderiens, conduites par un processus de Lévy $d$-dimensionnel $Z = (Z(t))_{t\geq 0}$, où, pour $t > 0$, la densité $f_t$ de la loi de $Z(t)$ existe et vérifie, pour un certain $(\alpha_i)_{i=1,\dots,d} \subset (0,2)$ et $C > 0$,

$$\limsup_{t\to 0} t^{1/\alpha_i} \int_{\mathbb{R}^d} \left| f_t(z + e_i h) - f_t(z) \right| dz \leq C|h|, \quad h \in \mathbb{R}, i = 1, \dots, d.$$

Ici, $e_1, \dots, e_d$ désignent les vecteurs de la base canonique de $R^d$. La précédente condition s'applique au cas de lois anisotropiques $(\alpha_1, \dots, \alpha_d)$-stables, mais aussi à des cas particuliers de mouvements browniens subordonnés. Pour démontrer ces résultats, nous utilisons certaines idées de (J. Funct. Anal. **264** (2013), 1757–1778).

*MSC2020 subject classifications:* 60H10; 60E07; 60G30

*Keywords:* Stochastic differential equation with jumps; Anisotropic Lévy process; Anisotropic Besov space; Transition density

## 1. Introduction

Let $d \geq 1$ and $X = (X(t))_{t\geq 0} \subset \mathbb{R}^d$ be a solution to the Lévy driven SDE

$$dX(t) = b\big(X(t)\big)\,dt + \sigma\big(X(t-)\big)\,dZ(t), \tag{1.1}$$

where $b \in \mathbb{R}^d$ is the drift, $\sigma \in \mathbb{R}^{d\times d}$ the diffusion coefficient and $Z$ a $d$-dimensional pure jump Lévy process with Lévy measure $\nu$, i.e.,

$$\mathbb{E}\big[\mathrm{e}^{\mathrm{i}\langle \xi, Z(t)\rangle}\big] = \mathrm{e}^{-t\Psi_\nu(\xi)}, \qquad \Psi_\nu(\xi) = \int_{\mathbb{R}^d} \big(1 + \mathbb{1}_{\{|z|\leq 1\}} \mathrm{i}\langle \xi, z\rangle - \mathrm{e}^{\mathrm{i}\langle \xi, z\rangle}\big)\nu(dz). \tag{1.2}$$

If the drift and diffusion coefficients $b$, $\sigma$ in (1.1) are smooth enough, then there exists a unique strong solution to (1.1). Moreover, under additional regularity properties of the driving noise Z, existence of a smooth density for the law of $X(t)$

can be obtained by the Malliavin calculus (see, e.g., [2,19,21,31,32]). For Hölder continuous coefficients, as studied in this work, the situation is much less developed. One possibility is to apply the parametrix method, see, e.g., [5,14,15,18], which under certain restrictions on the coefficients, typically provides a construction and two-sided estimates for the heat kernel and thus uniqueness in law for (1.1). For these cases one assumes, in addition, that $b$, $\sigma$ are bounded, $\sigma$ is uniformly non-degenerate, i.e., $\inf_{x\in\mathbb{R}^d}\inf_{|v|=1}|\sigma(x)v| > 0$, and most of the results are designed for a driving noise $Z$ which is comparable to a

(Z1) rotationally symmetric $\alpha \in (0,2)$-stable Lévy process with characteristic exponent $\Psi_\nu(\xi) = |\xi|^\alpha$, $\xi \in \mathbb{R}^d$.

Recently, in [8] a simple method for proving existence of a density on

$$\Gamma = \left\{x \in \mathbb{R}^d \mid \sigma(x) \text{ is invertible}\right\} \tag{1.3}$$

for solutions to (1.1) with bounded and Hölder continuous coefficients has been developed. Their main condition was formulated in terms of an integrability and non-degeneracy condition on the Lévy measure $\nu$ from which, in particular, the following crucial estimate was derived

$$c|\xi|^\alpha \le \operatorname{Re}\bigl(\Psi_\nu(\xi)\bigr) \le C|\xi|^\alpha, \quad |\xi| \gg 1, \tag{1.4}$$

where $c, C > 0$ and $\alpha \in (0,2)$. Similar ideas have been first applied in dimension $d = 1$, see [11], where the density was studied in terms of its characteristic function. The technique in [8] has been further successfully applied in [9,23,24] to the Navier–Stokes equations driven by Gaussian noises in $d = 3$, and in [10] to the space-homogeneous Boltzmann equation. A summary of this method and further extensions can be found in [25]. At this point it is worthwile to mention that this method, in contrast to the parametrix method, provides neither weak uniqueness for (1.1) nor estimates on the heat kernel.

In the recent years we observe an increasing interest in the study of (1.1) for a Lévy process $Z$ with anisotropic jumps (see, e.g., [1,3,4,16,17]). The most prominent example here is

(Z2) $Z = (Z_1^{\alpha_1}, \dots, Z_d^{\alpha_d})$, where $Z_1^{\alpha_1}, \dots, Z_d^{\alpha_d}$ are independent, one-dimensional symmetric $\alpha_1, \dots, \alpha_d \in (0,2)$-stable Lévy processes, i.e., one has for some constants $c_{\alpha_1}, \dots, c_{\alpha_d} > 0$

$$\nu(dz) = \sum_{j=1}^d c_{\alpha_j} \frac{dz_j}{|z_j|^{1+\alpha_j}} \otimes \prod_{k\neq j} \delta_0(dz_k), \qquad \Psi_\nu(\xi) = |\xi_1|^{\alpha_1} + \dots + |\xi_d|^{\alpha_d}. \tag{1.5}$$

Since (Z2) satisfies (1.4) only for the case where $\alpha_1 = \dots = \alpha_d$, one obtains existence of a density on $\Gamma$ from [8], provided that $b$, $\sigma$ are bounded, Hölder continuous and $\alpha_1 = \dots = \alpha_d$. If $\alpha_1, \dots, \alpha_d$ are possibly different, $b = 0$, $\sigma$ is Lipschitz continuous, bounded and uniformly non-degenerate, then existence of a heat kernel (being lower semi-continuous) and the strong Feller property are obtained in [17]. In contrast, this work provides the first result on the existence of a density for solutions to (1.1) with Hölder continuous coefficients, $\sigma$ being not necessarily uniformly non-degenerate and $Z$ being given by (Z2) and that has possibly different indices $\alpha_1, \dots, \alpha_d$. However, we neither study uniqueness for solutions to (1.1) nor investigate pointwise estimates on the obtained densities. Nevertheless, if one knows (by other methods) that (1.1) determines a Feller process, then the results of this work can be used to prove the strong Feller property, see also [12] for the particular case of the anisotropic stable JCIR process.

## 2. Statement of the results

### 2.1. *Assumption on the Lévy noise*

Consider (1.1) for a Lévy process $Z = (Z(t))_{t\ge 0}$ with Lévy measure $\nu$ and characteristic exponent (1.2). We suppose that $Z$ satisfies the following condition:

(A1) For each $t > 0$, the law of $Z(t)$ has density $f_t$ and there exist $\alpha_1, \dots, \alpha_d \in (0,2)$ and a constant $C > 0$ such that, for any $k \in \{1, \dots, d\}$,

$$\limsup_{t\to 0} t^{1/\alpha_k} \int_{\mathbb{R}^d} \bigl|f_t(z + e_k h) - f_t(z)\bigr| \, dz \le C|h|, \quad h \in \mathbb{R}. \tag{2.1}$$

The anisotropic regularization property of the noise is reflected by (2.1). It imposes a growth condition for the Besov norm of $f_t$ at the singularity in $t = 0$ (since $f_t|_{t=0} = \delta_0$).

**Remark 2.1.** Let $W$, $Z$ be independent Lévy processes. If $Z$ satisfies (A1) for some $(\alpha_k)_{k\in\{1,\dots,d\}}$, then also $Z+W$ satisfies (A1) for the same $(\alpha_k)_{k\in\{1,\dots,d\}}$.

The following proposition can be used to provide many examples of Lévy processes satisfying condition (A1).

**Proposition 2.1.** *Let $m \le d$ and $I_1,\dots,I_m \subset \{1,\dots,d\}$ be disjoint with $I_1 \cup \cdots \cup I_d = \{1,\dots,d\}$. For each $j \in \{1,\dots,m\}$, let $Z^j$ be a $|I_j|$-dimensional Lévy process with Lévy measure $\nu_j$ and characteristic exponent*

$$\Psi_{\nu_j}(\xi) = \int_{\mathbb{R}^{|I_j|}} \left(1 + \mathrm{i}\mathbb{1}_{\{|z|\le 1\}}\langle \xi, z\rangle - \mathrm{e}^{\mathrm{i}\langle \xi, z\rangle}\right)\nu_j(dz).$$

*Suppose that $Z^1,\dots,Z^m$ are independent and there exist constants $c, C > 0$ and $\alpha_1,\dots,\alpha_m \in (0,2)$ such that*

$$c|\xi|^{\alpha_j} \le \mathrm{Re}\big(\Psi_{\nu_j}(\xi)\big) \le C|\xi|^{\alpha_j}, \quad |\xi| \gg 1, j \in \{1,\dots,m\}.$$

*Then $Z = (Z^1,\dots,Z^m)$ has a smooth density $f_t$ such that*

$$\int_{\mathbb{R}^d} \left|\nabla_{I_j} f_t(z)\right| dz \le Ct^{-1/\alpha_j}, \quad t \to 0, j \in \{1,\dots,m\}.$$

*In particular, $Z$ satisfies condition (A1).*

The proof of this statement essentially follows from the arguments given in [8, Lemma 3.3] combined with the mutual independence of the processes $Z^1,\dots,Z^m$, and hence is left to the reader. Note that this proposition is applicable to the fully isotropic case (Z1) ($m = 1$, $\alpha := \alpha_1$ and $I_1 = \{1,\dots,d\}$) but also to the fully anisotropic case (Z2) ($m = d$ and $I_j = \{j\}$). Below we provide two examples of Lévy measures which satisfy condition (A1).

**Example 2.1.** Let $Z_1,\dots,Z_d$ be independent one-dimensional pure-jump Lévy processes with Lévy measures

$$\nu_k = c_k^+ r^{-1-\alpha_k^+}\mathbb{1}_{(0,1]}(r)\,dr + c_k^-|r|^{-1-\alpha_k^-}\mathbb{1}_{[-1,0)}(r)\,dr + \mu_k,$$

where $c_1^\pm,\dots,c_d^\pm \ge 0$, $\alpha_1^\pm,\dots,\alpha_d^\pm \in (0,2)$, and $\mu_1,\dots,\mu_d$ are one-dimensional Lévy measures. If $c_k^+ + c_k^- > 0$ holds for each $k = 1,\dots,d$, then previous proposition is applicable and hence, by Remark 2.1, (A1) holds for

$$\alpha_k = \begin{cases} \alpha_k^+, & \text{if } c_k^+ \ne 0, c_k^- = 0, \\ \alpha_k^-, & \text{if } c_k^+ = 0, c_k^+ \ne 0, \\ \max\{\alpha_k^+, \alpha_k^-\}, & \text{if } c_k^+ \ne 0, c_k^- \ne 0. \end{cases}$$

An example of processes $Z$ which satisfy condition (A1) with Lévy measures having no absolutely continuous component with respect to the Lebesgue measure is given in [8, Example 1.6]. Condition (A1) also holds true for the case where each $Z_k$ is a subordinate Brownian motion.

**Example 2.2.** Let $S_1(t),\dots,S_d(t)$ be independent subordinators with Laplace exponents

$$\psi_k(\lambda) = \lambda^{\alpha_k/2}\big(\log(1+\lambda)\big)^{\beta_k/2}, \quad \alpha_k \in (0,2), \beta_k \in (-\alpha_k, 2-\alpha_k), k = 1,\dots,d.$$

Let $B(t)$ a $d$-dimensional Brownian motion independent of $S_1,\dots,S_d$, and define $Z_k(t) = B_k(S_k(t))$, $k = 1,\dots,d$. Let $\Psi_k$, $k = 1,\dots,d$, be the corresponding characteristic exponents. Then $Z = (Z_1,\dots,Z_d)$ satisfies condition (A1) with

$$\widetilde{\alpha}_k = \begin{cases} \alpha_k, & \beta_k \in [0, 2-\alpha_k), \\ \alpha_k - \varepsilon_k, & \beta_k \in (-\alpha_k, 0), \end{cases} \quad k \in \{1,\dots,d\},$$

for any choice of $\varepsilon_k \in (0,\alpha_k)$.

A proof that this example satisfies condition (A1) is given in the Appendix.

### 2.2. *Some notation*

For a $d \times d$ matrix $A$, we set $|A| = \sup_{|x|=1} |Ax|$. Then $1/|A^{-1}| = \inf_{|x|=1} |Ax|$ with the convention that $1/|A^{-1}| := 0$ if $A$ is not invertible. Given another $d \times d$ matrix $B$, we deduce that

$$\frac{1}{|A^{-1}|} \le |A|, \qquad \left| \frac{1}{|A^{-1}|} - \frac{1}{|B^{-1}|} \right| \le |A - B|. \tag{2.2}$$

Here and below we denote by $C$ a generic positive constant which may vary from line to line. Possible dependencies on other parameters are denoted by $C = C(a, b, c, \ldots)$. For $a, b \ge 0$, we set $a \wedge b := \min\{a, b\}$.

Here and below we say that $X$ is a solution to (1.1), if it is a weak solution to (1.1) in the following sense: There exist a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\ge 0}, \mathbb{P})$ with the usual conditions, an $(\mathcal{F}_t)_{t\ge 0}$-Lévy process $Z$, and an $(\mathcal{F}_t)_{t\ge 0}$-adapted cádlág process $X$ such that (1.1) is satisfied. In order to simplify the notation, we simply write $X$ for the (weak) solution to (1.1). Note that under the conditions imposed in this work existence and uniqueness of solutions to (1.1) not fully established. The classical existence and uniqueness theory for stochastic equations with jumps is discussed, e.g., in [13,28], see also [6,22,30] for some recent results in this direction.

For $\lambda \in (0, 1)$ the (classical) Besov space $B^{\lambda}_{1,\infty}(\mathbb{R}^d)$ is defined as the Banach space of $L^1(\mathbb{R}^d)$-functions $f$ with finite norm

$$\|f\|_{B^{\lambda}_{1,\infty}} := \|f\|_{L^1(\mathbb{R}^d)} + \sup_{|h|\le 1} |h|^{-\lambda} \|\Delta_h f\|_{L^1(\mathbb{R}^d)},$$

where $\Delta_h f(x) = f(x + h) - f(x)$, $h \in \mathbb{R}^d$, see [29] for additional details and references.

### 2.3. *General case*

We study (1.1) for bounded and Hölder continuous coefficients and keep track of the moments of the Lévy process $Z$, i.e., we suppose that:

(A2) There exist constants $\gamma \in (0, 2]$ and $\delta \in (0, \gamma]$ such that

$$\int_{\mathbb{R}^d} \left( \mathbb{1}_{\{|z|\le 1\}} |z|^{\gamma} + \mathbb{1}_{\{|z|>1\}} |z|^{\delta} \right) \nu(dz) < \infty. \tag{2.3}$$

Moreover $b$, $\sigma$ are bounded, and there exist $\beta \in [0, 1]$, $\chi \in (0, 1)$ and $C > 0$ such that

$$\left| \widetilde{b}(x) - \widetilde{b}(y) \right| \le C|x - y|^{\beta}, \qquad \left| \sigma(x) - \sigma(y) \right| \le C|x - y|^{\chi}, \quad x, y \in \mathbb{R}^d,$$

where $\widetilde{b}(x) = b(x) - \mathbb{1}_{(0,1)}(\gamma)\sigma(x) \int_{|z|\le 1} z\nu(dz)$.

Note that $\beta = 0$ corresponds to the case where $\widetilde{b}$ is only bounded. Let $\alpha_1, \ldots, \alpha_d$ be as in condition (A1). Set $\alpha^{\min} = \min\{\alpha_1, \ldots, \alpha_d\}$ and $\alpha^{\max} = \max\{\alpha_1, \ldots, \alpha_d\}$. The following is our first main result.

**Theorem 2.1.** *Let $Z$ be a Lévy process with Lévy measure $\nu$ and characteristic exponent* (1.2). *Assume that $Z$ satisfies condition* (A1) *and that condition* (A2) *is satisfied. Suppose that the constants $\gamma$, $\delta$, $\beta$, $\chi$ satisfy*:

(a) *If $\gamma \in [1, 2]$, then suppose that*

$$\alpha^{\min} \left( 1 + \frac{\beta \wedge \delta}{\gamma} \right) > \mathbb{1}_{\{b \ne 0\}}, \qquad \frac{\alpha^{\min}}{\gamma} \left( 1 + \chi \wedge (\delta/\gamma) \right) > 1. \tag{2.4}$$

(b) *If $\gamma \in (0, 1)$, then suppose that*

$$\alpha^{\min} \left( 1 + \frac{\beta}{\gamma} \right) > 1, \qquad \alpha^{\min} \left( \frac{1}{\gamma} + \chi \right) > 1, \qquad \alpha^{\min} + \beta > 1. \tag{2.5}$$

*Let $(X(t))_{t\ge 0}$ satisfy* (1.1) *and denote by $\mu_t$ the law of $X(t)$. Then there exists $\lambda \in (0, 1)$, $C > 0$ and $g_t \in B^{\lambda}_{1,\infty}(\mathbb{R}^d)$, for $t > 0$, such that $|\sigma^{-1}(x)|^{-1} \mu_t(dx) = g_t(x)\, dx$ and*

$$\|g_t\|_{B^{\lambda}_{1,\infty}} \le \frac{C}{(1 \wedge t)^{1/\alpha^{\min}}}, \quad t > 0. \tag{2.6}$$

*In particular, $\mu_t$ has, for each $t > 0$, a density on $\Gamma$ defined in* (1.3).

The indicatior function in (2.4) indicates that, for $b = 0$, the first condition in (2.4) can be omitted. Likewise, for $b = 0$, we may choose $\beta = \chi$ in condition (2.5). The proof of this Theorem is given in Section 4 where it is actually shown that the constant $C$ is independent of the particular choice of solution $X$. Without assuming finite moments for the big jumps as in (2.3), we still get existence of a density but may lose Besov regularity.

**Corollary 2.1.** *Let $Z$ be a Lévy process with Lévy measure $\nu$ and characteristic exponent* (1.2). *Assume that $Z$ satisfies condition* (A1) *and condition* (A2) *is satisfied for $\delta = 0$. Finally, suppose that either condition* (2.4) *holds for $\delta = \gamma \in [1, 2]$ or* (2.5) *is satisfied for $\gamma \in (0, 1)$. Let $(X(t))_{t\geq 0}$ satisfy* (1.1) *and denote by $\mu_t$ the distribution of $X(t)$. Then $\mu_t$ has, for each $t > 0$, a density when restricted to $\Gamma$.*

**Proof.** Observe that Theorem 2.1 is applicable for the truncated Lévy measure $\nu^{(n)}(dz) = \mathbb{1}_{\{|z|\leq n\}}\nu(dz)$, $n \in \mathbb{N}$, which has finite moments of all orders for the big jumps. Repeating the arguments given in [8, Corollary 1.2] proves the assertion. □

Below we reformulate Theorem 2.1 for the particular case where $Z$ is given as in (Z1).

**Remark 2.2.** Suppose that the Lévy process $Z$ is given as in (Z1). Then (A1) is satisfied for $\alpha = \alpha_1 = \cdots = \alpha_d$, and (2.3) holds for any choice of $\gamma \in (\alpha, 2]$ and $\delta \in (0, \alpha)$. Hence conditions (2.4) and (2.5) are reduced to:

(a) If $\alpha \in [1, 2)$, then (2.4) is automatically satisfied.
(b) If $\alpha \in (0, 1)$, then (2.5) is implied by $\alpha + \beta > 1$ which coincides with the condition imposed in [8, Theorem 1.1].

In particular, we recover for $Z$ given as in (Z1) the results obtained in [8].

Let us briefly comment on the conditions (2.4) and (2.5). Our proof is based on a small-time Euler approximation of the process $X$ for which a precise convergence rate $\kappa > 0$ is obtained, i.e., we find an $\mathcal{F}_{t-\varepsilon}$-measurable process $U^\varepsilon(t)$ such that $\mathbb{E}[|X(t) - X^\varepsilon(t)|^\eta] \leq C\varepsilon^{\eta\kappa}$, where $\eta \in (0, 1)$, $0 \leq t < 1 \wedge \varepsilon$, $\varepsilon \ll 1$ and $X^\varepsilon(t) = U^\varepsilon(t) + \sigma(X(t-\varepsilon))(Z(t) - Z(t-\varepsilon))$. Based on this approximation, we deduce regularity of the law of $X(t)$ from the regularity of the law of $Z(\varepsilon)$. In this framework the above conditions (2.4) and (2.5) read as $\alpha^{\min}\kappa > 1$. It is clear that such a rate of convergence $\kappa$ depends on the moments of the Lévy measure $\nu$, i.e., on $\gamma$ and $\delta$, but also on the Hölder indices $\beta$ and $\chi$. Hence $\alpha^{\min}\kappa > 1$ imposes a balance condition between the regularity of coefficients represented by $(\beta, \chi)$, moments of the Lévy measure represented by $(\gamma, \delta)$ and smoothness index of the driving Lévy noise represented by $\alpha^{\min}$.

When $\gamma \in [1, 2]$, we may simply take $U^\varepsilon(t) = X(t-\varepsilon) + \varepsilon b(X(t-\varepsilon))$, while for $\gamma \in (0, 1)$ this choice is not optimal. In order to obtain a good convergence rate when $\gamma \in (0, 1)$ we employ some ideas taken from [8]. A similar problem was also encountered in [14,18] for the particular case with $Z$ being given as in (Z1). Indeed, in [18] such a problem was treated by a different approximation for which the same balance condition $\alpha + \beta > 1$ (compare with (2.5) and Remark 2.2(b)) was used. In the same spirit, cases A and C from [14] are covered by (2.4) or (2.5), respectively, when applied to $Z$ given as in (Z1). Finally, case B therein reads as $\alpha + \alpha\beta > 1$ and hence implies our condition $\alpha + \beta > 1$ when $\alpha \in (0, 1)$.

Although it follows from Remark 2.2 that our result covers [8], the following example shows that Theorem 2.1 may be applicable to cases which cannot be teated by [8, Theorem 1.1].

**Example 2.3.** Consider (1.1) with $d \geq 2$ and suppose that the Lévy process $Z$ has Lévy measure

$$\nu(dz) = \frac{dz}{|z|^{d+\alpha}} + \frac{dz_1}{|z_1|^{1+\widetilde{\alpha}}} \otimes \prod_{k\neq 1} \delta_0(dz_k), \qquad \alpha, \widetilde{\alpha} \in (0, 2).$$

If $\alpha \neq \widetilde{\alpha}$, then condition $(H_\alpha)$ from [8] fails to hold. However, condition (A1) is satisfied with $\alpha_1 = \max\{\alpha, \widetilde{\alpha}\}$ and $\alpha_2 = \cdots = \alpha_d = \alpha$, while (2.3) holds for each $\gamma > \max\{\alpha, \widetilde{\alpha}\}$ and each $\delta < \min\{\alpha, \widetilde{\alpha}\}$.

Note that in the above example one may take any $\gamma > \alpha^{\max}$. The next example shows that this does not need to be the case, i.e. the obligation to keep track of the parameters $\gamma \in (0, 2]$ and $\delta \in (0, \gamma]$ that determine the moments of the Lévy measure is not artificial.

**Example 2.4.** Consider (1.1) with $d = 1$ and suppose that the Lévy process $Z$ is given by $Z = Z_0 + Z_1$ where $Z_0$, $Z_1$ are indepent Lévy processes with Lévy measures $\nu_0$, $\nu_1$ given by

$$\nu_0(dz) = \frac{dz}{|z|^{1+\alpha}}, \qquad \nu_1(dz) = \sum_{n=1}^{\infty} a_n^{-\widetilde{\alpha}} \delta_{a_n}(dz),$$

where $\alpha, \widetilde{\alpha} \in (0,2)$ are such that $\alpha < \widetilde{\alpha}$, $a_n = 2^{-c^n}$ and $c$ is an integer greater than $2/(2-\widetilde{\alpha})$. It is not difficult to see that condition (A1) is satisfied for $\alpha_1 = \alpha$. Since the law of $Z_1(t)$ is, for $t > 0$, not absolutely continuous with respect to the Lebesgue measure (see [20] and [26, Exercise 29.12]), the particular choice of index $\alpha_1$ cannot be improved, i.e. condition (A1) fails to hold for any choice of $\alpha_1 > \alpha$. Moreover, it can be shown that (2.3) holds for each $\delta \in (0, \alpha)$ and each $\gamma > \widetilde{\alpha}$.

For the case where $Z$ is given as in (Z2), the situation is slightly more complicated.

**Remark 2.3.** Let $Z = (Z_1, \dots, Z_d)$ be a Lévy process given by (Z2). Suppose that (A2) is satisfied. Then (2.3) holds for any $\gamma \in (\alpha^{\max}, 2]$ and $\delta \in (0, \alpha^{\min})$. Hence the assumptions of Theorem 2.1 are satisfied, provided:

(a) if $\alpha^{\max} \in [1,2)$, it holds that

$$\alpha^{\min} + \frac{\alpha^{\min}}{\alpha^{\max}}\left(\beta \wedge \alpha^{\min}\right) > \mathbb{1}_{\{b \neq 0\}}, \qquad \frac{\alpha^{\min}}{\alpha^{\max}}\left(1 + \chi \wedge \left(\frac{\alpha^{\min}}{\alpha^{\max}}\right)\right) > 1.$$

(b) if $\alpha^{\max} \in (0,1)$, it holds that

$$\frac{\alpha^{\min}}{\alpha^{\max}} + \alpha^{\min}\chi > 1, \qquad \alpha^{\min} + \frac{\alpha^{\min}}{\alpha^{\max}}(\beta \wedge \chi) > 1.$$

Finally we consider the particular case

$$dX(t) = \sigma\left(X(t-)\right) dZ(t), \quad t > 0 \tag{2.7}$$

with $Z$ being given as in (Z2). If $\sigma$ is continuous, bounded, uniformly non-degenerate and $\alpha_1 = \cdots = \alpha_d$, then existence and uniqueness in law were shown in [1]. If $\sigma$ is Lipschitz continuous, bounded, uniformly non-degenerate and $\alpha_1, \dots, \alpha_d \in (0,2)$ satisfy $\frac{\alpha^{\min}}{\alpha^{\max}} > \frac{2}{3}$, then the corresponding heat kernel was constructed in [17] where it was also shown that the corresponding Markov process is strong Feller. Our result complements [17] by providing Besov regularity for the densities of any solution to (2.7) when $\sigma$ is merely Hölder continuous and not necessarily uniformly non-degenerate. The price we have to pay is that the conditions given below are more restrictive than their condition $\frac{\alpha^{\min}}{\alpha^{\max}} > \frac{2}{3}$.

**Corollary 2.2.** *Let $Z$ be given by* (Z2). *Suppose that $\sigma$ is bounded and there exists $\chi \in (0,1)$ such that $|\sigma(x) - \sigma(y)| \le C|x-y|^{\chi}$ for all $x, y \in \mathbb{R}^d$. Finally, suppose that*

(a) *if $\alpha^{\max} \in [1,2)$, assume that $\frac{\alpha^{\min}}{\alpha^{\max}} > \frac{1}{1+\chi\wedge(\frac{\alpha^{\min}}{\alpha^{\max}})}$;*

(b) *if $\alpha^{\max} \in (0,1)$, assume that $\alpha^{\min} > \frac{1}{1+\chi}$.*

*Then Theorem 2.1 is applicable.*

**Proof.** If $\alpha^{\max} \in (0,1)$, then $\alpha^{\min} > \frac{1}{1+\chi}$ implies that $\frac{\alpha^{\min}}{\alpha^{\max}} + \alpha^{\min}\chi > 1$ and $\alpha^{\min} + \chi > 1$ hold. □

Note that uniqueness for (2.7) when $\sigma$ is Hölder continuous, bounded and uniformly non-degenerate still remains a challenging open problem.

### 2.4. *Diagonal case*

If $\sigma$ in (1.1) is diagonal, then each component $X_k$ is only driven by one Lévy process $Z_k$ and hence one may expect that previous conditions are too rough. Below we show that this is indeed the case, i.e., we study the system of stochastic equations

$$dX_k(t) = b_k\left(X(t)\right) dt + \sigma_k\left(X(t-)\right) dZ_k(t), \quad t \ge 0, k \in \{1, \dots, d\}, \tag{2.8}$$

where the driving Lévy noise $Z, \sigma_1, \dots, \sigma_d : \mathbb{R}^d \longrightarrow \mathbb{R}$ and the drift $b$ satisfy:

(A3) For each $k \in \{1, \dots, d\}$ there exist $\gamma_k \in (0, 2]$ and $\delta_k \in (0, \gamma_k]$ such that

$$\int_{\mathbb{R}^d} \left(\mathbb{1}_{\{|z|\le 1\}}|z_k|^{\gamma_k} + \mathbb{1}_{\{|z|>1\}}|z_k|^{\delta_k}\right)\nu(dz) < \infty, \quad k \in \{1, \dots, d\}. \tag{2.9}$$

Moreover, $b_1, \dots, b_d$ and $\sigma_1, \dots, \sigma_d$ are bounded and there exist $\beta_1, \dots, \beta_d \in [0, 1]$, $\chi_1, \dots, \chi_d \in (0, 1)$ and $C > 0$ such that, for each $k \in \{1, \dots, d\}$,

$$\left|\widetilde{b}_k(x) - \widetilde{b}_k(y)\right| \le C|x - y|^{\beta_k}, \qquad \left|\sigma_k(x) - \sigma_k(y)\right| \le C|x - y|^{\chi_k}, \quad x, y \in \mathbb{R}^d,$$

where $\widetilde{b}_k(x) = b_k(x) - \mathbb{1}_{(0,1)}(\gamma_k)\sigma_k(x)\int_{|z|\le 1} z_k \nu(dz)$.

Set $\gamma^{\max} = \max\{\gamma_1, \dots, \gamma_d\}$, $\delta^{\min} = \min\{\delta_1, \dots, \delta_d\}$, $\chi^{\min} = \min\{\chi_1, \dots, \chi_d\}$ and $\beta^{\min} = \min\{\beta_1, \dots, \beta_d\}$. The following is our second main result of this work.

**Theorem 2.2.** *Let $Z$ be a Lévy process with Lévy measure $\nu$ and characteristic exponent as in* (1.2). *Assume that $Z$ satisfies condition* (A1) *and* (A3) *and suppose that, for each $k \in \{1, \dots, d\}$, the following conditions are satisfied*

(a) *if $\gamma_k \in [1, 2]$, suppose that*

$$\alpha_k\left(1 + \frac{\beta_k \wedge \delta^{\min}}{\gamma^{\max}}\right) > \mathbb{1}_{\{b_k \ne 0\}}, \qquad \alpha_k\left(\frac{1}{\gamma_k} + \frac{\chi_k \wedge (\delta^{\min}/\gamma_k)}{\gamma^{\max}}\right) > 1; \tag{2.10}$$

(b) *if $\gamma_k \in (0, 1)$, suppose that*

$$\alpha_k\left(1 + \frac{\beta_k}{\gamma^{\max}}\right) > 1, \qquad \alpha_k\left(\frac{1}{\gamma_k} + \frac{\chi_k \wedge (\delta^{\min}/\gamma_k)}{\max\{1, \gamma^{\max}\}}\right) > 1, \qquad \alpha^{\min} + \beta^{\min} > 1. \tag{2.11}$$

*Let $(X(t))_{t\ge 0}$ satisfy* (2.8) *and denote by $\mu_t$ the law of $X(t)$. Then there exists $\lambda \in (0, 1)$ and $g_t \in B^{\lambda}_{1,\infty}(\mathbb{R}^d)$, for $t > 0$, such that $|\sigma^{-1}(x)|^{-1}\mu_t(dx) = g_t(x)\,dx$ and* (2.6) *holds. In particular, $\mu_t$ has, for each $t > 0$, a density on*

$$\Gamma = \left\{x \in \mathbb{R}^d \mid \sigma_k(x) \ne 0, k \in \{1, \dots, d\}\right\}.$$

A proof of this Theorem is given in Section 5. In contrast to the proof of Theorem 2.1, we estimate the convergence rate $\kappa_k$ of the small-time Euler approximations now for each component $k \in \{1, \dots, d\}$ separately. The regularity result is then deduced from the regularity of the law of $Z(\varepsilon)$ for which we employ an anisotropic analogue of [8, Lemma 2.1] (see Section 3). To summarize, conditions (2.10) and (2.11) read as $\alpha_k\kappa_k > 1$, for $k \in \{1, \dots, d\}$. We could have studied this case also by directly applying [8, Lemma 2.1] instead of our anisotropic analogue given in Section 3. However, corresponding results would require to replace $\alpha_k$ by $\alpha^{\min}$, i.e., would reduce to $\alpha^{\min}\kappa_k > 1$, for $k \in \{1, \dots, d\}$. The latter condition is certainly not sufficient for some particular models, see the example given at the end of this section or [12] for the anisotropic stable JCIR process.

Note that in condition (A3) one may always replace $\beta_1, \dots, \beta_d$ by $\beta^{\min}$, $\chi_1, \dots, \chi_d$ by $\chi^{\min}$, $\gamma_1, \dots, \gamma_d$ by $\gamma^{\max}$ and $\delta_1, \dots, \delta_d$ by $\delta^{\min}$. However, conditions (2.10) and (2.11) would become more restrictive in these cases. Hence, it is not artificial to assume different Hölder regularity for different components and to keep track of the moments of the Lévy measure $\nu$ in different directions.

We have the following analogue of Corollary 2.1.

**Corollary 2.3.** *Let $Z$ be a Lévy process with Lévy measure $\nu$ and characteristic exponent as in* (1.2). *Assume that $Z$ satisfies condition* (A1) *and condition* (A3) *holds for $\delta_1 = \dots = \delta_d = 0$. Assume that conditions* (a) *and* (b) *from Theorem* 2.2 *hold with $\delta_k = \min\{\gamma_1, \dots, \gamma_d\}$. Let $(X(t))_{t\ge 0}$ satisfy* (2.8) *and denote by $\mu_t$ the law of $X_t$. Then $\mu_t$ has, for each $t > 0$, a density on*

$$\Gamma = \left\{x \in \mathbb{R}^d \mid \sigma_k(x) \ne 0, k \in \{1, \dots, d\}\right\}.$$

If $Z$ is given as in (Z2), we have the following.

**Remark 2.4.** Let $Z = (Z_1, \dots, Z_d)$ be a Lévy process given by (Z2). Suppose that (A3) is satisfied and $\sigma$ is diagonal. Then (2.9) is satisfied for $\gamma_k \in (\alpha_k, 2]$ and $\delta_k \in (0, \alpha_k)$. In particular, Theorem 2.2 is applicable, provided for each $k \in$

$\{1, \dots, d\}$ with $\alpha_k \in (0,1)$ one has

$$\alpha^{\min} + \beta^{\min} > 1, \qquad \alpha_k + \frac{\alpha_k}{\alpha^{\max}} \beta_k > 1.$$

Note that, for $k \in \{1, \dots, d\}$ and $\alpha_k \in [1,2)$, no additional restriction on the parameters have to be imposed.

Finally, let us consider the system of stochastic equations

$$dX_k(t) = \sigma_k\big(X(t-)\big)\, dZ_k(t), \quad t \geq 0, k \in \{1, \dots, d\}, \tag{2.12}$$

where $Z$ is given by (Z2). If $\sigma_1, \dots, \sigma_d$ are continuous, bounded and $\sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_d)$ is uniformly non-degenerate, then existence and uniqueness in law holds for (2.12), see [4]. Moreover, if in addition $\alpha_1 = \cdots = \alpha_d$, then the corresponding Markov process $X(t)$ has a transition density for which also sharp two-sided estimates have been obtained in [16]. Below we complement these results for the case where $\alpha_1, \dots, \alpha_d \in (0,2)$ may be different.

**Corollary 2.4.** *Let $Z$ be given by* (Z2) *and suppose that $\sigma_1, \dots, \sigma_d$ are bounded, $\chi$-Hölder continuous with $\chi \in (0,1)$ and $\sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_d)$ is uniformly non-degenerate. If $\alpha^{\max} \in (0,1)$ suppose, in addition, that*

$$\alpha^{\min} + \min\{\chi_1, \dots, \chi_d\} > 1.$$

*Then the corresponding transition probabilities $P_t(x, dy)$ have a heat kernel $p_t(x,y)$ and $\mathbb{R}^d \ni x \longmapsto p_t(x, \cdot) \in L^1(\mathbb{R}^d)$ is continuous. In particular, the Markov process determined by* (2.12) *has the strong Feller property.*

**Proof.** Since weak existence and uniqueness in law holds for (2.12) due to [4], the corresponding Markov process exists and one may check that it has a Markov transition kernel $P_t(x, dy)$. Using classical tightness arguments, it is not difficult to show that the Markov process determined by (2.12) has the Feller property. Applying a variant of Theorem 2.2 to this particular case (see also Remark 5.1) shows that $P_t(x, dy) = p_t(x,y)dy$ with $\|p_t(x, \cdot)\|_{B^{\lambda}_{1,\infty}} \leq C(1 \wedge t)^{-1/\alpha^{\min}}$ for each $x \in \mathbb{R}^d$ and $t > 0$. Note that the constant $C > 0$ can be choosen to be independent of $x$. Arguing like in [12, Lemma 4.3] (with the particular choice $\rho_\delta \equiv 1$ therein) proves the assertion. □

Comparing our result with [17] we do not impose the condition $\frac{\alpha^{\min}}{\alpha^{\max}} > \frac{2}{3}$ in order to prove the strong Feller property. It would be interesting to obtain similar results to [16] and [17] for this case.

### 2.5. *Structure of the work*

This work is organized as follows. In Section 3 we introduce our main technical tool of this work. Section 4 is devoted to the proof of Theorem 2.1, while Theorem 2.2 is proved in Section 5. Basic estimates on stochastic integrals with respect to Lévy processes, technical details for the small-time Euler approximations used in the proofs of our main results and a proof that Example 2.2 satisfies condition (A1) are given in the Appendix.

## 3. Main technical tool

### 3.1. *Anisotropic Besov space*

Anisotropic smoothness related to processes given as in (A1) will be measured by an anisotropic analogue of classical Besov spaces as introduced below. We call a collection of numbers $a = (a_1, \dots, a_d)$ an anisotropy if it satisfies

$$0 < a_1, \dots, a_d < \infty \quad \text{and} \quad a_1 + \cdots + a_d = d. \tag{3.1}$$

Let $a = (a_1, \dots, a_d)$ be an anisotropy and take $\lambda > 0$ with $\lambda / a_k \in (0,1)$ for all $k \in \{1, \dots, d\}$. The anisotropic Besov space $B^{\lambda,a}_{1,\infty}(\mathbb{R}^d)$ is defined as the Banach space of $L^1$-functions $f$ with finite norm

$$\|f\|_{B^{\lambda,a}_{1,\infty}} := \|f\|_{L^1(\mathbb{R}^d)} + \sum_{k=1}^{d} \sup_{h \in [-1,1]} |h|^{-\lambda/a_k} \|\Delta_{he_k} f\|_{L^1(\mathbb{R}^d)}, \tag{3.2}$$

where $\Delta_h f(x) = f(x+h) - f(x)$, $h \in \mathbb{R}^d$, and $e_k \in \mathbb{R}^d$ denotes the canonical basis vector in the $k$-the direction (see [7] and [29] for additional details and references). In the above definition, $\lambda/a_k$ describes the smoothness in the $k$-th coordinate, its restriction to $(0,1)$ is not essential. Without this restriction we should use iterated differences in (3.2) instead (see [29, Theorem 5.8(ii)]). Note that, for $a_1 = \cdots = a_d = 1$, we recover the classical Besov space in which case we write $B^{\lambda,1}_{1,\infty}(\mathbb{R}^d) = B^{\lambda}_{1,\infty}(\mathbb{R}^d)$. Moreover, one clearly has $B^{\lambda,a}_{1,\infty}(\mathbb{R}^d) \subset B^{\lambda'}_{1,\infty}(\mathbb{R}^d)$ with $\lambda' = \lambda/\max\{a_1,\ldots,a_d\}$.

### 3.2. *Discrete integration by parts*

Existence of a density for solutions to (1.1) is, in the isotropic case, essentially based on a discrete integration by parts lemma formulated for the difference operator $\Delta_h$ acting on the classical Hölder–Zygmund space (see [8, Lemma 2.1]). Such density belongs, by construction, to an isotropic Besov space. In this work we use an anisotropic version of this lemma which is designed for Lévy processes satisfying condition (A1).

The anisotropic Hölder–Zygmund space $C_b^{\lambda,a}(\mathbb{R}^d)$ is defined as the Banach space of functions $\phi$ with finite norm

$$\|\phi\|_{C_b^{\lambda,a}} = \|\phi\|_\infty + \sum_{k=1}^d \sup_{h\in[-1,1]} |h|^{-\lambda/a_k} \|\Delta_{he_k}\phi\|_\infty. \tag{3.3}$$

The following is our main technical tool for the existence of a density.

**Lemma 3.1.** *Let $a = (a_1,\ldots,a_d)$ be an anisotropy in the sense of (3.1) and $\lambda, \eta > 0$ be such that $(\lambda+\eta)/a_k \in (0,1)$ for all $k = 1,\ldots,d$. Suppose that $q$ is a finite measure over $\mathbb{R}^d$ and there exists $A > 0$ such that, for all $\phi \in C_b^{\eta,a}(\mathbb{R}^d)$ and all $k = 1,\ldots,d$,*

$$\left|\int_{\mathbb{R}^d} \big(\phi(x+he_k) - \phi(x)\big) q(dx)\right| \le A\|\phi\|_{C_b^{\eta,a}} |h|^{(\lambda+\eta)/a_k}, \quad \forall h \in [-1,1]. \tag{3.4}$$

*Then $q(dx)$ has a density $g$ with respect to the Lebesgue measure and*

$$\|g\|_{B^{\lambda,a}_{1,\infty}} \le \mu\big(\mathbb{R}^d\big) + 3dA(2d)^{\eta/\lambda}\left(1 + \frac{\lambda}{\eta}\right)^{1+\frac{\eta}{\lambda}}. \tag{3.5}$$

**Proof.** Given an anisotropy $a$ as above, define the corresponding anisotropic (maximum-)norm on $\mathbb{R}^d$ by $|x|_a = \max\{|x_1|^{1/a_1},\ldots,|x_d|^{1/a_d}\}$. We show the assertion in 3 steps.

*Step 1.* For $r \in (0,1]$ let $\varphi_r(x) = (2r)^{-d}\mathbb{1}_{\{|x|_a<r\}} = (2r)^{-d}\mathbb{1}_{\{|x_1|<r^{a_1}\}}\cdots\mathbb{1}_{\{|x_d|<r^{a_d}\}}$. Fix $\psi \in L^\infty(\mathbb{R}^d)$. Then for each $j$ and $h \in [-1,1]$ we have

$$\begin{aligned}
\big|(\psi*\varphi_r)(x+he_j) - (\psi*\varphi_r)(x)\big| &\le \|\psi\|_\infty \int_{\mathbb{R}^d} \big|\varphi_r(x+he_j-z) - \varphi_r(x-z)\big|\,dz \\
&\le \frac{\|\psi\|_\infty}{2r^{a_j}} \int_{\mathbb{R}} |\mathbb{1}_{\{|x_j+h-z|<r^{a_j}\}} - \mathbb{1}_{\{|x_j-z|<r^{a_j}\}}|\,dz \\
&\le 2\min\left\{1, \frac{|h|}{r^{a_j}}\right\}\|\psi\|_\infty \le 2r^{-\eta}|h|^{\eta/a_j}\|\psi\|_\infty.
\end{aligned}$$

Since also $\|\psi*\varphi_r\|_\infty \le \|\psi\|_\infty$, we obtain $\|\psi*\varphi_r\|_{C_b^{\eta,a}} \le 3d\|\psi\|_\infty r^{-\eta}$ and hence

$$\begin{aligned}
\left|\int_{\mathbb{R}^d} \psi(x)\big[(q*\varphi_r)(x+he_j) - (q*\varphi_r)(x)\big]dx\right| &= \left|\int_{\mathbb{R}^d} \big[(\psi*\varphi_r)(z-he_j) - (\psi*\varphi_r)(z)\big]q(dz)\right| \\
&\le A\|\psi*\varphi_r\|_{C_b^{\eta,a}}|h|^{(\eta+\lambda)/a_j} \\
&\le 3dA\|\psi\|_\infty r^{-\eta}|h|^{(\eta+\lambda)/a_j}.
\end{aligned}$$

By duality this implies that for all $j = 1,\ldots,d$ and $h \in [-1,1]$,

$$\int_{\mathbb{R}^d} \big|(q*\varphi_r)(x+he_j) - (q*\varphi_r)(x)\big|\,dx \le 3dAr^{-\eta}|h|^{(\eta+\lambda)/a_j}. \tag{3.6}$$

*Step 2.* Let us first suppose that $q$ has a density $g \in C^1(\mathbb{R}^d)$ with $\nabla g \in L^1(\mathbb{R}^d)$. Then we obtain, for any $h \in \mathbb{R}^d$ with $|h|_a \leq 1$,

$$\begin{aligned}\int_{\mathbb{R}^d} \big|g(x+h) - g(x)\big|\,dx &\leq \sum_{k=1}^{d} \int_{\mathbb{R}^d} \big|g(x+h_k e_k) - g(x)\big|\,dx \\ &\leq \sum_{k=1}^{d} \int_{\mathbb{R}^d} \big|(g*\varphi_r)(x+h_k e_k) - (g*\varphi_r)(x)\big|\,dx + 2d\int_{\mathbb{R}^d}\big|(g*\varphi_r)(x) - g(x)\big|\,dx \\ &\leq 3dAr^{-\eta}\sum_{k=1}^{d}|h_k|^{(\eta+\lambda)/a_k} + \frac{2d}{(2r)^d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\big|g(y)-g(x)\big|\mathbb{1}_{\{|x-y|_a<r\}}\,dy\,dx \\ &\leq 3d^2Ar^{-\eta}|h|_a^{\eta+\lambda} + \frac{2d}{(2r)^d}\int_{\{|u|_a<r\}}\int_{\mathbb{R}^d}\big|g(x+u)-g(x)\big|\,dx\,du,\end{aligned}$$

where we have used (3.6). For $t \in (0,1]$, set $I_t := \sup_{|h|_a=t}\int_{\mathbb{R}^d}|g(x+h)-g(x)|\,dx$ and $S_t := \sup_{s\in(0,t]} s^{-\lambda} I_s$. Then, using $\int_{\mathbb{R}^d}|u|_a^\lambda \mathbb{1}_{\{|u|_a<r\}}\,du \leq (2r)^d r^\lambda$ and $S_{|u|_a} \leq S_1$ for $|u|_a \leq 1$, we obtain

$$\begin{aligned}t^{-\lambda}I_t &\leq 3d^2A\left(\frac{t}{r}\right)^\eta + \frac{2d}{(2r)^d}t^{-\lambda}\int_{\mathbb{R}^d}|u|_a^\lambda\mathbb{1}_{\{|u|_a<r\}}S_{|u|_a}\,du \\ &\leq 3d^2A\left(\frac{t}{r}\right)^\eta + 2d\left(\frac{r}{t}\right)^\lambda S_1.\end{aligned}$$

Letting $r = \varepsilon t$ with $0 < \varepsilon^\lambda < (2d)^{-1}$ and then taking the supremum over $t \in (0,1]$ on both sides yield $S_1 \leq 3d^2A\varepsilon^{-\eta} + 2d\varepsilon^\lambda S_1$ and hence $S_1 \leq \frac{3d^2A}{1-2d\varepsilon^\lambda}\frac{1}{\varepsilon^\eta}$. A simple extreme value analysis shows that the right-hand side of the last inequality attains its minimum for $\varepsilon^\lambda = \frac{\eta}{\lambda+\eta}\frac{1}{2d}$, which gives

$$\sup_{|h|_a\leq 1}|h|_a^{-\lambda}\int_{\mathbb{R}^d}\big|g(x+h)-g(x)\big|\,dx = S_1 \leq 3d^2A(2d)^{\eta/\lambda}\left(1+\frac{\lambda}{\eta}\right)^{1+\frac{\eta}{\lambda}}.$$

*Step 3.* For the general case let $q_n = q * G_n$, where $G_n(x) = n^d G(nx)$ and $0 \leq G \in C_c^\infty(\mathbb{R}^d)$ satisfies $\|G\|_{L^1(\mathbb{R}^d)} = 1$ with support in $\{x \in \mathbb{R}^d \,|\, |x| \leq 1\}$. Then $q_n$ has a density $g_n \in C^1(\mathbb{R}^d)$ with $\nabla g_n \in L^1(\mathbb{R}^d)$. Moreover, for any $\phi \in C_b^{\eta,a}(\mathbb{R}^d)$ and $h \in [-1,1]$,

$$\begin{aligned}\left|\int_{\mathbb{R}^d}\big(\phi(x+he_j)-\phi(x)\big)q_n(dx)\right| &= \left|\int_{\mathbb{R}^d}\big[(\phi*G_n)(x+he_j)-(\phi*G_n)(x)\big]q(dx)\right| \\ &\leq A\|\phi*G_n\|_{C_b^{\eta,a}}|h|^{(\lambda+\eta)/a_j} \leq A\|\phi\|_{C_b^{\eta,a}}|h|^{(\lambda+\eta)/a_j}.\end{aligned}$$

Applying step 2 for $g_n$ gives

$$\begin{aligned}\|g_n\|_{B_{1,\infty}^{\lambda,a}} &\leq q\big(\mathbb{R}^d\big) + \frac{1}{d}\sup_{|h|_a\leq 1}|h|_a^{-\lambda}\int_{\mathbb{R}^d}\big|g_n(x+h)-g_n(x)\big|\,dx \\ &\leq q\big(\mathbb{R}^d\big) + 3dA(2d)^{\eta/\lambda}\left(1+\frac{\lambda}{\eta}\right)^{1+\frac{\eta}{\lambda}}. \qquad (3.7)\end{aligned}$$

The particular choice of $G_n$ implies that $g_n$ is also tight, i.e.

$$\sup_{n\in\mathbb{N}}\int_{|x|\geq R} g_n(x)\,dx \longrightarrow 0, \quad R \to \infty.$$

Hence we may apply the Kolmogorov-Riesz compactness criterion for $L^1(\mathbb{R}^d)$ to conclude that $(g_n)_{n\in\mathbb{N}}$ is relatively compact in $L^1(\mathbb{R}^d)$. Let $g \in L^1(\mathbb{R}^d)$ be the limit of a subsequence of $(g_n)_{n\in\mathbb{N}}$. Since also $g_n(x)\,dx \longrightarrow q(dx)$ weakly as $n \to \infty$, we conclude that $q(dx) = g(x)\,dx$. The estimate (3.5) is now a consequence of (3.7) and the Lemma of Fatou. □

Note that the restriction $(\lambda+\eta)/a_k \in (0,1)$, $k=1,\dots,d$, in the above lemma is not essential since we may always replace $\lambda,\eta>0$ by some smaller values which satisfy this condition and (3.4).

## 4. Proof of Theorem 2.1

### 4.1. *Small-time Euler approximation*

Below we provide a small-time Euler approximation of processes $X$ obtained from (1.1). Their proofs follow essentially the arguments given in [8, Lemma 3.1] (for $\gamma\in[1,2]$) and [8, Lemma 3.2] (for $\gamma\in(0,1)$), where we use instead of [8, Lemma 5.2] now Lemma A.1. Hence we omit the proofs of the statements given below. We start with the case $\gamma\in[1,2]$ and define $\kappa=\kappa(\gamma,\delta,\beta,\chi)$ by

$$\kappa=\begin{cases}\min\{1+\frac{\beta\wedge\delta}{\gamma},\frac{1}{\gamma}+\frac{\chi\wedge(\delta/\gamma)}{\gamma}\}, & b\neq 0,\\ \frac{1}{\gamma}+\frac{\chi\wedge(\delta/\gamma)}{\gamma}, & b=0.\end{cases}\tag{4.1}$$

Then we can prove that the following proposition.

**Proposition 4.1.** *Let $Z$ be a Lévy process with Lévy measure $\nu$ and characteristic exponent* (1.2). *Suppose that* (2.3) *holds with $\gamma\in[1,2]$ and condition* (A2) *is satisfied. Let $X$ be given as in* (1.1). *Then*

(a) *For each $\eta\in(0,\delta]$ there exists a constant $C=C(\eta,b,\sigma,Z)>0$ such that*

$$\mathbb{E}\big[\big|X(t)-X(s)\big|^{\eta}\big]\le C(t-s)^{\eta/\gamma},\quad 0\le s\le t\le s+1.$$

(b) *For each $\eta\in(0,1\wedge\delta]$ there exists a constant $C=C(\eta,b,\sigma,Z)>0$ such that*

$$\mathbb{E}\big[\big|X(t)-X^{\varepsilon}(t)\big|^{\eta}\big]\le C\varepsilon^{\eta\kappa},\quad t>0,\varepsilon\in(0,1\wedge t),$$

*where $X^{\varepsilon}(t)$ is defined by*

$$X^{\varepsilon}(t)=U^{\varepsilon}(t)+\sigma\big(X(t-\varepsilon)\big)\big(Z(t)-Z(t-\varepsilon)\big),\qquad U^{\varepsilon}(t)=X(t-\varepsilon)+b\big(X(t-\varepsilon)\big)\varepsilon.$$

The next statement treats the case $\gamma\in(0,1)$ for which we let $\kappa=\kappa(\gamma,\delta,\beta,\chi)$ be given by

$$\kappa=\min\left\{1+\frac{\beta}{\gamma},\frac{1}{\gamma}+\chi,\frac{1}{1-\beta}\right\}.\tag{4.2}$$

Then we can prove the following.

**Proposition 4.2.** *Let $Z$ be a Lévy process with Lévy measure $\nu$ and characteristic exponent* (1.2). *Suppose that* (2.3) *holds for $\gamma\in(0,1)$ and condition* (A2) *is satisfied. Let $X$ be as in* (1.1). *Then*

(a) *For each $\eta\in(0,\delta]$ there exists a constant $C=C(\eta,b,\sigma,Z)>0$ such that*

$$\mathbb{E}\big[\big|X(t)-X(s)\big|^{\eta}\big]\le C(t-s)^{\eta},\quad 0\le s\le t\le s+1.$$

*Moreover, if $\eta\in(0,1)$, then*

$$\mathbb{E}\big[\big|X(t)-X(s)\big|^{\eta}\wedge 1\big]\le C(t-s)^{\eta},\quad 0\le s\le t\le s+1.$$

(b) *For each $t>0$ and $\varepsilon\in(0,1\wedge t)$ there exists an $\mathcal{F}_{t-\varepsilon}$-measurable random variable $U^{\varepsilon}(t)$ such that, setting $X^{\varepsilon}(t)=U^{\varepsilon}(t)+\sigma(X(t-\varepsilon))(Z(t)-Z(t-\varepsilon))$, for any $\eta\in(0,\delta]$ there exists a constant $C=C(\eta,b,\sigma,Z)>0$ with*

$$\mathbb{E}\big[\big|X(t)-X^{\varepsilon}(t)\big|^{\eta}\big]\le C\varepsilon^{\eta\kappa},\quad t>0,\varepsilon\in(0,1\wedge t).$$

Let us stress that the constant $C$ is independent of the particular choice of solution $X$ as well as the parameter $\varepsilon$.

4.2. *Concluding the proof of Theorem* 2.1

Below we use the isotropic Hölder–Zygmund space defined in (3.3) with $a_1 = \cdots = a_d = 1$. To simpify the notation we let $C_b^{\eta,1}(\mathbb{R}^d) = C_b^{\eta}(\mathbb{R}^d)$. Recall that $\kappa$ is defined by (4.1) or (4.2), respectively. Based on the previous approximation we prove the following.

**Proposition 4.3.** *Let $Z$ be a Lévy process with Lévy measure $\nu$ and characteristic exponent* (1.2). *Suppose that* (2.3), (A1) *and* (A2) *are satisfied. Let $X$ be as in* (1.1) *and let $\eta \in (0, 1 \wedge \delta)$. Then there exists a constant $C = C(\eta, b, \sigma, Z) > 0$ and $\varepsilon_0 \in (0, 1 \wedge t)$ such that, for any $\varepsilon \in (0, \varepsilon_0)$, $|h| \le 1$ and $\phi \in C_b^{\eta}(\mathbb{R}^d)$,*

$$\big|\mathbb{E}\big[\big|\sigma^{-1}\big(X(t)\big)\big|^{-1}\Delta_h\phi\big(X(t)\big)\big]\big| \le C\|\phi\|_{C_b^{\eta}}\big(|h|^{\eta}\varepsilon^{\frac{\chi\wedge\delta}{\max\{1,\gamma\}}} + |h|\varepsilon^{-1/\alpha^{\min}} + \varepsilon^{\eta\kappa}\big).$$

*Moreover, the constant $C$ is independent of the particular choice of solution $X$ and of $\varepsilon$.*

**Proof.** For $\varepsilon \in (0, 1 \wedge t)$ let $X^{\varepsilon}(t)$ be the approximation from Proposition 4.1 or Proposition 4.2, respectively. Then $|\mathbb{E}[|\sigma^{-1}(X(t))|^{-1}\Delta_h\phi(X(t))]| \le R_1 + R_2 + R_3$ where

$$\begin{aligned}
R_1 &= \big|\mathbb{E}\big[\Delta_h\phi\big(X(t)\big)\big(\big|\sigma^{-1}\big(X(t)\big)\big|^{-1} - \big|\sigma^{-1}\big(X(t-\varepsilon)\big)\big|^{-1}\big)\big]\big|,\\
R_2 &= \mathbb{E}\big[\big|\Delta_h\phi\big(X(t)\big) - \Delta_h\phi\big(X^{\varepsilon}(t)\big)\big|\big|\sigma^{-1}\big(X(t-\varepsilon)\big)\big|^{-1}\big],\\
R_3 &= \big|\mathbb{E}\big[\big|\sigma^{-1}\big(X(t-\varepsilon)\big)\big|^{-1}\Delta_h\phi\big(X^{\varepsilon}(t)\big)\big]\big|.
\end{aligned}$$

For the first term we use (2.2) to get $||\sigma^{-1}(x)|^{-1} - |\sigma^{-1}(y)|^{-1}| \le |\sigma(x) - \sigma(y)| \le C|x-y|^{\chi\wedge\delta}$ and then Proposition 4.1(a) or Propostion 4.2(a), respectively, to obtain

$$R_1 \le \|\phi\|_{C_b^{\eta}}|h|^{\eta}\mathbb{E}\big[\big|X(t) - X(t-\varepsilon)\big|^{\chi\wedge\delta}|\big] \le C\|\phi\|_{C_b^{\eta}}|h|^{\eta}\varepsilon^{\frac{\chi\wedge\delta}{\max\{1,\gamma\}}}.$$

For $R_2$ we use again (2.2), i.e. $|\sigma^{-1}(x)|^{-1} \le |\sigma(x)| \le C$, to obtain

$$R_2 \le C\|\phi\|_{C_b^{\eta}}\mathbb{E}\big[\big|\sigma^{-1}\big(X(t-\varepsilon)\big)\big|^{-1}\big|X(t) - X^{\varepsilon}(t)\big|^{\eta}\big] \le C\|\phi\|_{C_b^{\eta}}\varepsilon^{\eta\kappa},$$

where in the second inequality we have used $\eta \in (0, 1 \wedge \delta)$ so that Proposition 4.1(b) or Proposition 4.2(b) is applicable. Let us turn to $R_3$. Let $f_t$ be the density given by (A1) and write $X^{\varepsilon}(t) = U^{\varepsilon}(t) + \sigma(X(t-\varepsilon))(Z(t) - Z(t-\varepsilon))$, where $U^{\varepsilon}(t)$ is either given by Proposition 4.1(b) or Proposition 4.2(b), respectively. Then we obtain

$$\begin{aligned}
R_3 &= \left|\mathbb{E}\left[\int_{\mathbb{R}^d}\big|\sigma^{-1}\big(X(t-\varepsilon)\big)\big|^{-1}(\Delta_h\phi)\big(U^{\varepsilon}(t) + \sigma\big(X(t-\varepsilon)\big)z\big)f_{\varepsilon}(z)\,dz\right]\right|\\
&= \left|\mathbb{E}\left[\int_{\mathbb{R}^d}\big|\sigma^{-1}\big(X(t-\varepsilon)\big)\big|^{-1}\phi\big(U^{\varepsilon}(t) + \sigma\big(X(t-\varepsilon)\big)z\big)(\Delta_{-\sigma^{-1}(X(t-\varepsilon))h}f_{\varepsilon})(z)\,dz\right]\right|\\
&\le \|\phi\|_{\infty}\mathbb{E}\left[\big|\sigma^{-1}\big(X(t-\varepsilon)\big)\big|^{-1}\int_{\mathbb{R}^d}\big|(\Delta_{-\sigma^{-1}(X(t-\varepsilon))h}f_{\varepsilon})(z)\big|\,dz\right]\\
&\le C\|\phi\|_{C_b^{\eta}}|h|\varepsilon^{-1/\alpha^{\min}}\mathbb{E}\big[\big|\sigma^{-1}\big(X(t-\varepsilon)\big)\big|^{-1}\big|\sigma^{-1}\big(X(t-\varepsilon)\big)\big|\big]\\
&\le C\|\phi\|_{C_b^{\eta}}|h|\varepsilon^{-1/\alpha^{\min}},
\end{aligned}$$

where we have used (2.1), i.e.

$$\int_{\mathbb{R}^d}\big|f_{\varepsilon}(z+w) - f_{\varepsilon}(z)\big|\,dz \le \sum_{i=1}^{d}\int_{\mathbb{R}^d}\big|f_{\varepsilon}(z+w_ie_i) - f_{\varepsilon}(z)\big|\,dz \le C\sum_{i=1}^{d}\varepsilon^{-1/\alpha_i} \le C\varepsilon^{-1/\alpha^{\min}},$$

for $\varepsilon \in (0, \varepsilon_0)$, $\varepsilon_0 \in (0, 1)$ small enough, $w = -\sigma^{-1}(X(t-\varepsilon))h$ and,

$$\mathbb{E}\big[\big|\sigma^{-1}\big(X(t-\varepsilon)\big)\big|^{-1}\big|\sigma^{-1}\big(X(t-\varepsilon)\big)\big|\big] \le 1. \tag{4.3}$$

Summing up the estimates for $R_0$, $R_1$, $R_2$, $R_3$ yields the assertion. □

Below we provide the proof of Theorem 2.1.

**Proof of Theorem 2.1.** Fix $t>0$. It suffices to show that Lemma 3.1 is applicable to the finite measure

$$q_t(dx)=\left|\sigma^{-1}(x)\right|^{-1}\mu_t(dx), \tag{4.4}$$

where $\mu_t(dx)$ denotes the law of $X(t)$. Using (2.4) or (2.5), respectively, we obtain $\kappa\alpha^{\min}>1$. Choose $c>0$ and $\eta\in(0,1\wedge\delta)$ such that $\frac{1}{\kappa}<c<\alpha^{\min}(1-\eta)$ and define

$$\lambda=\left\{c\frac{\chi\wedge\delta}{\max\{1,\gamma\}},1-\eta-\frac{c}{\alpha^{\min}},\eta(c\kappa-1)\right\}>0.$$

Let $\phi\in C_b^\eta(\mathbb{R}^d)$. By Proposition 4.3 we obtain, for $|h|\le 1$ and $\varepsilon=|h|^c(1\wedge t)\varepsilon_0$,

$$\begin{aligned}\int_{\mathbb{R}^d}\Delta_h\phi(x)q_t(dx)&=\left|\mathbb{E}\left[\left|\sigma^{-1}\big(X(t)\big)\right|^{-1}\Delta_h\phi\big(X(t)\big)\right]\right|\\&\le C\|\phi\|_{C_b^\eta}\left(|h|^\eta\varepsilon^{\frac{\chi\wedge\delta}{\max\{1,\gamma\}}}+|h|\varepsilon^{-1/\alpha^{\min}}+\varepsilon^{\eta\kappa}\right)\\&\le\frac{C\|\phi\|_{C_b^\eta}}{(1\wedge t)^{1/\alpha^{\min}}}\left(|h|^{\eta+c\frac{\chi\wedge\delta}{\max\{1,\gamma\}}}+|h|^{1-c/\alpha^{\min}}+|h|^{c\eta\kappa}\right)\\&\le\frac{C\|\phi\|_{C_b^\eta}}{(1\wedge t)^{1/\alpha^{\min}}}|h|^{(\eta+\lambda)}.\end{aligned}$$

This shows that Lemma 3.1 is applicable for $a_1=\cdots=a_d$, i.e. Theorem 2.1 is proved. □

## 5. Proof of Theorem 2.2

### 5.1. *Small time Euler approximation*

Recall that $\gamma^{\max}=\max\{\gamma_1,\dots,\gamma_d\}$, $\delta^{\min}=\min\{\delta_1,\dots,\delta_d\}$, $\chi^{\min}=\min\{\chi_1,\dots,\chi_d\}$ and $\beta^{\min}=\min\{\beta_1,\dots,\beta_d\}$. Fix $i\in\{1,\dots,d\}$. If $\gamma_i\in[1,2]$ define

$$\kappa_i=\begin{cases}\min\{1+\frac{\beta_i\wedge\delta^{\min}}{\gamma^{\max}},\frac{1}{\gamma_i}+\frac{\chi_i\wedge(\delta^{\min}/\gamma_i)}{\gamma^{\max}}\}, & b_i\neq 0,\\ \frac{1}{\gamma_i}+\frac{\chi_i\wedge(\delta^{\min}/\gamma_i)}{\gamma^{\max}}, & b_i=0,\end{cases}$$

while for $\gamma_i\in(0,1)$ define

$$\kappa_i=\min\left\{1+\frac{\beta_i}{\gamma^{\max}},\frac{1}{\gamma_i}+\frac{\chi_i\wedge(\delta^{\min}/\gamma_i)}{\max\{1,\gamma^{\max}\}},\frac{1}{1-\beta^{\min}}\right\}.$$

We start with the case where $\gamma_i\in[1,2]$ and provide a refined version of Proposition 4.1.

**Proposition 5.1.** *Let $Z$ be a Lévy process with Lévy measure $\nu$ and characteristic exponent* (1.2). *Suppose that* (2.9) *and* (A3) *are satisfied. Let $i\in\{1,\dots,d\}$ be such that $\gamma_i\in[1,2]$, and let $X$ be given as in* (2.8). *Then*

(a) *For each $\eta\in(0,\delta_i]$ there exists a constant $C=C(\eta,b,\sigma,Z)>0$ such that*

$$\mathbb{E}\left[\left|X_i(t)-X_i(s)\right|^\eta\right]\le C(t-s)^{\eta/\gamma_i},\quad 0\le s\le t\le s+1.$$

(b) *For each $\eta\in(0,1\wedge\delta_i]$ there exists a constant $C=C(\eta,b,\sigma,Z)>0$ such that*

$$\mathbb{E}\left[\left|X_i(t)-X_i^\varepsilon(t)\right|^\eta\right]\le C\varepsilon^{\eta\kappa_i},\quad t>0,\varepsilon\in(0,1\wedge t),$$

*where $X_i^\varepsilon(t)$ is defined by*

$$X_i^\varepsilon(t)=U_i^\varepsilon(t)+\sigma_i\big(X(t-\varepsilon)\big)\big(Z_i(t)-Z_i(t-\varepsilon)\big),\qquad U_i^\varepsilon(t)=X_i(t-\varepsilon)+b_i\big(X(t-\varepsilon)\big)\varepsilon.$$

A proof of this statement is given in the Appendix. Below we provide a similar statement for the case $\gamma_i \in (0,1)$.

**Proposition 5.2.** *Let $Z$ be a Lévy process with Lévy measure $\nu$ and characteristic exponent* (1.2). *Suppose that* (2.9) *and* (A3) *are satisfied. Let $i \in \{1,\dots,d\}$ be such that $\gamma_i \in (0,1)$ and let $X$ be as in* (2.8). *Then*

(a) *For each $\eta \in (0,\delta_i]$ there exists a constant $C = C(\eta,b,\sigma,Z) > 0$ such that*

$$\mathbb{E}\big[\big|X_i(t)-X_i(s)\big|^{\eta}\big] \le C(t-s)^{\eta}, \quad 0 \le s \le t \le s+1.$$

*Moreover, if $\eta \in (0,1)$, then*

$$\mathbb{E}\big[\big|X_i(t)-X_i(s)\big|^{\eta} \wedge 1\big] \le C(t-s)^{\eta}, \quad 0 \le s \le t \le s+1.$$

(b) *For each $t > 0$ and $\varepsilon \in (0,1\wedge t)$ there exists a $\mathcal{F}_{t-\varepsilon}$-measurable random variable $U_i^{\varepsilon}(t)$ such that, setting $X_i^{\varepsilon}(t) = U_i^{\varepsilon}(t) + \sigma_i(X(t-\varepsilon))(Z_i(t)-Z_i(t-\varepsilon))$, for any $\eta \in (0,\delta_i]$ there exists a constant $C = C(\eta,b,\sigma,Z) > 0$ with*

$$\mathbb{E}\big[\big|X_i(t)-X_i^{\varepsilon}(t)\big|^{\eta}\big] \le C\varepsilon^{\eta\kappa_i}, \quad t>0, \varepsilon \in (0,1\wedge t).$$

A proof of this statement is given in the Appendix. Let us mention that the constant $C$ is independent of the particular choice of solution $X$ and of $\varepsilon$.

### 5.2. *Concluding the proof of Theorem* 2.2

For $\alpha_1,\dots,\alpha_d$ given as in condition (A1), we define an anisotropy $(a_1,\dots,a_d)$ and mean order of smoothness $\overline{\alpha} > 0$ by

$$\frac{1}{\overline{\alpha}} = \frac{1}{d}\left(\frac{1}{\alpha_1}+\cdots+\frac{1}{\alpha_d}\right), \quad a_i = \frac{\overline{\alpha}}{\alpha_i}, i = 1,\dots,d. \tag{5.1}$$

As in the proof of Theorem 2.1, we proceed to prove a refined version of Proposition 4.3.

**Proposition 5.3.** *Let $Z$ be a Lévy process with Lévy measure $\nu$ and characteristic exponent* (1.2). *Suppose that* (2.9), (A1) *and* (A3) *are satisfied. Let $X$ be as in* (2.8), *take an anisotropy $a = (a_i)_{i\in\{1,\dots,d\}}$ and $\eta \in (0,1)$ with*

$$\frac{\eta}{a_j} \le 1 \wedge \delta^{\min}, \quad j \in \{1,\dots,d\}. \tag{5.2}$$

*Then there exists a constant $C = C(\eta,b,\sigma,Z) > 0$ and $\varepsilon_0 \in (0,1\wedge t)$ such that, for any $\varepsilon \in (0,\varepsilon_0)$, $h \in [-1,1]$, $\phi \in C_b^{\eta,a}(\mathbb{R}^d)$ and $i \in \{1,\dots,d\}$,*

$$\big|\mathbb{E}\big[\big|\sigma^{-1}\big(X(t)\big)\big|^{-1}\Delta_{he_i}\phi\big(X(t)\big)\big]\big| \le C\|\phi\|_{C_b^{\eta,a}}\Big(|h|^{\eta/a_i}\varepsilon^{\frac{\chi^{\min}\wedge\delta^{\min}}{\max\{1,\gamma^{\max}\}}} + |h|\varepsilon^{-1/\alpha_i} + \max_{j\in\{1,\dots,d\}}\varepsilon^{\eta\kappa_j/a_j}\Big).$$

*Moreover the constant $C$ is independent of the particular choice of solution $X$ and of $\varepsilon$.*

**Proof.** For $\varepsilon \in (0,1\wedge t)$ let $X^{\varepsilon}(t)$ be the approximation from Proposition 5.1 or Proposition 5.2, respectively. Then $|\mathbb{E}[|\sigma^{-1}(X(t))|^{-1}\Delta_{he_i}\phi(X(t))]| \le R_1 + R_2 + R_3$ where

$$R_1 = \big|\mathbb{E}\big[\Delta_{he_i}\phi\big(X(t)\big)\big(\big|\sigma\big(X(t)\big)\big|^{-1} - \big|\sigma^{-1}\big(X(t-\varepsilon)\big)\big|^{-1}\big)\big]\big|,$$

$$R_2 = \mathbb{E}\big[\big|\Delta_{he_i}\phi\big(X(t)\big) - \Delta_{he_i}\phi\big(X^{\varepsilon}(t)\big)\big|\big|\sigma^{-1}\big(X(t-\varepsilon)\big)\big|^{-1}\big],$$

$$R_3 = \big|\mathbb{E}\big[\big|\sigma^{-1}\big(X(t-\varepsilon)\big)\big|^{-1}\Delta_{he_i}\phi\big(X^{\varepsilon}(t)\big)\big]\big|.$$

For the first term we use (2.2) to get $||\sigma^{-1}(x)|^{-1} - |\sigma^{-1}(y)|^{-1}| \le |\sigma(x)-\sigma(y)| \le C|x-y|^{\chi^{\min}\wedge\delta^{\min}}$ and then Proposition 5.1(a) or Propostion 5.2(a), respectively, to obtain

$$R_1 \le \|\phi\|_{C_b^{\eta,a}}|h|^{\eta/a_i}\mathbb{E}\big[\big|X(t)-X(t-\varepsilon)\big|^{\chi^{\min}\wedge\delta^{\min}}|\big] \le C\|\phi\|_{C_b^{\eta,a}}|h|^{\eta/a_i}\varepsilon^{\frac{\chi^{\min}\wedge\delta^{\min}}{\max\{1,\gamma^{\max}\}}}.$$

For $R_2$ we use again (2.2), i.e. $|\sigma^{-1}(x)|^{-1} \le |\sigma(x)| \le C$, to obtain

$$R_2 \le C\|\phi\|_{C_b^{\eta,a}} \max_{j\in\{1,\dots,d\}} \mathbb{E}\big[\big|\sigma^{-1}\big(X(t-\varepsilon)\big)\big|^{-1}\big|X_j(t)-X_j^{\varepsilon}(t)\big|^{\eta/a_j}\big] \le C\|\phi\|_{C_b^{\eta,a}} \max_{j\in\{1,\dots,d\}} \varepsilon^{\eta\kappa_j/a_j},$$

where in the last inequality we have used (5.2) so that Proposition 5.1(b) or Proposition 5.2(b) is applicable. Let us turn to $R_3$. Let $f_t$ be the density given by (A1) and write $X^{\varepsilon}(t) = U^{\varepsilon}(t) + \sigma(X(t-\varepsilon))(Z(t)-Z(t-\varepsilon))$, where $U^{\varepsilon}(t)$ is either given by Proposition 5.1(b) or Proposition 5.2(b), respectively. Then we obtain

$$\begin{aligned}
R_3 &= \left|\mathbb{E}\left[\int_{\mathbb{R}^d}\big|\sigma^{-1}\big(X(t-\varepsilon)\big)\big|^{-1}\phi\big(U^{\varepsilon}(t)+\sigma\big(X(t-\varepsilon)\big)z\big)(\Delta_{-h\sigma^{-1}(X(t-\varepsilon))e_i}f_{\varepsilon})(z)\,dz\right]\right| \\
&\le \|\phi\|_{\infty}\mathbb{E}\left[\big|\sigma^{-1}\big(X(t-\varepsilon)\big)\big|^{-1}\int_{\mathbb{R}^d}\big|(\Delta_{-h\sigma^{-1}(X(t-\varepsilon))e_i}f_{\varepsilon})(z)\big|\,dz\right] \\
&\le C\|\phi\|_{C_b^{\eta,a}}|h|\varepsilon^{-1/\alpha_i}\mathbb{E}\big[\big|\sigma^{-1}\big(X(t-\varepsilon)\big)\big|^{-1}\big|\sigma^{-1}\big(X(t-\varepsilon)\big)e_i\big|\big] \\
&\le C\|\phi\|_{C_b^{\eta,a}}|h|\varepsilon^{-1/\alpha_i},
\end{aligned}$$

where we have used (2.1) for $\varepsilon \in (0,\varepsilon_0)$, $\varepsilon_0 \in (0,1)$ small enough and (4.3). Summing up the estimates for $R_0$, $R_1$, $R_2$, $R_3$ yields the assertion. □

Finally we give the proof of Theorem 2.2.

**Proof of Theorem 2.2.** Fix $t>0$. It suffices to show that Lemma 3.1 is applicable to the finite measure $q_t$ given by (4.4). For this purpose we use a refinement of the arguments given in in the proof of Theorem 2.1 applied to the anisotropy $a=(a_1,\dots,a_d)$ and mean order of smoothness defined in (5.1). Using (2.10) or (2.11), respectively, we obtain $\kappa_j\alpha_j > 1$ and hence $\kappa_j/a_j > 1/\overline{\alpha}$ for all $j\in\{1,\dots,d\}$. This implies

$$\frac{a_j}{\kappa_j}\frac{1}{a_i} < \frac{\overline{\alpha}}{a_i} = \alpha_i, \quad i,j\in\{1,\dots,d\}.$$

Hence we find $\eta\in(0,1)$ and $c_1,\dots,c_d>0$ such that, for all $i,j\in\{1,\dots,d\}$,

$$0 < \frac{\eta}{a_i} < 1\wedge\delta^{\min}, \qquad \frac{a_j}{\kappa_j}\frac{1}{a_i} < c_i < \alpha_i\left(1-\frac{\eta}{a_i}\right).$$

Define

$$\lambda = \min_{i,j\in\{1,\dots,d\}}\left\{c_i\frac{\chi^{\min}\wedge\delta^{\min}}{\max\{1,\gamma^{\max}\}}a_i, a_i-\eta-\frac{a_ic_i}{\alpha_i}, \eta\left(c_ia_i\frac{\kappa_j}{a_j}-1\right)\right\} > 0.$$

Let $\phi\in C_b^{\eta,a}(\mathbb{R}^d)$. By Proposition 4.3 we obtain, for $h\in[-1,1]$, $\varepsilon=|h|^{c_i}(1\wedge t)\varepsilon_0$ and $i\in\{1,\dots,d\}$,

$$\begin{aligned}
&\big|\mathbb{E}\big[\big|\sigma^{-1}\big(X(t)\big)\big|^{-1}\Delta_{he_i}\phi\big(X(t)\big)\big]\big| \\
&\le C\|\phi\|_{C_b^{\eta,a}}\left(|h|^{\eta/a_i}\varepsilon^{\frac{\chi^{\min}\wedge\delta^{\min}}{\max\{1,\gamma^{\max}\}}} + |h|\varepsilon^{-1/\alpha_i} + \max_{j\in\{1,\dots,d\}}\varepsilon^{\eta\kappa_j/a_j}\right) \\
&\le \frac{C\|\phi\|_{C_b^{\eta,a}}}{(1\wedge t)^{1/\alpha_i}}\left(|h|^{\eta/a_i+c_i\frac{\chi^{\min}\wedge\delta^{\min}}{\max\{1,\gamma^{\max}\}}} + |h|^{1-c_i/\alpha_i} + \max_{j\in\{1,\dots,d\}}|h|^{c_i\eta\kappa_j/a_j}\right) \\
&\le \frac{C\|\phi\|_{C_b^{\eta,a}}}{(1\wedge t)^{1/\alpha_i}}|h|^{(\eta+\lambda)/a_i}.
\end{aligned}$$

This shows that Lemma 3.1 is applicable. □

By inspection of the proof we obtained the following simple observation.

**Remark 5.1.** Suppose that, in addition to the conditions of Theorem 2.2, $\sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_d)$ is uniformly non-degenerate. Then above arguments can also be applied to the measure $q_t(dx) = \mu_t(dx)$. In this case one needs to adapt the proof of Proposition 5.3 by showing that

$$\left|\mathbb{E}\left[\Delta_{he_i}\phi\left(X(t)\right)\right]\right| \le C\|\phi\|_{C_b^{\eta,a}}\left(|h|\varepsilon^{-1/\alpha_i} + \max_{j\in\{1,\ldots,d\}} \varepsilon^{\eta\kappa_j/a_j}\right).$$

## Appendix A: Small time estimates on stochastic integrals with respect to Lévy processes

Below we state some useful estimates on stochastic integrals with respect to Lévy processes. Similar results were obtained in [8, Lemma 5.2].

**Lemma A.1.** *Let $Z$ be a Lévy process with Lévy measure $\nu$ and characteristic exponent given by (1.2). Suppose that there exist $\gamma \in (0, 2]$ and $\delta \in (0, \gamma]$ such that*

$$\int_{\mathbb{R}^d} \left(\mathbb{1}_{\{|z|\le 1\}}|z|^{\gamma} + \mathbb{1}_{\{|z|>1\}}|z|^{\delta}\right)\nu(dz) < \infty.$$

*Then the following assertions hold.*

(a) *Let $0 < \eta \le \delta \le \gamma$ and $1 \le \gamma \le 2$. Then there exists a constant $C > 0$ such that, for any predictable process $H(u)$ and $0 \le s \le t \le s + 1$,*

$$\mathbb{E}\left[\left|\int_s^t H(u)\,dZ(u)\right|^{\eta}\right] \le C(t-s)^{\eta/\gamma} \sup_{u\in[s,t]} \left(\mathbb{E}\left[\left|H(u)\right|^{\gamma}\right]\right)^{\eta/\gamma}.$$

(b) *Let $0 < \eta \le \delta \le \gamma < 1$ and define $Y(t) := Z(t) + t\int_{|z|\le 1} z\nu(dz)$. Then there exists a constant $C > 0$ such that, for any predictable process $H(u)$ and $0 \le s \le t \le s + 1$,*

$$\mathbb{E}\left[\left|\int_s^t H(u)\,dY(u)\right|^{\eta}\right] \le C(t-s)^{\eta/\gamma} \sup_{u\in[s,t]} \left(\mathbb{E}\left[\left|H(u)\right|^{\gamma}\right]\right)^{\eta/\gamma}.$$

*Moreover, setting $\Delta Y(u) = \lim_{r\nearrow u}(Y(u) - Y(r))$ and $\Delta Y(0) = 0$, we obtain*

$$\mathbb{E}\left[\left(\sum_{u\in[s,t]} \left|\Delta Y(u)\right|\right)^{\eta}\right] \le C(t-s)^{\eta/\gamma}.$$

**Proof.** (a) Let $N(du, dz)$ be a Poisson random measure with compensator $\widehat{N}(du, dz) = du\,m(dz)$ such that

$$Z(t) = \int_0^t \int_{|z|\le 1} z\widetilde{N}(du, dz) + \int_0^t \int_{|z|>1} zN(du, dz),$$

where $\widetilde{N}(du, dz) = N(du, dz) - \widehat{N}(du, dz)$ denotes the corresponding compensated Poisson random measure. Then

$$\mathbb{E}\left[\left|\int_s^t H(u)\,dZ(u)\right|^{\eta}\right] \le C\mathbb{E}\left[\left|\int_s^t \int_{|z|\le 1} H(u)z\widetilde{N}(du, dz)\right|^{\eta}\right] + C\mathbb{E}\left[\left|\int_s^t \int_{|z|>1} H(u)zN(du, dz)\right|^{\eta}\right].$$

If $\eta \ge 1$, then by the BDG-inequality, sub-additivity of $x \longmapsto x^{\frac{\gamma}{2}}$ and Hölder inequality we obtain

$$\begin{aligned}\mathbb{E}\left[\left|\int_s^t \int_{|z|\le 1} H(u)z\widetilde{N}(du, dz)\right|^{\eta}\right] &\le C\mathbb{E}\left[\left|\int_s^t \int_{|z|\le 1} \left|H(u)z\right|^2 N(du, dz)\right|^{\eta/2}\right] \\ &\le C\mathbb{E}\left[\left|\int_s^t \int_{|z|\le 1} \left|H(u)z\right|^{\gamma} N(du, dz)\right|^{\eta/\gamma}\right]\end{aligned}$$

$$\le C\left(\mathbb{E}\left[\int_s^t \int_{|z|\le 1} \left|H(u)z\right|^{\gamma} du\nu(dz)\right]\right)^{\eta/\gamma}$$
$$\le C(t-s)^{\frac{\eta}{\gamma}}\left(\int_{|z|\le 1} |z|^{\gamma}\nu(dz)\right)^{\eta/\gamma} \sup_{u\in[s,t]} \left(\mathbb{E}\left[\left|H(u)\right|^{\gamma}\right]\right)^{\eta/\gamma}.$$

If $0<\eta\le 1\le\gamma\le 2$, then the Hölder inequality and previous estimates imply

$$\mathbb{E}\left[\left|\int_s^t \int_{|z|\le 1} H(u)z\widetilde{N}(du,dz)\right|^{\eta}\right] \le \left(\mathbb{E}\left[\left|\int_s^t \int_{|z|\le 1} H(u)z\widetilde{N}(du,dz)\right|\right]\right)^{\eta}$$
$$\le C(t-s)^{\frac{\eta}{\gamma}} \sup_{u\in[s,t]} \left(\mathbb{E}\left[\left|H(u)\right|^{\gamma}\right]\right)^{\eta/\gamma}.$$

Let us turn to the integral involving the big jumps. If $\delta\in(0,1]$, then by sub-additivity of $x\longmapsto x^{\delta}$ and Hölder inequality we get

$$\mathbb{E}\left[\left|\int_s^t \int_{|z|>1} H(u)zN(du,dz)\right|^{\eta}\right] \le \mathbb{E}\left[\left|\int_s^t \int_{|z|>1} \left|H(u)z\right|^{\delta} N(du,dz)\right|^{\eta/\delta}\right]$$
$$\le \left(\mathbb{E}\left[\int_s^t \int_{|z|>1} \left|H(u)z\right|^{\delta} du\nu(dz)\right]\right)^{\eta/\delta} \le (t-s)^{\frac{\eta}{\delta}} \sup_{u\in[s,t]} \left(\mathbb{E}\left[\left|H(u)\right|^{\delta}\right]\right)^{\eta/\delta}.$$

If $\delta\in(1,\gamma]$, then

$$\mathbb{E}\left[\left|\int_s^t \int_{|z|>1} H(u)zN(du,dz)\right|^{\eta}\right] \le C\mathbb{E}\left[\left|\int_s^t \int_{|z|>1} H(u)z\widetilde{N}(du,dz)\right|^{\eta}\right] + C\mathbb{E}\left[\left|\int_s^t \int_{|z|>1} H(u)z\,du\nu(dz)\right|^{\eta}\right].$$

The stochastic integral can be estimated similarly as before, which gives

$$\mathbb{E}\left[\left|\int_s^t \int_{|z|>1} H(u)z\widetilde{N}(du,dz)\right|^{\eta}\right] \le C(t-s)^{\eta/\delta} \sup_{u\in[s,t]} \left(\mathbb{E}\left[\left|H(u)\right|^{\delta}\right]\right)^{\eta/\delta}.$$

The second integral can be estimated by

$$\mathbb{E}\left[\left|\int_s^t \int_{|z|>1} H(u)z\,du\nu(dz)\right|^{\eta}\right] \le \left(\mathbb{E}\left[\left|\int_s^t \int_{|z|>1} H(u)z\,du\nu(dz)\right|^{\delta}\right]\right)^{\eta/\delta} \le (t-s)^{\eta} \sup_{u\in[s,t]} \left(\mathbb{E}\left[\left|H(u)\right|^{\delta}\right]\right)^{\eta/\delta}.$$

Collecting all estimates and using $t-s\le 1$, so that $(t-s)^{\eta/\delta}\le(t-s)^{\eta/\gamma}$ and $(t-s)^{\eta}\le(t-s)^{\eta/\gamma}$, gives

$$\mathbb{E}\left[\left|\int_s^t H(u)\,dZ(u)\right|^{\eta}\right] \le C(t-s)^{\frac{\eta}{\gamma}}\left(\sup_{u\in[s,t]} \left(\mathbb{E}\left[\left|H(u)\right|^{\gamma}\right]\right)^{\eta/\gamma} + \sup_{u\in[s,t]} \left(\mathbb{E}\left[\left|H(u)\right|^{\delta}\right]\right)^{\eta/\delta}\right).$$

Applying Hölder inequality with $p=\gamma/\delta$ and $q=\gamma/(\gamma-\delta)$ gives the assertion.

(b) Consider the decomposition

$$Z(t)=\int_0^t \int_{|z|\le 1} zN(du,dz)+\int_0^t \int_{|z|>1} zN(du,dz)-(t-s)A,$$

where $A=\int_{|z|\le 1} z\nu(dz)$. Then we obtain

$$\mathbb{E}\left[\left|\int_s^t H(u)\,dZ(u)\right|^{\eta}\right] \le \mathbb{E}\left[\left|\int_s^t \int_{|z|\le 1} H(u)zN(du,dz)\right|^{\eta}\right] + \mathbb{E}\left[\left|\int_s^t \int_{|z|>1} H(u)zN(du,dz)\right|^{\eta}\right]$$
$$+\mathbb{E}\left[\left|\int_s^t H(u)A\,du\right|^{\eta}\right]$$
$$\le C(t-s)^{\eta/\gamma} \sup_{u\in[s,t]} \left(\mathbb{E}\left[\left|H(u)\right|^{\gamma}\right]\right)^{\eta/\gamma} + C(t-s)^{\eta/\delta} \sup_{u\in[s,t]} \left(\mathbb{E}\left[\left|H(u)\right|^{\delta}\right]\right)^{\eta/\delta}$$

$$+C(t-s)^{\eta}\sup_{u\in[s,t]}\mathbb{E}\big[\big|H(u)\big|^{\eta}\big]$$
$$\le C(t-s)^{\eta/\gamma}\sup_{u\in[s,t]}\big(\mathbb{E}\big[\big|H(u)\big|^{\gamma}\big]\big)^{\eta/\gamma},$$

where we have used the same estimates as in part (a). □

## Appendix B: Proof of Proposition 5.1

**Proof.** (a) Write

$$\mathbb{E}\big[\big|X_i(t)-X_i(s)\big|^{\eta}\big]\le C\mathbb{E}\bigg[\bigg|\int_s^t b_i\big(X(u)\big)\,du\bigg|^{\eta}\bigg]+C\mathbb{E}\bigg[\bigg|\int_s^t \sigma_i\big(X(u)\big)\,dZ_i(u)\bigg|^{\eta}\bigg]=:R_1+R_2.$$

Using the boundedness of $b_i$ and $\eta\le\gamma_i$ yields for the first term $R_1\le C(t-s)^{\eta}\le C(t-s)^{\eta/\gamma_i}$. For the second term we apply Lemma A.1(a) and the boundedness of $\sigma_i$ to obtain

$$R_2\le C(t-s)^{\eta/\gamma_i}\sup_{u\in[s,t]}\mathbb{E}\big[\big|\sigma_i\big(X(u)\big)\big|^{\gamma_i}\big]^{\eta/\gamma_i}\le C(t-s)^{\eta/\gamma_i}.$$

This proves the assertion.

(b) Write $\mathbb{E}[|X_i(t)-X_i^{\varepsilon}(t)|^{\eta}]\le R_1+R_2$, where

$$R_1=\mathbb{E}\bigg[\bigg|\int_{t-\varepsilon}^t \big(b_i\big(X(u)\big)-b_i\big(X(t-\varepsilon)\big)\big)\,du\bigg|^{\eta}\bigg],$$
$$R_2=\mathbb{E}\bigg[\bigg|\int_{t-\varepsilon}^t \big(\sigma_i\big(X(t-)\big)-\sigma_i\big(X(t-\varepsilon)\big)\big)\,dZ_i(u)\bigg|^{\eta}\bigg].$$

For the first term we use $|b_i(x)-b_i(y)|\le C|x-y|^{\beta_i\wedge\delta^{\min}}$ and part (a) to obtain

$$R_1\le\mathbb{E}\bigg[\int_{t-\varepsilon}^t \big|b_i\big(X(u)\big)-b_i\big(X(t-\varepsilon)\big)\big|\,du\bigg]^{\eta}$$
$$\le C\varepsilon^{\eta}\sup_{u\in[t-\varepsilon,t]}\big(\mathbb{E}\big[\big|X(u)-X(t-\varepsilon)\big|^{\beta_i\wedge\delta^{\min}}\big]\big)^{\eta}$$
$$\le C\varepsilon^{\eta}\sum_{j=1}^d\sup_{u\in[t-\varepsilon,t]}\big(\mathbb{E}\big[\big|X_j(u)-X_j(t-\varepsilon)\big|^{\beta_i\wedge\delta^{\min}}\big]\big)^{\eta}\le C\varepsilon^{\eta+\frac{\eta(\beta_i\wedge\delta^{\min})}{\gamma^{\max}}}.$$

For the second term we use Lemma A.1(a), $|\sigma_i(x)-\sigma_i(y)|\le C|x-y|^{\chi_i\wedge(\delta^{\min}/\gamma_i)}$ and then part *(a)* to obtain

$$R_2\le C\varepsilon^{\eta/\gamma_i}\sup_{u\in[t-\varepsilon,t]}\big(\mathbb{E}\big[\big|\sigma_i\big(X(u)\big)-\sigma_i\big(X(t-\varepsilon)\big)\big|^{\gamma_i}\big]\big)^{\eta/\gamma_i}$$
$$\le C\varepsilon^{\eta/\gamma_i}\sup_{u\in[t-\varepsilon,t]}\big(\mathbb{E}\big[\big|X(u)-X(t-\varepsilon)\big|^{\gamma_i(\chi_i\wedge(\delta^{\min}/\gamma_i))}\big]\big)^{\eta/\gamma_i}$$
$$\le C\varepsilon^{\eta/\gamma_i}\sum_{j=1}^d\sup_{u\in[t-\varepsilon,t]}\big(\mathbb{E}\big[\big|X_j(u)-X_j(t-\varepsilon)\big|^{\gamma_i(\chi_i\wedge(\delta^{\min}/\gamma_i))}\big]\big)^{\eta/\gamma_i}\le C\varepsilon^{\frac{\eta}{\gamma_i}+\frac{\eta}{\gamma^{\max}}\chi_i\wedge(\delta^{\min}/\gamma_i)}.$$

This proves the assertion. □

## Appendix C: Proof of Proposition 5.2

**Proof.** (a) This can be shown exactly in the same way as Proposition 5.1(a).

(b) Define $Y_i(t) = Z_i(t) - t\int_{\mathbb{R}^d} \mathbb{1}_{\{|z|\le 1\}} z_i \nu(dz)$ and

$$\widetilde{b}_j(x) = \begin{cases} b_j(x), & \gamma_j \in [1,2], \\ b_j(x) - \sigma_j(x)\int_{\mathbb{R}^d} \mathbb{1}_{\{|z|\le 1\}} z_j \nu(dz), & \gamma_j \in (0,1), \end{cases} \quad j \in \{1,\dots,d\}.$$

Set $\tau = \varepsilon^{1/(1-\beta^{\min})}$ and define, for $s \in [t-\varepsilon, t]$, $s_\tau = t - \varepsilon + \tau\lfloor (s-(t-\varepsilon))/\tau \rfloor$. Let $W^\varepsilon$ be the solution to

$$W^\varepsilon(u) = X(t-\varepsilon) + \int_{t-\varepsilon}^u \widetilde{b}\big(W^\varepsilon(s_\tau)\big)\,ds, \quad u \in [t-\varepsilon, t].$$

Then $W^\varepsilon(t)$ is well-defined and $\mathcal{F}_{t-\varepsilon}$-measurable, because it is a deterministic continuous function of $X(t-\varepsilon)$. Define

$$\begin{aligned} X_i^\varepsilon(t) &= U_i^\varepsilon(t) + \sigma_i\big(X(t-\varepsilon)\big)\big(Z_i(t) - Z_i(t-\varepsilon)\big), \\ U_i^\varepsilon(t) &= W_i^\varepsilon(t) + \varepsilon\sigma_i\big(X(t-\varepsilon)\big)\int_{\mathbb{R}^d} \mathbb{1}_{\{|z|\le 1\}} z_i \nu(dz). \end{aligned} \tag{C.1}$$

It remains to prove the desired estimate. Thus write

$$X_i^\varepsilon(t) = X_i(t-\varepsilon) + \int_{t-\varepsilon}^t \widetilde{b}_i\big(W^\varepsilon(s)\big)\,ds + \int_{t-\varepsilon}^t \sigma_i\big(X(t-\varepsilon)\big)\,dY_i(s) + \int_{t-\varepsilon}^t \big(\widetilde{b}_i\big(W^\varepsilon(s_\tau)\big) - \widetilde{b}_i\big(W^\varepsilon(s)\big)\big)\,ds,$$

so that $\mathbb{E}[|X_i(t) - X_i^\varepsilon(t)|^\eta] \le \mathbb{E}[I_\varepsilon^\eta] + \mathbb{E}[J_\varepsilon^\eta] + \mathbb{E}[J_\varepsilon^\eta]$ with

$$\begin{aligned} I_\varepsilon &= \int_{t-\varepsilon}^t \big|\widetilde{b}_i\big(X(s)\big) - \widetilde{b}_i\big(W^\varepsilon(s)\big)\big|\,ds, \\ J_\varepsilon &= \bigg|\int_{t-\varepsilon}^t \big(\sigma_i\big(X(s-)\big) - \sigma_i\big(X(t-\varepsilon)\big)\big)\,dY_i(s)\bigg|, \\ K_\varepsilon &= \int_{t-\varepsilon}^t \big|\widetilde{b}_i\big(W^\varepsilon(s_\tau)\big) - \widetilde{b}_i\big(W^\varepsilon(s)\big)\big|\,ds. \end{aligned}$$

For the second term we apply Lemma A.1(b), the $\chi_i \wedge (\delta^{\min}/\gamma_i)$-Hölder continuity of $\sigma$ and part *(a)*, to obtain

$$\begin{aligned} \mathbb{E}\big[J_\varepsilon^\eta\big] &\le C\varepsilon^{\eta/\gamma_i} \sup_{s\in[t-\varepsilon,t]} \big(\mathbb{E}\big[\big|\sigma_i\big(X(s)\big) - \sigma_i\big(X(t-\varepsilon)\big)\big|^{\gamma_i}\big]\big)^{\eta/\gamma_i} \\ &\le C\varepsilon^{\eta/\gamma_i} \sup_{s\in[t-\varepsilon,t]} \big(\mathbb{E}\big[\big|X(s) - X(t-\varepsilon)\big|^{\gamma_i \min\{\chi_i, \delta^{\min}/\gamma_i\}} \wedge 1\big]\big)^{\eta/\gamma_i} \\ &\le C\varepsilon^{\eta/\gamma_i} \sum_{j=1}^d \sup_{s\in[t-\varepsilon,t]} \big(\mathbb{E}\big[\big|X_j(s) - X_j(t-\varepsilon)\big|^{\gamma_i \min\{\chi_i, \delta^{\min}/\gamma_i\}} \wedge 1\big]\big)^{\eta/\gamma_i} \\ &\le C\sum_{j=1}^d \big(\mathbb{1}_{[1,2]}(\gamma_j)\varepsilon^{\frac{\eta}{\gamma_i} + \frac{\eta\min\{\chi_i,\delta^{\min}/\gamma_i\}}{\gamma_j}} + \mathbb{1}_{(0,1)}(\gamma_j)\varepsilon^{\frac{\eta}{\gamma_i} + \min\{\chi_i,\delta^{\min}/\gamma_i\}}\big) \le C\varepsilon^{\eta\kappa_i}. \end{aligned}$$

For the last term we use the $\beta_i$-Hölder continuity of $\widetilde{b}_i$ to obtain

$$\mathbb{E}\big[K_\varepsilon^\eta\big] \le C\mathbb{E}\bigg[\int_{t-\varepsilon}^t \big|W^\varepsilon(s_\tau) - W^\varepsilon(s)\big|^{\beta_i}\,ds\bigg]^\eta \le C\varepsilon\tau^{\eta\beta_i} \le C\varepsilon^{\frac{\eta}{1-\beta^{\min}}},$$

where we have used $|W^\varepsilon(s) - W^\varepsilon(s_\tau)| \le C\tau$ (since $\widetilde{b}$ is bounded), and $\beta_i \ge \beta^{\min}$. Hence it remains to show that

$$\mathbb{E}\big[I_\varepsilon^\eta\big] \le C\big(\varepsilon^{\eta(1+\frac{\beta_i}{\gamma^{\max}})} + \varepsilon^{\frac{\eta}{1-\beta^{\min}}}\big).$$

For this purpose, write

$$W^{\varepsilon}(u) = X(t-\varepsilon) + \int_{t-\varepsilon}^{u} \widetilde{b}\big(W^{\varepsilon}(s)\big)\,ds + \int_{t-\varepsilon}^{u} \big(\widetilde{b}\big(W^{\varepsilon}(s_\tau)\big) - \widetilde{b}\big(W^{\varepsilon}(s)\big)\big)\,ds,$$

and let, for each $j \in \{1, \dots, d\}$,

$$R_j^{\varepsilon}(t) = \begin{cases} |\int_{t-\varepsilon}^{t} \sigma_j(X(s-))\,dY_j(s)|, & \gamma_j \in (0,1), \\ |\int_{t-\varepsilon}^{t} \sigma_j(X(s-))\,dZ_j(s)|, & \gamma_j \in [1,2]. \end{cases}$$

Then, for each $j \in \{1, \dots, d\}$, we obtain

$$\big|X_j(t) - W_j^{\varepsilon}(t)\big| \le \int_{t-\varepsilon}^{t} \big|\widetilde{b}_j\big(W^{\varepsilon}(s_\tau)\big) - \widetilde{b}_j\big(W^{\varepsilon}(s)\big)\big|\,ds + \int_{t-\varepsilon}^{t} \big|\widetilde{b}_j\big(X(s)\big) - \widetilde{b}_j\big(W^{\varepsilon}(s)\big)\big|\,ds + R_j^{\varepsilon}(t).$$

Setting $S_j^{\varepsilon}(t) = \sup_{s\in[t-\varepsilon,t]} |X_j(s) - W_j^{\varepsilon}(s)|$, using the $\beta_j$-Hölder continuity of $\widetilde{b}_j$ and that $|W^{\varepsilon}(s) - W^{\varepsilon}(s_\tau)| \le C\tau$, we obtain

$$\begin{aligned} S_j^{\varepsilon}(t) &\le R_j^{\varepsilon}(t) + C\varepsilon\tau^{\beta_j} + \sum_{k=1}^{d} (C\varepsilon) S_k^{\varepsilon}(t)^{\beta_j} \\ &\le R_j^{\varepsilon}(t) + C\varepsilon^{\frac{1}{1-\beta^{\min}}} + \beta_j \sum_{k=1}^{d} S_k^{\varepsilon}(t), \end{aligned}$$

where we have used $\beta_j \ge \beta^{\min}$, so that $\varepsilon\tau^{\beta_j} \le \varepsilon^{\frac{1}{1-\beta^{\min}}}$, and the Young inequality $xy \le (1-\beta_j)x^{1/(1-\beta_j)} + \beta_j y^{1/\beta_j}$ with $x = C\varepsilon$ and $y = S_k^{\varepsilon}(t)^{\beta_j}$. Taking the sum over $j \in \{1, \dots, d\}$ and using that $\beta_j < 1$, we obtain

$$\sum_{j=1}^{d} S_j^{\varepsilon}(t) \le C \sum_{j=1}^{d} R_j^{\varepsilon}(t) + C\varepsilon^{\frac{1}{1-\beta^{\min}}}.$$

Using this inequality we obtain

$$I_\varepsilon \le C \int_{t-\varepsilon}^{t} \big|X(s) - W^{\varepsilon}(s)\big|^{\beta_i}\,ds \le C\varepsilon \left(\sum_{j=1}^{d} S_j^{\varepsilon}(t)\right)^{\beta_i} \le C\varepsilon \sum_{j=1}^{d} \big(R_j^{\varepsilon}(t)^{\beta_i} + \varepsilon^{\frac{\beta_i}{1-\beta^{\min}}}\big).$$

Using now Lemma A.1(a) we obtain

$$\mathbb{E}\big[I_\varepsilon^{\eta}\big] \le C\varepsilon^{\eta}\left(\varepsilon^{\eta\frac{\beta_i}{1-\beta^{\min}}} + \sum_{j=1}^{d} \mathbb{E}\big[R_j^{\varepsilon}(t)^{\eta\beta_i}\big]\right) \le C\big(\varepsilon^{\frac{\eta}{1-\beta^{\min}}} + \varepsilon^{\eta+\eta\frac{\beta_i}{\gamma^{\max}}}\big).$$

This proves the assertion. □

## Appendix D: Proof of Example 2.2

Before we show that Example 2.2 satisfies condition (A1), we prove the following simple modification of [8, Lemma 3.3] which provides a general estimate on the derivative of the density of a Lévy process.

**Proposition D.1.** *Let $Z$ be a Lévy process with Lévy measure $\nu$ and characteristic exponent* (1.2). *Suppose that*

$$\liminf_{|\xi|\longrightarrow\infty} \frac{\mathrm{Re}(\Psi_\nu(\xi))}{\log(1+|\xi|)} = \infty, \tag{D.1}$$

*and assume that there exists $t_0 > 0$ and $C_1 > 0$ such that*

$$\int_{\mathbb{R}^d} \mathrm{e}^{-t\,\mathrm{Re}(\Psi_\nu(\xi))} |\xi|^{d+2} d\xi \le C_1 \Xi(t)^{2d+2}, \quad t \in (0, t_0), \tag{D.2}$$

*where* $\delta(\eta) = \sup_{|\xi|\leq \eta} \mathrm{Re}(\Psi_\nu(\xi))$ *and* $\Xi(t) = \delta^{-1}(1/t)$. *Then* $Z(t)$ *has a density* $g_t \in C^1(\mathbb{R}^d)$ *such that* $\nabla g_t \in L^1(\mathbb{R}^d)$ *and for some constant* $C_2 > 0$ *and* $t_1 > 0$,

$$\|\nabla g_t\|_{L^1(\mathbb{R}^d)} \leq C_2 \Xi(t \wedge t_1), \quad t > 0. \tag{D.3}$$

**Proof.** From (D.1) it follows that $Z(t)$ has a density $p_t$. For $r > 0$ write $\Psi_\nu = \Psi_{\nu_r} + \Psi_{\nu_r'}$ where $\nu_r'(dz) = \mathbb{1}_{\{|z|>r\}}\nu(dz)$ and $\nu_r(dz) = \mathbb{1}_{\{|z|\leq r\}}\nu(dz)$. Then $p_t = q_t^r * p_t^r$, where $q_t^r$ is the infinite divisible distribution with characteristic exponent $\Psi_{\nu_r'}$ and $p_t^r$ the density with characteristic exponent $\Psi_{\nu_r}$. It follows from [27, Proposition 2.3] that there exist $t_1 > 0$ and $C > 0$ such that for all $t \in (0, t_1]$,

$$\left|\nabla p_t^{1/\Xi(t)}(z)\right| \leq C\,\Xi(t)^{d+1}\big(1 + |z|\Xi(t)\big)^{-d-1}, \quad z \in \mathbb{R}^d,$$

and hence $\|\nabla p_t^{1/\Xi(t)}\|_{L^1(\mathbb{R}^d)} \leq C\,\Xi(t) \int_{\mathbb{R}^d}(1+|z|)^{-d-1}\,dz < \infty$. By Young inequality we obtain for $t \in (0, t_1]$ and some constant $C' > 0$,

$$\|\nabla p_t\|_{L^1(\mathbb{R}^d)} = \left\|q_t^{1/\Xi(t)} * \left(\nabla p_t^{1/\Xi(t)}\right)\right\|_{L^1(\mathbb{R}^d)} \leq \left\|\nabla p_t^{1/\Xi(t)}\right\|_{L^1(\mathbb{R}^d)} \leq C'\,\Xi(t).$$

Now let $t > t_1$, then using the infinite divisibility of $p_t$, we obtain $p_t = p_{t-t_1} * p_{t_1}$ and hence

$$\|\nabla p_t\|_{L^1(\mathbb{R}^d)} = \left\|p_{t-t_1} * (\nabla p_{t_1})\right\|_{L^1(\mathbb{R}^d)} \leq \|\nabla p_{t_1}\|_{L^1(\mathbb{R}^d)} \leq C'\,\Xi(t_1). \qquad \square$$

**Proof that Example 2.2 satisfies condition (A1).** It suffices to show that Proposition D.1 is applicable in $d = 1$ to each $Z_k$ with characteristic exponent $\Psi_k$, $k \in \{1, \dots, d\}$. Fix $k \in \{1, \dots, d\}$ and observe that, by [27, Example 1.4], one has

$$\mathrm{Re}\big(\Psi_k(\xi)\big) \simeq |\xi|^{\alpha_k}\big(\log\big(1 + |\xi|\big)\big)^{\beta_k/2}, \quad \text{as } |\xi| \to \infty.$$

Hence we easily see that (D.1) is satisfied. It follows from the proof of [27, Example 1.4] and [27, Theorem 1.3] that (D.2) is satisfied and, moreover, one has

$$\Xi_k(t) \simeq t^{-\frac{1}{\alpha_k}}\left(\log\left(1 + \frac{1}{t}\right)\right)^{-\frac{\beta_k}{2\alpha_k}}, \quad \text{as } |\xi| \to \infty.$$

If $\beta_k \in [0, 2 - \alpha_k)$, then we may use $\log(1 + t^{-1}) \geq \log(2)$ for $t \in (0, 1]$, to obtain

$$\Xi_k(t) \leq Ct^{-\frac{1}{\alpha_k}}\big(\log(2)\big)^{-\frac{\beta_k}{2\alpha_k}}, \quad \text{as } t \to 0.$$

If $\beta_k \in (-\alpha_k, 0)$, then we may use $\log(1 + t^{-1}) \leq Ct^{-r_k}$, where $r_k = \frac{2\alpha_k}{|\beta_k|}\frac{\varepsilon_k}{\alpha_k(\alpha_k - \varepsilon_k)}$, to obtain

$$\Xi_k(t) \leq Ct^{-\frac{1}{\alpha_k}}\left(\log\left(1 + \frac{1}{t}\right)\right)^{\frac{|\beta_k|}{2\alpha_k}} \leq Ct^{-1/\widetilde{\alpha}_k}, \quad \text{as } t \to 0. \qquad \square$$

## Acknowledgements

The authors would like to thank the anonymous referees for useful remarks and comments. Peng Jin is supported by the STU Scientific Research Foundation for Talents (No. NTF18023).

## References

[1] R. Bass and Z.-Q. Chen. Systems of equations driven by stable processes. *Probab. Theory Related Fields* **134** (2) (2006) 175–214. MR2222382
[2] R. Bass and M. Cranston. The Malliavin calculus for pure jump processes and applications to local time. *Ann. Probab.* **14** (2) (1986) 490–532. MR0832021
[3] K. Bogdan, P. Sztonyk and V. Knopova. Heat kernel of anisotropic nonlocal operators, 2017. Available at arXiv:1704.03705v1 [math.AP].
[4] J. Chaker. The martingale problem for a class of nonlocal operators of diagonal type. *Math. Nachr.* **292** (2019) 2316–2337. MR4033009 https://doi.org/10.1002/mana.201800452
[5] Z.-Q. Chen and X. Zhang. Heat kernels and analyticity of non-symmetric jump diffusion semigroups. *Probab. Theory Related Fields* **165** (1–2) (2016) 267–312. MR3500272

[6] Z.-Q. Chen, X. Zhang and G. Zhao. Well-posedness of supercritical SDE driven by Lev́y processes with irregular drifts, 2017. Available at arXiv:1709.04632v1 [math.PR].
[7] S. Dachkovski. Anisotropic function spaces and related semi-linear hypoelliptic equations. *Math. Nachr.* **248/249** (2003) 40–61. MR1950714
[8] A. Debussche and N. Fournier. Existence of densities for stable-like driven SDE's with Hölder continuous coefficients. *J. Funct. Anal.* **264** (8) (2013) 1757–1778. MR3022725
[9] A. Debussche and M. Romito. Existence of densities for the 3D Navier–Stokes equations driven by Gaussian noise. *Probab. Theory Related Fields* **158** (3–4) (2014) 575–596. MR3176359
[10] N. Fournier. Finiteness of entropy for the homogeneous Boltzmann equation with measure initial condition. *Ann. Appl. Probab.* **25** (2) (2015) 860–897. MR3313757
[11] N. Fournier and J. Printems. Absolute continuity for some one-dimensional processes. *Bernoulli* **16** (2) (2010) 343–360. MR2668905
[12] M. Friesen and P. Jin. On the anisotropic stable JCIR process, 2019. Available at arXiv:1908.05473 [math.PR].
[13] J. Jacod and A. Shiryaev. *Limit Theorems for Stochastic Processes*, 2nd edition. *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **288**. Springer-Verlag, Berlin, 2003. MR1943877
[14] V. Knopova and A. Kulik. Parametrix construction of the transition probability density of the solution to an SDE driven by $\alpha$-stable noise. *Ann. Inst. Henri Poincaré Probab. Stat.* **54** (1) (2018) 100–140. MR3765882
[15] A. Kohatsu-Higa and L. Li. Regularity of the density of a stable-like driven SDE with Hölder continuous coefficients. *Stoch. Anal. Appl.* **34** (6) (2016) 979–1024. MR3544166
[16] T. Kulczycki and M. Ryznar. Transition density estimates for diagonal systems of SDEs driven by cylindrical $\alpha$-stable processes. *ALEA Lat. Am. J. Probab. Math. Stat.* **15** (2) (2018) 1335–1375. MR3877025
[17] T. Kulczycki and M. Ryznar. Semigroup properties of solutions of SDEs driven by Lévy processes with independent coordinates, 2019. Available at arXiv:1906.07173 [math.PR].
[18] A. Kulik. On weak uniqueness and distributional properties of a solution to an SDE with $\alpha$-stable noise. *Stochastic Process. Appl.* **129** (2019) 473–506. MR3907007 https://doi.org/10.1016/j.spa.2018.03.010
[19] O. M. Kulik. Malliavin calculus for Lévy processes with arbitrary Lévy measures. *Teor. Ĭmovīr. Mat. Stat.* **72** (2005) 67–83. MR2168138
[20] S. Orey. On continuity properties of infinitely divisible distribution functions. *Ann. Math. Stat.* **39** (1968) 936–937. MR0226701
[21] J. Picard. On the existence of smooth densities for jump processes. *Probab. Theory Related Fields* **105** (4) (1996) 481–511. MR1402654
[22] E. Priola. Stochastic flow for SDEs with jumps and irregular drift term. In *Stochastic Analysis* 193–210. *Banach Center Publ.* **105**. Polish Acad. Sci. Inst. Math., Warsaw, 2015. MR3445537
[23] M. Romito. Hölder regularity of the densities for the Navier–Stokes equations with noise. *Stoch. Partial Differ. Equ. Anal. Comput.* **4** (3) (2016) 691–711. MR3538013
[24] M. Romito. Time regularity of the densities for the Navier–Stokes equations with noise. *J. Evol. Equ.* **16** (3) (2016) 503–518. MR3551234
[25] M. Romito. A simple method for the existence of a density for stochastic evolutions with rough coefficients. *Electron. J. Probab.* **23** (2018) Paper no. 113, 43. MR3885546
[26] K. Sato. *Lévy Processes and Infinitely Divisible Distributions*. *Cambridge Studies in Advanced Mathematics* **68**. Cambridge University Press, Cambridge, 2013. Translated from the 1990 Japanese original, Revised edition of the 1999 English translation. MR3185174
[27] R. Schilling, P. Sztonyk and J. Wang. Coupling property and gradient estimates of Lévy processes via the symbol. *Bernoulli* **18** (4) (2012) 1128–1149. MR2995789
[28] R. Situ. *Theory of Stochastic Differential Equations with Jumps and Applications, Mathematical and Analytical Techniques with Applications to Engineering*. Springer, New York, 2005. MR2160585
[29] H. Triebel. *Theory of Function Spaces. III*. *Monographs in Mathematics* **100**. Birkhäuser Verlag, Basel, 2006. MR2250142
[30] L. Xie and X. Zhang. Ergodicity of stochastic differential equations with jumps and singular coefficients, 2017. Available at arXiv:1709.04632v1 [math.PR]. MR4058986 https://doi.org/10.1214/19-AIHP959
[31] X. Zhang. Fundamental solutions of nonlocal Hörmander's operators. *Commun. Math. Stat.* **4** (3) (2016) 359–402. MR3554922
[32] X. Zhang. Fundamental solutions of nonlocal Hörmander's operators II. *Ann. Probab.* **45** (3) (2017) 1799–1841. MR3650416