Stéphane Menozzi & Xicheng Zhang

**Heat kernel of supercritical nonlocal operators with unbounded drifts**

# HEAT KERNEL OF SUPERCRITICAL NONLOCAL
# OPERATORS WITH UNBOUNDED DRIFTS

Stéphane Menozzi & Xicheng Zhang

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

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

where $L^{(\alpha)}$ is a $d$-dimensional rotationally invariant $\alpha$-stable process, $b : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ and $a : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ are Hölder continuous functions in space, with respective order $\beta, \gamma \in (0, 1)$ such that $(\beta \wedge \gamma) + \alpha > 1$, uniformly in $t$. Here $b$ may be unbounded. When $a$ is bounded and uniformly elliptic, we show that the unique solution $X_t(x)$ of the above SDE admits a continuous density, which enjoys sharp two-sided estimates. We also establish sharp upper-bound for the logarithmic derivative. In particular, we cover the whole *supercritical* range $\alpha \in (0, 1)$. Our proof is based on ad hoc parametrix expansions and probabilistic techniques.

Résumé (Noyau de la chaleur pour des EDS surcritiques à dérive non bornée)
   Soit $\alpha \in (0, 2)$ et $d \in \mathbb{N}$. Considérons l'équation différentielle stochastique (EDS) suivante dans $\mathbb{R}^d$ :

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

où $L^{(\alpha)}$ est un processus $\alpha$-stable isotrope de dimension $d$, $b : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ et $a : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ sont des fonctions Hölder continues en espace, d'indices respectifs $\beta, \gamma \in (0, 1)$ tels que $(\beta \wedge \gamma) + \alpha > 1$, uniformément en $t$. En particulier $b$ peut être non bornée. Lorsque $a$ est bornée et uniformément elliptique, nous montrons que la solution $X_t(x)$ de l'EDS admet une densité continue, que l'on peut encadrer, à constante multiplicative près, par une même quantité. Nous obtenons également une borne supérieure précise pour la dérivée logarithmique de la densité. En particulier, nous traitons complètement le régime *surcritique* $\alpha \in (0, 1)$. Notre approche se base sur des développements parametrix ad hoc et des techniques probabilistes.

## Contents

## 1. Introduction

Throughout this paper we fix $\alpha \in (0,2)$. Let $L^{(\alpha)}$ be a $d$-dimensional rotationally invariant $\alpha$-stable process. We consider the following stochastic differential equation:

$$\text{(1.1)} \qquad \mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + a(t, X_{t-})\,\mathrm{d}L_t^{(\alpha)},$$

where $b : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ and $a : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$ are Borel measurable functions and satisfy that for some $\beta \in ((1-\alpha)^+, 1]$ and $\kappa_0 \geqslant 1$,

$$(\mathbf{H}_b^\beta) \qquad |b(t,0)| \leqslant \kappa_0,\; |b(t,x) - b(t,y)| \leqslant \kappa_0(|x-y|^\beta \vee |x-y|),$$

and for some $\gamma \in ((1-\alpha)^+, 1]$ and $\kappa_1 \geqslant 1$,

$$(\mathbf{H}_a^\gamma) \qquad \kappa_1^{-1}\mathbb{I} \leqslant (aa^*)(t,x) \leqslant \kappa_1\mathbb{I}, \quad |a(t,x) - a(t,y)| \leqslant \kappa_1|x-y|^\gamma,$$

where $a^*$ stands for the transpose of $a$ and $\mathbb{I}$ is the identity matrix. Under $(\mathbf{H}_b^\beta)$ and $(\mathbf{H}_a^\gamma)$, it is well known that for each $(s,x) \in \mathbb{R}_+ \times \mathbb{R}^d$, there is a unique weak solution $X_{s,t}(x)$ to SDE (1.1) starting from $x$ at time $s$ (see e.g. [12, Th. 1.1]), and the generator of SDE (1.1) writes as

$$\mathscr{L}_s f(x) := \tfrac{1}{2}\mathcal{L}_s f(x) + b(s,x) \cdot \nabla f(x),$$

where $\mathcal{L}_s$ is given by

$$\text{(1.2)} \qquad \mathcal{L}_s f(x) = \int_{\mathbb{R}^d} \delta_f^{(2)}(x; a(s,x)z)\,\frac{\mathrm{d}z}{|z|^{d+\alpha}} = \int_{\mathbb{R}^d} \delta_f^{(2)}(x; z)\,\frac{\kappa(s,x,z)}{|z|^{d+\alpha}}\,\mathrm{d}z$$

with

$$\text{(1.3)} \qquad \delta_f^{(2)}(x;z) := f(x+z) + f(x-z) - 2f(x)$$

and

$$\text{(1.4)} \qquad \kappa(s,x,z) := \det(a^{-1}(s,x))(|z|/|a^{-1}(s,x)z|)^{d+\alpha}.$$

Clearly, by $(\mathbf{H}_a^\gamma)$ we have for some $\overline{\kappa}_1 \geqslant 1$,

$$\text{(1.5)} \qquad \overline{\kappa}_1^{-1} \leqslant \kappa(s,x,z) \leqslant \overline{\kappa}_1, \quad |\kappa(s,x,z) - \kappa(s,y,z)| \leqslant \overline{\kappa}_1|x-y|^\gamma.$$

The operator $\mathscr{L}_s$ is called *supercritical* for $\alpha \in (0,1)$ since in this case, the drift term plays a dominant role. Namely, from the self-similarity properties of the driving process $L^{(\alpha)}$ in (1.1), it holds that for any $s > 0$, $L_s^{(\alpha)} \overset{\text{(law)}}{=} s^{1/\alpha}L_1^{(\alpha)}$ and for $s \in (0,1)$, $\alpha \in (0,1)$, $s^{1/\alpha} < s$. This precisely means that the fluctuations induced by the noise are smaller than the typical order of the drift term in (1.1). For $\alpha \in (1,2)$, the converse phenomenon happens. Since for $s \in (0,1)$, $s^{1/\alpha} > s$, the fluctuations of the noise prevail in the SDE. From the operator viewpoint, $\mathcal{L}_s$ plays a dominant role and we say that $\mathcal{L}_s$ is *subcritical*. For the remaining case $\alpha = 1$, the noise and drift both have the same typical order and the operator $\mathscr{L}_s$ is called *critical*. Note that for $\alpha \in (0,1)$, since $z \mapsto \kappa(s,x,z)$ is symmetric, we have

$$\mathcal{L}_s f(x) = 2 \int_{\mathbb{R}^d} \delta_f^{(1)}(x;z)\,\frac{\kappa(s,x,z)\,\mathrm{d}z}{|z|^{d+\alpha}},$$

where

$$\delta_f^{(1)}(x;z) := f(x+z) - f(x).$$

Let us now indicate that there is a quite large literature concerning stable driven SDEs. We can first mention the seminal work of Kolokoltsov [22] from which one can derive that for an SDE driven by a symmetric stable process with smooth non-degenerate spectral measure, Lipschitz non-degenerate diffusion coefficient and non trivial Lipschitz bounded drifts when $\alpha > 1$, two sided estimates for the density of the type:

$$(1.6) \qquad p(s, x, t, y) \asymp_C (t - s)^{-d/\alpha} \Big( 1 + \frac{|x - y|}{(t - s)^{1/\alpha}} \Big)^{-(d+\alpha)},$$

where $C \geqslant 1$ depends on the non-degeneracy and Lipschitz constants of the coefficients and the final considered time horizon $T$. Here and below, $Q_1 \asymp_C Q_2$ means that $C^{-1} Q_2 \leqslant Q_1 \leqslant C Q_2$.

Going to weaker regularity of the coefficients in (1.1) then first leads to investigate the well-posedness of the martingale problem associated with the formal generator associated with the dynamics (1.1). In [2], Bass and Chen showed the weak well-posedness for SDE (1.1) when $a$ is only continuous and uniformly elliptic, $b$ is Lipschitz and $L_t^{(\alpha)}$ is cylindrical $\alpha$-stable process. In the subcritical case we can mention the work by Mikulevicius and Pragarauskas [35] who derived that weak uniqueness holds for equation (1.1) for bounded Hölder coefficients when $\alpha \geqslant 1$ and a non degenerate $a$. The martingale problem was in their framework studied from some related Schauder estimates established on the associated Integro Partial Differential Equation (IPDE). We can also refer to [6] for parabolic Schauder estimates in the super-critical stable case and to the work by Kühn [24] for interior elliptic Schauder estimates for a larger class of Lévy operators satisfying the Hartman-Wintner growth condition and suitable $L^1$ gradient estimates for the corresponding heat kernel.

In the super-critical case, the well-posedness of the martingale problem was recently investigated by Kulik et al. [19], [28] (see also [12]). In [28], the authors consider SDEs of type (1.1) with bounded Hölder drift and non-degenerate scalar diffusion coefficients under the natural condition $\alpha + \beta > 1$[(1)] and obtain the existence of the heat kernel, a corresponding two-sided estimate of the form (1.6) as well as some estimates corresponding to the time derivative through parametrix type expansions. Let us emphasize that in the super-critical regime, the time derivative of the heat kernel roughly typically behaves as $t^{-1}$ at time $t$ whereas the spatial gradient is then more singular, as it is expected to have typical behavior of order $t^{-1/\alpha} \geqslant t^{-1}$ for $t \in (0, 1]$.

Concerning other results related to stable heat kernel estimates we can refer e.g. for driftless operators of the form (1.2) to [9] [11] in which the authors consider non-symmetric functions $\kappa$ with possible unbounded dependence in the jump variable in the second paper. Therein, two-sided heat kernel and gradient estimates are derived, and for a bounded, even $\kappa := \kappa(x)$ the gradient estimate has been shown to be sharp

---

[(1)]It is indeed well known from the seminal work of Tanaka et al. [36] that weak uniqueness may fail if this condition is not met.

in [15]. We can also quote [8] for similar results when the operator also has a diffusion part, or [10] for more general absolutely continuous Lévy measures satisfying suitable scaling properties.

We insist that, beyond the symmetry and the absolute continuity condition of the Lévy measure with respect to the one of the isotropic stable process, the behavior of the stable densities or stable driven SDEs can be very different from the above description. This can be already seen from the seminal work of Watanabe [38] for stable processes and from the recent work [20] for SDEs, see Section 4 therein, and in particular Example 4.2 which relates to a cylindrically stable driven SDE whose density turns out to be unbounded.

For non zero drifts and rather general, possibly non-symmetric, stable Lévy measures, estimates of the heat kernel in Besov norms were obtained in [13]. We also refer to the work [29] and [7] for gradient estimates on the semigroup associated with additive and multiplicative cylindrical noises in (1.1) respectively, for $\alpha \in (0,2)$ and appropriate assumptions on the drift. Still in the cylindrical case, we can quote the papers by Kulczycki, Ryznar et al. [27] for semigroup estimates in the stable supercritical case, [26], [25] for semigroup and diagonal estimates on the density for more general driving processes whose characteristic exponents satisfy upper and lower scaling conditions.

In the current work, we stick to operators of the form (1.2), but anyhow face two difficulties: we want to establish density and gradient estimates for all $\alpha \in (0,2)$ and for unbounded drifts. It is known, and somehow intuitive, that for unbounded drifts the heat kernel bounds must reflect somehow the transport induced by the drift. This was for instance observed for a Lipschitz drift in [14] for degenerate Kolmogorov SDEs which can be viewed as ODEs perturbed on some components by a Brownian noise propagating through the whole chain thanks to a weak type Hörmander condition on the drift. Before going further let us also mention the work by Huang [17] which establishes two-sided estimates for stable driven SDE with unbounded Lipschitz drift and $\alpha \in (1,2)$, which then read as: for $s < t$ and $x, y \in \mathbb{R}^d$,

$$(1.7) \qquad p(s,x,t,y) \asymp_C (t-s)^{-d/\alpha} \Big( 1 + \frac{|\theta_{s,t}(x) - y|}{(t-s)^{1/\alpha}} \Big)^{-(d+\alpha)},$$

where $\theta_{s,t}(x)$ denotes the flow associated to the drift in (1.1). Namely,

$$\dot{\theta}_{s,t}(x) = b(t, \theta_{s,t}(x)), \theta_{s,s}(x) = x.$$

In the non-degenerate Brownian case, the type of heat kernel estimate in [14] has recently been extended to drifts satisfying a linear-growth without a priori smoothness assumptions on the drift, see [34]. In the quoted work, under additional Hölder continuity of the drift for the second derivatives, the estimates also extend to the derivatives up to order two with the corresponding additional parabolic singularity.

All the previously quoted works use the *parametrix* method. Initiated by E. E. Levi [30], it has become a classical tool for the study of heat kernels of operators with non constant coefficients, see e.g. [16], [33] in the diffusive case or [21] in a much larger

framework. The idea of the method is that, in *small* time, assuming that the heat kernel of the operator with variable coefficients exists, it should *somehow* behave as a well understood one corresponding to an operator with *constant* coefficient. The difference between the two densities is then controlled through a Duhamel type formula. In its usual formulation this approach is designed for operators with spatially bounded coefficients. Assume for a while that, additionally to the stated assumptions, the coefficients in (1.1) are smooth and bounded. It is then known that the density of $(X_t)_{t \geqslant s}$ exists for $t > s$. Denote it by $p(s, x, t, \cdot)$ for the starting point $x$ at time $s$. The *parametrix* is a zero order approximation of the density $p(s, x, t, \cdot)$ to be estimated and for which one precisely knows the behavior, e.g. Gaussian in [16], stable isotropic with possible shift in [22]. The most common choice consists in considering, for a fixed terminal point $y$, as parametrix the density of the process with dynamics

$$(1.8) \qquad \mathrm{d}\widetilde{X}_t^y = b(t, y)\,\mathrm{d}t + a(t, y)\,\mathrm{d}L_t^{(\alpha)}, \quad t \geqslant s, \ \widetilde{X}_s^y = x.$$

From the uniform ellipticity of $a$ and the specific form (1.2) it is possible to derive that $\widetilde{X}_t^y$ admits a density for $t > s$ and that for any $k \in \mathbb{N}$, $w \in \mathbb{R}^d$,

$$(1.9) \quad |\nabla_x^k \widetilde{p}^y(s, x, t, w)| \leqslant C(t-s)^{-(k+d)/\alpha} \left( 1 + \frac{|w - (x + \int_s^t b(r, y)\,\mathrm{d}r)|}{(t-s)^{1/\alpha}} \right)^{-(d+\alpha+k)},$$

see e.g. [3], [4] and Section 2.4 below. The difference between the density of the SDE and the *proxy* one is then investigated through the Duhamel formula which formally follows from the Kolmogorov equations satisfied by the initial density and the *parametrix* in (1.8). Namely, $\widetilde{p}^y(\cdot, \cdot, t, w)$ is a classical solution of:

$$(1.10) \qquad \partial_s \widetilde{p}^y(s, x, t, w) + \widetilde{\mathscr{L}}_s^y \widetilde{p}^y(s, \cdot, t, w)(x) = 0, \ (s, x) \in [0, t) \times \mathbb{R}^d,$$

where

$$\widetilde{\mathscr{L}}_s^y f(x) := \tfrac{1}{2} \widetilde{\mathcal{L}}_s^y f(x) + b(s, y) \cdot \nabla f(x), \ \widetilde{\mathcal{L}}_s^y f(x) = \int_{\mathbb{R}^d} \delta_f^{(2)}(x; z) \frac{\kappa(s, y, z)}{|z|^{d+\alpha}} \,\mathrm{d}z$$

with

$$\kappa(s, y, z) := \frac{\det(a^{-1}(s, y)|z|^{d+\alpha}}{|a^{-1}(s, y)z|^{d+\alpha}}.$$

And $\widetilde{p}^y(s, \cdot, t, w) \underset{s \uparrow t}{\longrightarrow} \delta_w(\cdot)$, weakly identifying the density with the induced measure. Equation (1.10) is the *backward* Kolmogorov equation. If the density of the SDE is now a classical solution of the forward Kolmogorov equation

$$(1.11) \qquad \begin{aligned} &\partial_t p(s, x, t, w) - (\mathscr{L}_t)^* p(s, x, t, \cdot)(w) = 0, \ (t, w) \in (s, +\infty] \times \mathbb{R}^d, \\ &p(s, x, t, w) \underset{t \downarrow s}{\longrightarrow} \delta_x(\cdot), \end{aligned}$$

where $(\mathscr{L}_t)^*$ denotes the adjoint of $\mathscr{L}_t$, it formally follows that

$$
(p - \widetilde{p}^y)(s, x, t, y) = \int_s^t \mathrm{d}r \partial_r \Big( \int_{\mathbb{R}^d} p(s, x, r, w) \widetilde{p}^y(r, w, t, y) \, \mathrm{d}w \Big)
$$

$$
= \int_s^t \mathrm{d}r \int_{\mathbb{R}^d} \partial_r p(s, x, r, w) \widetilde{p}^y(r, w, t, y) \, \mathrm{d}w
$$

(1.12)

$$
+ \int_s^t \mathrm{d}r \int_{\mathbb{R}^d} p(s, x, r, w) \partial_r \widetilde{p}^y(r, w, t, y) \, \mathrm{d}w
$$

$$
\underset{(1.10),(1.11)}{=} \int_s^t \mathrm{d}r \int_{\mathbb{R}^d} p(s, x, r, w)(\mathscr{L}_r - \widetilde{\mathscr{L}_r^y}) \widetilde{p}^y(r, \cdot, t, y)(w) \, \mathrm{d}w,
$$

provided all the previous computations can be justified, which can be delicate for the density of the SDE. The idea is then to repeat the approximation procedure in the integral of (1.12), i.e., approximate $p(s, x, r, w)$ at order 0 with $\widetilde{p}^w(s, x, r, w)$ and control again the difference through the associate integral. Iterating infinitely many times leads to the so-called parametrix series. To obtain some quantitative bounds from this procedure it is clear that one has to understand the behavior of the so-called *parametrix kernel*:

$$
q_0(r, w, t, y) := (\mathscr{L}_r - \widetilde{\mathscr{L}_r^y}) \widetilde{p}^y(r, \cdot, t, y)(w)
$$

$$
= [\tfrac{1}{2}(\widetilde{\mathcal{L}}_r - \widetilde{\mathcal{L}}_r^y) + (b(r, w) - b(r, y)) \cdot \nabla] \widetilde{p}^y(r, \cdot, t, y)(w),
$$

and its smoothing properties. Let us here focus on the drift term. For a bounded Lipschitz drift and $\alpha \in [1, 2)$, it is seen from (1.9) that

$$
|b(r, w) - b(r, y)||\nabla_w \widetilde{p}^y(r, w, t, y)|
$$

(1.13)

$$
\leqslant C|w - y|(t - r)^{-(1+d)/\alpha} \Big( 1 + \frac{|y - (w + \int_r^t b(u, y) \, \mathrm{d}u)|}{(t - r)^{1/\alpha}} \Big)^{-(d+\alpha+1)}
$$

$$
\leqslant \widetilde{C}(t - r)^{-d/\alpha} \Big( 1 + \frac{|y - w|}{(t - r)^{1/\alpha}} \Big)^{-(d+\alpha)},
$$

where for the last inequality we also used that, in small time and for a bounded drift $b$ and $\alpha \geqslant 1$, the frozen drift $\int_r^t b(u, y) \, \mathrm{d}u$ is negligible with respect to to the characteristic time scale, i.e., $|\int_r^t b(u, y) \, \mathrm{d}u| \leqslant \|b\|_\infty (t - r) \leqslant \|b\|_\infty (t - r)^{1/\alpha}$ so that, up to a modification of $C$,

$$
\Big( 1 + \frac{|y - (w + \int_r^t b(u, y) \, \mathrm{d}u)|}{(t - r)^{1/\alpha}} \Big)^{-(d+\alpha+1)} \leqslant C \Big( 1 + \frac{|y - w|}{(t - r)^{1/\alpha}} \Big)^{-(d+\alpha+1)}.
$$

We have thus bounded in (1.13) the drift part of the parametrix kernel by a quantity which serves as upper and lower bound for the isotropic stable density, see again (1.6). Hence, the regularity of the drift allowed to absorb the time singularity induced by the spatial derivative of the *parametrix*. Observe that if $\alpha \in (1, 2)$ and since the drift is bounded, we could even have taken a driftless parametrix. In that case, the contribution associated with the difference of the drifts, between the generator of the

SDE and the one of the frozen process would write:

$$
\begin{aligned}
|b(r,w)||\nabla_w \widetilde{p}^y(r,w,t,y)| &\leqslant C\|b\|_\infty (t-r)^{-(1+d)/\alpha}\Big(1+\frac{|y-w|}{(t-r)^{1/\alpha}}\Big)^{-(d+\alpha+1)} \\
&\leqslant \widetilde{C}(t-r)^{-(d+1)/\alpha}\Big(1+\frac{|y-w|}{(t-r)^{1/\alpha}}\Big)^{-(d+\alpha)},
\end{aligned}
\tag{1.14}
$$

which is homogeneous to an isotropic stable density multiplied by a time integrable singularity. Having in mind that in (1.12) the parametrix kernel is integrated in time, this is sufficient to provide a smoothing effect in small time, i.e., the integral term in (1.12) will be negligible with respect to to the first order approximation, provided we manage to control the series expansion deriving from the iteration of the procedure. This requires to control convolutions of the density of the parametrix and iterated convolutions of the parametrix kernel.

Turning now to the case $\alpha \in (0,1)$, even if $b$ is bounded, the last inequality of (1.13) fails precisely because $t^{1/\alpha} \leqslant t$, $t \leqslant 1$, i.e., the drift prevails with respect to characteristic time of the noise in small time. Thus, freezing at the terminal point does not provide a *good* parametrix. On the other hand, taking a driftless proxy is not an option either, since the associated bound for the drift part of the parametrix kernel would yield a non-integrable singularity in (1.14). We thus need to consider a proxy which can absorb the time singularity of the gradient through the difference of the drift coefficients in the operators. If $b$ is Lipschitz, this can e.g. be done through the backward flow associated with the differential system. Namely, introducing for fixed final point $y \in \mathbb{R}^d$ and final time $t$,

$$
\dot{\theta}_{t,r}(y) = b(r, \theta_{t,r}(y)), \quad r \in [0,t], \ \theta_{t,t}(y) = y,
\tag{1.15}
$$

and considering as parametrix the density of the process with dynamics

$$
\mathrm{d}\widetilde{X}_r^{(t,y)} = b(r, \theta_{t,r}(y))\,\mathrm{d}t + a(r, \theta_{t,r}(y))\,\mathrm{d}L_t^{(\alpha)}, \quad r \geqslant s, \ \widetilde{X}_s^{(t,y)} = x,
\tag{1.16}
$$

one derives similarly to (1.9)

$$
\begin{aligned}
|\nabla_w^k \widetilde{p}^{(t,y)}(r,w,t,y)| &\leqslant C(t-r)^{-(k+d)/\alpha}\Big(1+\frac{|y-(w+\int_r^t b(u,\theta_{t,u}(y))\,\mathrm{d}u)|}{(t-r)^{1/\alpha}}\Big)^{-(d+\alpha+k)} \\
&\leqslant C(t-r)^{-(k+d)/\alpha}\Big(1+\frac{|\theta_{t,r}(y)-w|}{(t-r)^{1/\alpha}}\Big)^{-(d+\alpha+k)},
\end{aligned}
$$

exploiting the differential dynamics (1.15) in backward time for the last inequality. Writing now the difference of the drift parts of the generators applied to the frozen density $\widetilde{p}^{(t,y)}(\cdot,\cdot,t,y)$ yields:

$$
\begin{aligned}
|b(r,w) - b(r,\theta_{t,r}(y))||&\nabla_w \widetilde{p}^y(r,w,t,y)| \\
&\leqslant C|w - \theta_{t,r}(y)|(t-r)^{-(1+d)/\alpha}\Big(1+\frac{|\theta_{t,r}(y)-w|}{(t-r)^{1/\alpha}}\Big)^{-(d+\alpha+1)} \\
&\leqslant \widetilde{C}(t-r)^{-d/\alpha}\Big(1+\frac{|\theta_{t,r}(y)-w|}{(t-r)^{1/\alpha}}\Big)^{-(d+\alpha)},
\end{aligned}
$$

and this choice then allows to absorb the time singularity similarly to the previously considered sub-critical case $\alpha \in [1,2)$, see (1.13). For parametrix expansions and density estimates, this idea of using the backward flow associated with the first order term was already used in the diffusive setting in [23], [14] and for stable driven SDE in [17] for Lipschitz drifts with linear growth in the subcritical case and [19] for bounded Lipschitz drifts in the supercritical case. In this last setting, an underlying flow was also used in [28] where the author addresses the case of a bounded, time-homogeneous Hölder drift coefficient[(2)]. In this framework, there is no uniqueness to (1.15) but only existence (thanks to the Peano theorem). Anyhow, for a $\beta$ Hölder drift, bounded or not, and a given associated Peano flow $\theta_{r,t}(y)$ solving (1.15),

$$|b(r,w) - b(r, \theta_{t,r}(y))||\nabla_w \widetilde{p}^y(r,w,t,y)|$$

$$(1.17) \qquad \leqslant C|w - \theta_{t,r}(y)|^\beta (t-r)^{-(1+d)/\alpha}\Big(1 + \frac{|\theta_{t,r}(y) - w|}{(t-r)^{1/\alpha}}\Big)^{-(d+\alpha+1)}$$

$$\leqslant \widetilde{C}(t-r)^{-(d/\alpha + (1-\beta)/\alpha)}\Big(1 + \frac{|\theta_{t,r}(y) - w|}{(t-r)^{1/\alpha}}\Big)^{-(d+\alpha)},$$

which yields an integrable singularity provided $\alpha + \beta > 1$ (which is the natural condition we assumed and already discussed above). This type of estimate is crucial for the parametrix approach to work in our framework.

In the following, in order to benefit from Lipschitz properties of an associate flow, which is e.g. needed to make convolutions of terms which have the form of the r.h.s. of (1.17) and which will naturally appear from the parametrix expansion, we will not use exactly (1.16) as zero order approximation but replace therein $b$ by a suitable mollification observed along the associated backward (Lipschitz) flow. Rather naturally, the mollification parameter is chosen to correspond to the characteristic scaling time, see equation (1.21) below

With respect to the various previously described steps, let us indicate that some properties of mollified flows, that appear in our main results, will be stated at the end of the section. The bounds on the frozen densities and the related convolution estimates, which will appear when iterating the first order expansion in (1.12) will be discussed in Section 2.

We will here somehow follow the main line of [34] but are faced with many additional difficulties. In particular, a common feature to both the Gaussian SDEs considered in [34] and the stable driven here is that we first need to establish the density and gradient estimates for smooth coefficients and a bounded drift, in order to justify that the previous expansions can indeed be performed (see in particular Lemma 3.1). In the current strictly stable framework we cannot rely on the Malliavin calculus arguments of [34] because of integrability issues. We thus establish here some direct bounds on the associated semi-group and its derivatives when the coefficients are smooth and the

---

[(2)]Let us mention that the boundedness of the drift coefficient in [28] follows from the specific chosen zero order approximation. Namely the diffusion coefficient therein is not frozen along the flow but at the final spatial point, see e.g. the proof of Lemma 4.1 in that reference.

drift bounded which serve as a starting point to derive the estimates of Theorem 1.1. This part is crucial and quite intricate (see Theorems 3.5 and 4.1 below). Since the two-sided and gradient estimates hold independently of the smoothness of the coefficients and the boundedness of the drift, we eventually conclude through compactness arguments (see Section 5).

To state our main result, we introduce the regularized flow associated with drift $b$. For $\varepsilon > 0$, let $b_\varepsilon(t,x) := b(t,\cdot) * \rho_\varepsilon(x)$, where $\rho_\varepsilon(x) = \varepsilon^{-d}\rho(x/\varepsilon)$ and $\rho$ is a smooth density function with support in the unit ball $B(0,1)$. Note that under $(\mathbf{H}_b^\beta)$,

$$
\begin{aligned}
(1.18) \quad |b_\varepsilon(t,x) - b(t,x)| &= \left| \int_{\mathbb{R}^d} \big(b(t,x-y) - b(t,x)\big)\rho_\varepsilon(y)\,\mathrm{d}y \right| \\
&\leqslant \kappa_0 \int_{B(0,\varepsilon)} (|y|^\beta + |y|)\rho_\varepsilon(y)\,\mathrm{d}y \leqslant \kappa_0(\varepsilon^\beta + \varepsilon),
\end{aligned}
$$

since $\int_{B(0,\varepsilon)} \rho_\varepsilon(y)\,\mathrm{d}y = 1$. Recalling now that $\nabla_x \underbrace{\int_{\mathbb{R}^d} \rho_\varepsilon(x-y)\,\mathrm{d}y}_{=1} = 0$,

$$
\begin{aligned}
(1.19) \quad |\nabla b_\varepsilon(t,x)| &= \left| \int_{\mathbb{R}^d} (b(t,y) - b(t,x))\nabla\rho_\varepsilon(x-y)\,\mathrm{d}y \right| \\
&\leqslant \kappa_0 \int_{B(x,\varepsilon)} (|x-y|^\beta + |x-y|)\varepsilon^{-(d+1)}|\nabla\rho(z)|_{z=(x-y)/\varepsilon}\,\mathrm{d}y \\
&\leqslant \kappa_0(\varepsilon^{\beta-1} + 1)\|\nabla\rho\|_{L^1}.
\end{aligned}
$$

In particular, since $\alpha + \beta > 1$, for any $T > 0$, there is a $C > 0$ such that for any $0 \leqslant s < t \leqslant T$,

$$
(1.20) \quad \int_s^t \|\nabla b_{|r-s|^{1/\alpha}}(r,\cdot)\|_\infty\,\mathrm{d}r \leqslant C \int_s^t \left( |r-s|^{(\beta-1)/\alpha} + 1 \right)\mathrm{d}r \leqslant C(t-s)^{(\alpha+\beta-1)/\alpha}.
$$

Thus, for fixed $s > 0$, the following ODE admits a unique solution $\theta_{s,t}(x)$:

$$
(1.21) \quad \dot\theta_{s,t} = b_{|t-s|^{1/\alpha}}(t,\theta_{s,t}), \quad \theta_{s,s} = x, \ t \geqslant 0.
$$

Note that for $t > s$, $\theta_{s,t}(x)$ denotes the forward solution of the above ODE, while for $t < s$, it denotes the backward solution. We carefully mention that our main results will be stated with respect to to the flow $\theta$ in (1.21) which is precisely associated with a mollified drift with parameter corresponding to the typical scale of the driving process of the SDE (1.1) at the current considered time.

For notational simplicity, we introduce the following parameter set

$$
(1.22) \quad \Theta := (\kappa_0, \kappa_1, d, \alpha, \beta, \gamma).
$$

We also denote for $T \in (0, \infty]$,

$$
\mathbb{D}_T := \{(s,x,t,y) : 0 \leqslant s < t < T, \ x,y \in \mathbb{R}^d\}.
$$

We will frequently use from now on the notation $\lesssim$. For two quantities $Q_1$ and $Q_2$, we mean by $Q_1 \lesssim Q_2$ that there exists $C := C(T, \Theta)$ such that $Q_1 \leqslant CQ_2$. Other

possible dependencies for the constants will be explicitly specified. Moreover, we also use the following notation

$$(1.23) \qquad |\mathcal{D}^{(\alpha)} f|(x) := \int_{\mathbb{R}^d} \frac{|\delta_f^{(2)}(x;z)|}{|z|^{d+\alpha}} \, \mathrm{d}z.$$

The aim of this paper is to show the following result.

Theorem 1.1. — *Under* $(\mathbf{H}_b^\beta)$ *and* $(\mathbf{H}_a^\gamma)$, *for each* $0 \leqslant s < t < \infty$ *and* $x \in \mathbb{R}^d$, $X_{s,t}(x)$ *admits a density* $p(s,x,t,y)$ *(called heat kernel of* $\mathscr{L}_s$*) that is continuous as a function of* $x, y$, *and such that for each* $t > 0$ *and* $x, y \in \mathbb{R}^d$ *and Lebesgue almost all* $s \in [0,t)$,

$$\partial_s p(s,x,t,y) = \mathscr{L}_s p(s,\cdot,t,y)(x), \quad p(s,x,t,\cdot) \to \delta_{\{x\}}(\cdot) \text{ weakly as } s \uparrow t,$$

*where* $\delta_{\{x\}}(\mathrm{d}y)$ *denotes the Dirac measure concentrated at* $x$. *Moreover, we have*

(i) *(Two-sides estimate) For any* $T > 0$, *there is a constant* $C_1 = C_1(T,\Theta) \geqslant 1$ *such that for all* $(s,x,t,y) \in \mathbb{D}_T$,

$$(1.24) \qquad p(s,x,t,y) \asymp_{C_1} (t-s)((t-s)^{1/\alpha} + |\theta_{s,t}(x) - y|)^{-d-\alpha},$$

*where* $\theta_{s,t}(x)$ *is defined by ODE* (1.21).

(ii) *(Fractional derivative estimate) For any* $T > 0$, *there is a constant* $C_2 = C_2(T,\Theta) > 0$ *such that for all* $(s,x,t,y) \in \mathbb{D}_T$,

$$|\mathcal{D}^{(\alpha)} p(s,\cdot,t,y)|(x) \leqslant C_2((t-s)^{1/\alpha} + |\theta_{s,t}(x) - y|)^{-d-\alpha}.$$

(iii) *(Gradient estimate in* $x$*) For any* $T > 0$, *there is a constant* $C_3 = C_3(T,\Theta) > 0$ *such that for all* $(s,x,t,y) \in \mathbb{D}_T$,

$$(1.25) \qquad |\nabla_x \log p(s,x,t,y)| \leqslant C_3 (t-s)^{-1/\alpha}.$$

Remark 1.2. — If $|b(t,x) - b(t,y)| \leqslant \kappa_0 |x-y|^\beta$ for any $x, y \in \mathbb{R}^d$ with $|x-y| \leqslant 1$ and $|b(0,t)| \leqslant \kappa_0$ for all $t \geqslant 0$, then $(\mathbf{H}_b^\beta)$ holds. In particular, for $c(x)$ being a bounded $\beta$-Hölder continuous function, $b(x) := x + c(x)$ satisfies $(\mathbf{H}_b^\beta)$.

Remark 1.3. — For $\alpha \in [1,2)$, we can replace $\theta_{s,t}(x)$ in (1.24) by any *regularized* flow $\theta_{s,t}^{(\varepsilon)}(x)$ defined in (2.1) below. When $\alpha \in (0,1)$, we choose the regularizing parameter $\varepsilon = (t-s)^{1/\alpha}$ since we need to use $\varepsilon$ to compensate the time singularity in the supercritical case. For $\alpha \in (0,1]$, since $b$ is continuous in $x$, we can replace $\theta_{s,t}(x)$ in (1.24) by any measurable *Peano* flow $\vartheta_{s,t}(x)$ of the ODE $\dot{\vartheta}_{s,t}(x) = b(t, \vartheta_{s,t}(x))$.

Remark 1.4. — When $b \equiv 0$ and $L^{(\alpha)}$ is a general $\alpha$-stable-like generator of the form (1.2), it was proved in [31] that the gradient estimate (1.25) holds for $\alpha \in (1/2, 2)$. See also [10] for more general absolutely continuous Lévy measures enjoying suitable scaling properties. It seems that our gradient estimate (1.25) is the first result for SDEs of the form (1.1) driven by a rotationally invariant $\alpha$-stable process with $\alpha \in (0, 1/2)$.

The paper is organized as follows. We give in Section 2 some preliminary estimates needed for the main analysis. This concerns the mollified flow, some exit probabilities, convolution inequalities and the density of the *proxy process* involved in the parametrix (which has a dynamic similar to (1.1) with coefficient frozen along a suitable deterministic flow). Section 3 is then devoted to the derivation of the two-sided bound and the fractional derivative estimate under $(\mathbf{H}_b^\beta)$ and $(\mathbf{H}_a^\gamma)$ when the coefficients are additionally supposed to be smooth and the drift bounded. We specifically address the gradient estimate under those same assumptions in Section 4. We eventually present in Section 5 some compactness arguments to derive the results of Theorem 1.1 under the sole conditions $(\mathbf{H}_b^\beta)$, $(\mathbf{H}_a^\gamma)$.

*Acknowledgements.* — The authors would like to thank the referee for his/her very useful comments.

## 2. Preliminaries

2.1. ODE flow. — We first present some basic properties about the solution $\theta_{s,t}(x)$ of the ODE (1.21). Since the drift coefficient therein depends on the initial time $s$, the following flow property does no longer hold:

$$\theta_{r,t} \circ \theta_{s,r}(x) = \theta_{s,t}(x), \quad s < r < t.$$

However, the above flow property holds for the following regularized ODE:

$$(2.1) \qquad \dot{\theta}_{s,t}^{(\varepsilon)}(x) = b_\varepsilon(t, \theta_{s,t}^{(\varepsilon)}(x)), \quad \theta_{s,s}^{(\varepsilon)}(x) = x,$$

for any fixed regularizing parameter $\varepsilon > 0$. Below we fix $\alpha \in (0,2)$ and always assume $(\mathbf{H}_b^\beta)$. The following lemma is easy.

Lemma 2.1

(i) *For each $\varepsilon > 0$ and $s, t \geqslant 0$, $x \mapsto \theta_{s,t}^{(\varepsilon)}(x)$ is a $C^1$-diffeomorphism and*

$$(2.2) \qquad (\theta_{s,t}^{(\varepsilon)})^{-1}(y) = \theta_{t,s}^{(\varepsilon)}(y).$$

*Moreover, for all $s, r, t \geqslant 0$, it holds that*

$$(2.3) \qquad \theta_{s,t}^{(\varepsilon)}(x) = \theta_{r,t}^{(\varepsilon)} \circ \theta_{s,r}^{(\varepsilon)}(x).$$

(ii) *For all $\varepsilon, \varepsilon' > 0$ and $s, t \geqslant 0$, $x \in \mathbb{R}^d$, it holds that*

$$(2.4) \qquad |\theta_{s,t}^{(\varepsilon')}(x) - \theta_{s,t}^{(\varepsilon)}(x)| \leqslant 2\kappa_0 (\varepsilon \vee \varepsilon')^\beta |t - s| e^{\kappa_0 \|\nabla \rho\|_{L^1} ((\varepsilon \vee \varepsilon')^{\beta-1}+1)|t-s|},$$

(iii) *For any $T > 0$, there is a constant $C = C(T, d, \kappa_0) > 0$ such that for all $s, t \in [0, T]$, $x, y \in \mathbb{R}^d$ and $\varepsilon = |t - s|^{1/\alpha}$,*

$$(2.5) \qquad |\theta_{s,t}^{(\varepsilon)}(x) - y| \asymp_C |x - \theta_{t,s}^{(\varepsilon)}(y)|, \quad |\theta_{s,t}^{(\varepsilon)}(x) - \theta_{s,t}^{(\varepsilon)}(y)| \asymp_C |x - y|.$$

*Proof*

(i) Note that by (2.1), for $0 \leqslant s < t$:

$$\theta_{s,t}^{(\varepsilon)}(x) = x + \int_s^t b_\varepsilon(r, \theta_{s,r}^{(\varepsilon)}(x))\,\mathrm{d}r, \quad \theta_{t,s}^{(\varepsilon)}(y) = y - \int_s^t b_\varepsilon(r, \theta_{t,r}^{(\varepsilon)}(y))\,\mathrm{d}r.$$

Let $y = \theta_{s,t}^{(\varepsilon)}(x)$. By the flow property, we have

$$y = \left(\theta_{s,t}^{(\varepsilon)}\right)^{-1}(y) + \int_s^t b_\varepsilon(r, \theta_{s,r}^{(\varepsilon)} \circ \left(\theta_{s,t}^{(\varepsilon)}\right)^{-1}(y))\,\mathrm{d}r$$

$$= (\theta_{s,t}^{(\varepsilon)})^{-1}(y) + \int_s^t b_\varepsilon(r, \left(\theta_{r,t}^{(\varepsilon)}\right)^{-1}(y))\,\mathrm{d}r.$$

Since the ODE has a unique solution, we immediately have $\left(\theta_{s,t}^{(\varepsilon)}\right)^{-1}(y) = \theta_{t,s}^{(\varepsilon)}(y)$. As for (2.3), it follows from (2.2) and the flow property.

(ii) Without loss of generality, we assume $\varepsilon' < \varepsilon$. Since by (1.18) and (1.19),

$$(2.6) \qquad |b_\varepsilon(t,x) - b_{\varepsilon'}(t,x)| \leqslant 2\kappa_0 \varepsilon^\beta, \quad \|\nabla b_\varepsilon\|_\infty \leqslant \kappa_0 \|\nabla\rho\|_{L^1}(\varepsilon^{\beta-1} + 1),$$

by definition we have

$$|\theta_{s,t}^{(\varepsilon')}(x) - \theta_{s,t}^{(\varepsilon)}(x)| \leqslant \int_s^t |b_{\varepsilon'}(r, \theta_{s,r}^{(\varepsilon')}(x)) - b_\varepsilon(r, \theta_{s,r}^{(\varepsilon')}(x))|\,\mathrm{d}r$$

$$+ \int_s^t |b_\varepsilon(r, \theta_{s,r}^{(\varepsilon')}(x) - b_\varepsilon(r, \theta_{s,r}^{(\varepsilon)}(x)|\,\mathrm{d}r$$

$$\leqslant 2\kappa_0 \varepsilon^\beta(t - s) + \kappa_0 \|\nabla\rho\|_{L^1}(\varepsilon^{\beta-1} + 1) \int_s^t |\theta_{s,r}^{(\varepsilon')}(x) - \theta_{s,r}^{(\varepsilon)}(x)|\,\mathrm{d}r.$$

Using Gronwall's inequality, we obtain (2.4).

(iii) Without loss of generality, we assume $0 \leqslant s < t \leqslant T$. Note that for $u \in [s,t]$,

$$|\theta_{s,u}^{(\varepsilon)}(x) - \theta_{s,u}^{(\varepsilon)}(y)| \leqslant |x - y| + \int_s^u \|\nabla b_\varepsilon(r,\cdot)\|_\infty |\theta_{s,r}^{(\varepsilon)}(x) - \theta_{s,r}^{(\varepsilon)}(y)|\,\mathrm{d}r.$$

For $\varepsilon = |t - s|^{1/\alpha}$, it follows from the Gronwall inequality and (1.19) that

$$|\theta_{s,t}^{(\varepsilon)}(x) - \theta_{s,t}^{(\varepsilon)}(y)| \leqslant \mathrm{e}^{\kappa_0 \|\nabla\rho\|_{L^1}(|t-s|^{(\beta-1)/\alpha}+1)|t-s|}|x - y|.$$

Similarly, from the backwards dynamics, one can derive that for $x', y' \in \mathbb{R}^d$,

$$|\theta_{t,s}^{(\varepsilon)}(x') - \theta_{t,s}^{(\varepsilon)}(y')| \leqslant \mathrm{e}^{\kappa_0 \|\nabla\rho\|_{L^1}(|t-s|^{(\beta-1)/\alpha}+1)|t-s|}|x' - y'|.$$

Thus, by (2.2) (inverse flow property) and putting $x' = \theta_{s,t}^{(\varepsilon)}(x)$ and $y' = \theta_{s,t}^{(\varepsilon)}(y)$ we obtain the second equivalence in (2.5). The first equivalence follows from the second one replacing therein $y$ by $\theta_{t,s}^\varepsilon(y)$ and using (2.2). $\qquad\square$

The following result is a consequence of the above lemma, which plays a crucial role below.

**Lemma 2.2**

(i) *For each $s, t \geqslant 0$, the map $x \mapsto \theta_{s,t}(x)$, where $(\theta_{s,u}(x))_{u \in [s,t]}$ solves (1.21), is a $C^1$-diffeomorphism and there is a constant $C_0 = C_0(T, \Theta) > 0$ such that*

$$|\det(\nabla\theta_{s,t}^{-1}(x)) - 1| \leqslant C_0|t - s|^{(\alpha+\beta-1)/\alpha}.$$

(ii) *For any $T > 0$, there is a constant $C_1 = C_1(T, \Theta) \geqslant 1$ such that for all $s, t \in [0, T]$ and $x, y \in \mathbb{R}^d$,*

$$(2.7) \qquad |t - s|^{1/\alpha} + |\theta_{s,t}(x) - y| \asymp_{C_1} |t - s|^{1/\alpha} + |x - \theta_{t,s}(y)|.$$

(iii) *For any $T > 0$, there is a constant $C_2 = C_2(T, \Theta) > 0$ such that for all $s, r, t \in [0, T]$ and $x \in \mathbb{R}^d$,*

$$(2.8) \qquad |\theta_{s,t}(x) - \theta_{r,t} \circ \theta_{s,r}(x)| \leqslant C_2 |(r \vee s \vee t) - (r \wedge s \wedge t)|^{1/\alpha}.$$

*Proof*

(i) It is well known that (cf. [32, Prop. 1.2])

$$\det(\nabla \theta_{s,t}(x)) = 1 + \int_s^t \operatorname{div} b_{|r-s|^{1/\alpha}}(r, \theta_{s,r}(x)) \det(\nabla \theta_{s,r}(x)) \, dr.$$

Thus

$$\det(\nabla \theta_{s,t}(x)) = \exp\left\{ \int_s^t \operatorname{div} b_{|r-s|^{1/\alpha}}(r, \theta_{s,r}(x)) \, dr \right\}.$$

The desired estimate follows by (1.20) and $\nabla \theta_{s,t}^{-1}(x) = (\nabla \theta_{s,t})^{-1}(\theta_{s,t}^{-1}(x))$.

(ii) Fix $s < t$. For $u \in [s, t]$, by definition we have

$$\theta_{s,u}(x) = x + \int_s^u b_{|r-s|^{1/\alpha}}(r, \theta_{s,r}(x)) \, dr,$$

and for $\varepsilon = |t - s|^{1/\alpha}$,

$$\theta_{s,u}^{(\varepsilon)}(x) = x + \int_s^u b_\varepsilon(r, \theta_{s,r}^{(\varepsilon)}(x)) \, dr.$$

By (2.6) with $\varepsilon' = |r - s|^{1/\alpha}$ and $\varepsilon = |t - s|^{1/\alpha}$, we have for all $u \geqslant s$,

$$|\theta_{s,u}(x) - \theta_{s,u}^{(\varepsilon)}(x)| \leqslant \int_s^u \left| b_{|r-s|^{1/\alpha}}(r, \theta_{s,r}(x)) - b_\varepsilon(r, \theta_{s,r}^{(\varepsilon)}(x)) \right| dr$$

$$\lesssim |t - s|^{\beta/\alpha + 1} + (t - s)^{(\beta-1)/\alpha} \int_s^u |\theta_{s,r}(x) - \theta_{s,r}^{(\varepsilon)}(x)| \, dr,$$

which yields by Gronwall's inequality that

$$(2.9) \qquad |\theta_{s,t}(x) - \theta_{s,t}^{(\varepsilon)}(x)| \lesssim \mathrm{e}^{(t-s)^{(\beta-1)/\alpha+1}} |t - s|^{\beta/\alpha+1} \lesssim |t - s|^{1/\alpha},$$

where the second inequality is due to $\alpha + \beta > 1$. Thus, by (2.5) and (2.9), we have

$$|\theta_{s,t}(x) - y| \leqslant |\theta_{s,t}^{(\varepsilon)}(x) - y| + |\theta_{s,t}(x) - \theta_{s,t}^{(\varepsilon)}(x)|$$

$$\lesssim |x - \theta_{s,t}^{(\varepsilon)}(y)| + |t - s|^{1/\alpha} \lesssim |x - \theta_{t,s}(y)| + |t - s|^{1/\alpha}.$$

The right hand side inequality of (2.7) follows. By symmetry, we also have the left hand side inequality.

(iii) Let $s, r, t \geqslant 0$ and $\varepsilon := |r \vee s \vee t - r \wedge s \wedge t|^{1/\alpha}$. By (2.3) we have

$$|\theta_{s,t}(x) - \theta_{r,t} \circ \theta_{s,r}(x)| \leqslant |\theta_{s,t}(x) - \theta_{s,t}^{(\varepsilon)}(x)| + |\theta_{r,t}^{(\varepsilon)} \circ \theta_{s,r}^{(\varepsilon)}(x) - \theta_{r,t}^{(\varepsilon)} \circ \theta_{s,r}(x)|$$

$$+ |\theta_{r,t}^{(\varepsilon)} \circ \theta_{s,r}(x) - \theta_{r,t} \circ \theta_{s,r}(x)|.$$

The desired estimate (2.8) again follows by (2.5) and (2.9). $\qquad \square$

Remark 2.3. — By (2.8), we have

$$|x - \theta_{t,s} \circ \theta_{s,t}(x)| \lesssim_{C_2} |t - s|^{1/\alpha},$$

and by (2.7),

$$(2.10) \qquad |t - s|^{1/\alpha} + |x - y| \asymp_{C_1} |t - s|^{1/\alpha} + |\theta_{s,t}(x) - \theta_{t,s}(y)|.$$

Put it differently, we have an *approximate* flow property for the ODE (1.21). Namely, the flow property holds up to an additive time factor which has the same magnitude as the current typical time (self-similarity index of the driving process).

2.2. Probability estimates. — We need the following master formula.

Lemma 2.4 (Lévy system). — *Let $X_t := X_{0,t}$ be any solution of SDE* (1.1). *For any nonnegative measurable function $f : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ and finite stopping time $\tau$,*

$$\mathbb{E} \sum_{r \in (0,\tau]} f(r, X_{r-}, \Delta X_r) = \mathbb{E} \int_0^\tau \int_{\mathbb{R}^d} f(r, X_{r-}, z) \frac{\kappa(r, X_{r-}, z)}{|z|^{d+\alpha}} \, \mathrm{d}z \, \mathrm{d}r,$$

*where $\Delta X_r := X_r - X_{r-}$ and $\kappa(r, x, z)$ is defined in* (1.4).

Let us emphasize that the Lévy system is well known for a time-homogeneous process, we can e.g. refer to Chapter 2.4 in [5] and references therein. The extension to the current framework under the uniform ellipticity condition is rather natural. We provide a proof for the sake of completeness.

*Proof.* — Let $N(\mathrm{d}t, \mathrm{d}z)$ be the counting measure associated with $L_t^{(\alpha)}$, i.e.,

$$N((0,t] \times E) := \sum_{s \in [0,t]} \mathbf{1}_E(\Delta L_s^{(\alpha)}), \quad E \in \mathscr{B}(\mathbb{R}^d).$$

Noting that

$$\Delta X_t = a(t, X_{t-})\Delta L_t^{(\alpha)},$$

we have for any $\varepsilon > 0$,

$$\sum_{r \in (0,\tau]} f(r, X_{r-}, \Delta X_r)\mathbf{1}_{|\Delta X_r| \geqslant \varepsilon} = \int_0^\tau \int_{|a(r, X_{r-})z| \geqslant \varepsilon} f(r, X_{r-}, a(r, X_{r-})z)N(\mathrm{d}r, \mathrm{d}z).$$

Since the compensated measure of $N(\mathrm{d}t, \mathrm{d}z)$ is $\mathrm{d}z \, \mathrm{d}t/|z|^{d+\alpha}$, by the change of variable, we have

$$\mathbb{E} \sum_{r \in (0,\tau]} f(r, X_{r-}, \Delta X_r)\mathbf{1}_{|\Delta X_r| \geqslant \varepsilon} = \mathbb{E} \int_0^\tau \int_{|a(r, X_{r-})z| \geqslant \varepsilon} f(r, X_{r-}, a(r, X_{r-})z) \frac{\mathrm{d}z \, \mathrm{d}r}{|z|^{d+\alpha}}$$

$$= \mathbb{E} \int_0^\tau \int_{|z| \geqslant \varepsilon} f(r, X_{r-}, z) \frac{\kappa(r, X_{r-}, z)}{|z|^{d+\alpha}} \, \mathrm{d}z \, \mathrm{d}r,$$

where $\kappa(r, x, z)$ is given in (1.4), which in turn gives the desired formula by the monotone convergence theorem. □

Fix $(s,x) \in \mathbb{R}_+ \times \mathbb{R}^d$. For $\eta > 0$, define the stopping time

$$(2.11) \qquad \tau_{s,x}^\eta := \inf\{t > s : |X_{s,t}(x) - \theta_{s,t}(x)| > \eta\},$$

which corresponds to the exit time of the diffusion from a tube around the deterministic ODE introduced in (1.21). We now give a *tube estimate* which roughly says that, for a given spatial threshold $\eta$, the probability that the difference between the process $X_{s,\bullet}(x)$ and the deterministic regularized flow $\theta_{s,\cdot}(x)$ leaves the tube of radius $\eta$ before a certain fraction $\varepsilon\eta^\alpha$ of the corresponding typical time scale $\eta^\alpha$ is somehow *small*.

LEMMA 2.5. — *Under* $(\mathbf{H}_b^\beta)$ *and* $(\mathbf{H}_a^0)$, *for any* $T > 0$, *there is an* $\varepsilon \in (0,1)$ *only depending on* $T, \Theta$ *such that for all* $(s,x) \in \mathbb{R}_+ \times \mathbb{R}^d$ *and* $\eta \in (0, T^{1/\alpha}]$,

$$\mathbb{P}(\tau_{s,x}^\eta < s + \varepsilon\eta^\alpha) \leqslant 1/2.$$

*Proof.* — Without loss of generality, we assume $(s,x) = (0,0)$, and for simplicity write

$$X_t := X_{0,t}(0), \quad \theta_t := \theta_{0,t}(0), \quad \tau := \tau_{0,0}^\eta.$$

Let $f \in C_b^2(\mathbb{R}^d)$ with $f(0) = 0$ and $f(x) = 1$ for $|x| > 1$. For $\eta > 0$, set

$$f_\eta(x) := f(x/\eta), \quad u := \varepsilon\eta^\alpha.$$

Note that

$$Y_t := X_t - \theta_t = \int_0^t (b(r, X_r) - b_{r^{1/\alpha}}(r, \theta_r))\,\mathrm{d}r + \int_0^t a(r, X_{r-})\,\mathrm{d}L_r^{(\alpha)}.$$

By Itô's formula, we have

$$\mathbb{E}f_\eta(Y_{u \wedge \tau}) = \mathbb{E}\int_0^{u \wedge \tau}\Big[(b(r, X_r) - b_{r^{1/\alpha}}(r, \theta_r)) \cdot \nabla f_\eta(Y_r) + \tfrac{1}{2}\mathcal{L}_r f_\eta(Y_r)\Big]\,\mathrm{d}r.$$

Note that

$$|b(r, X_r) - b_{r^{1/\alpha}}(r, \theta_r)| \leqslant |b(r, X_r) - b(r, \theta_r)| + |b(r, \theta_r) - b_{r^{1/\alpha}}(r, \theta_r)|$$
$$\underset{(\mathbf{H}_b^\beta),(2.6)}{\leqslant} \kappa_0(|X_r - \theta_r|^\beta + |X_r - \theta_r|) + \kappa_0 r^{\beta/\alpha}.$$

Also, with align the notation of (1.3),

$$|\delta_{f_\eta}^{(2)}(x; z)| \leqslant (|z|^2 \|\nabla^2 f_\eta\|_\infty) \wedge (4\|f_\eta\|_\infty) \lesssim (|z|^2/\eta^2) \wedge 1.$$

Hence,

$$\mathbb{E}f_\eta(Y_{u \wedge \tau}) \lesssim \mathbb{E}\int_0^u\left[(|Y_r|^\beta + |Y_r| + r^{\beta/\alpha}) \cdot \frac{\mathbf{1}_{|Y_r| \leqslant \eta}}{\eta} + \int_{\mathbb{R}^d} \frac{(|z|^2/\eta^2) \wedge 1}{|z|^{d+\alpha}}\,\mathrm{d}z\right]\mathrm{d}r$$

$$\lesssim \int_0^u (\eta^{\beta-1} + 1 + r^{\beta/\alpha}\eta^{-1} + \eta^{-\alpha})\,\mathrm{d}r$$

$$\underset{u = \varepsilon\eta^\alpha}{\lesssim} \varepsilon(\eta^{\alpha+\beta-1} + \eta^\alpha + \varepsilon^{\beta/\alpha}\eta^{\alpha+\beta-1} + 1) \lesssim \varepsilon(T^{1+(\beta-1)/\alpha} + T + 1),$$

where the last inequality is due to $\alpha + \beta > 1$ and $\eta \leqslant T^{1/\alpha}$. Importantly, the implicit constant is independent of $\varepsilon$. Note that

$$\mathbb{P}(\tau < u) = \mathbb{E}\mathbf{1}_{\tau < u} \leqslant \mathbb{E}f_\eta\big(Y_{u \wedge \tau}\big) \lesssim \varepsilon(T^{1+(\beta-1)/\alpha} + T + 1).$$

The desired estimate follows by choosing $\varepsilon$ small enough. $\qquad\square$

The following lemma will be used to show the lower bound estimate of the heat kernel. It gives a lower bound estimate for the probability that, considered two points $x, y$ in the *off-diagonal regime* between times $s$ and $t$, namely such that $|x - \theta_{t,s}(y)| \geqslant K(t-s)^{1/\alpha}$, after a time $\varepsilon(t-s)$ with $\varepsilon$ as in Lemma 2.5, the stochastic forward transport of $x$ by the SDE, i.e., $X_{s,s+\varepsilon(t-s)}(x)$ and the backward deterministic transport of $y$ by the regularized flow $\theta_{t,s+\varepsilon(t-s)}(y)$ belong to a *diagonal* tube, with radius $K(t-s)^{1/\alpha}$, where again $(t-s)^{1/\alpha}$ corresponds to the current typical scale between times $s$ and $t$.

LEMMA 2.6. — *Suppose that $(\mathbf{H}_b^\beta)$ and $(\mathbf{H}_a^0)$ hold. Let $\varepsilon \in (0,1)$ be as in Lemma 2.5. For any $T > 0$, there are constants $c_0 \in (0,1), K \geqslant 1$ depending only on $T, \Theta$ such that for all $0 \leqslant s < t \leqslant T$ and $|x - \theta_{t,s}(y)| \geqslant K(t-s)^{1/\alpha}$,*

$$\mathbb{P}\Big(|X_{s,s+\varepsilon(t-s)}(x) - \theta_{t,s+\varepsilon(t-s)}(y)| \leqslant K(t-s)^{1/\alpha}\Big) \geqslant \frac{c_0(t-s)^{1+d/\alpha}}{|x - \theta_{t,s}(y)|^{d+\alpha}}.$$

This lemma will be crucial for the lower bound estimate of the heat kernel, since it precisely gives the control needed for a chaining argument, see Theorem 3.5 below. As opposed to the continuous case, for SDEs driven by pure jump processes, a single intermediate time, associated with a *large jump*, is needed for the chaining. Roughly speaking between times $s+\varepsilon(t-s)$ and $t$ we will use the global *diagonal* bound of order $(t-s)^{-d/\alpha}$, since $\varepsilon$ is meant to be *small enough*, and the above lemma controls the probability that the process enters a good neighborhood of the backward flow to do so. The lower bound is the sought one in the sense that when multiplying it by $(t-s)^{-d/\alpha}$ exactly makes the expression in (1.24), $(t-s)((t-s)^{1/\alpha} + |\theta_{s,t}(x) - y|)^{-(d+\alpha)} \asymp_C (t-s)|\theta_{s,t}(x) - y|^{-(d+\alpha)}$ appear since $|x - \theta_{s,t}(y)| \geqslant K(t-s)^{1/\alpha}$ (*off-diagonal regime*).

*Proof of Lemma 2.6.* — Without loss of generality, we assume $s = 0$ and for simplicity, we write

$$\eta := t^{1/\alpha}, \ u := \varepsilon\eta^\alpha = \varepsilon t, \quad X_r(x) := X_{0,r}(x).$$

Define a stopping time

$$\sigma := \inf\big\{r > 0 : |X_r(x) - \theta_{t,r}(y)| \leqslant \eta\big\}.$$

By the right continuity of $r \mapsto X_r(x) - \theta_{t,r}(y)$, one sees that

$$|X_\sigma(x) - \theta_{t,\sigma}(y)| \leqslant \eta, \ \text{a.s.}$$

In particular, for $\sigma \leqslant u$, by (2.10) and (2.8), there is a constant $C_0 = C_0(\Theta) \geqslant 1$ such that

$$|X_u(x) - \theta_{t,u}(y)| \leqslant |X_{\sigma,u}(X_\sigma(x)) - \theta_{\sigma,u}(X_\sigma(x))| + |\theta_{\sigma,u}(X_\sigma(x)) - \theta_{\sigma,u}(\theta_{t,\sigma}(y))|$$
$$+ |\theta_{\sigma,u}(\theta_{t,\sigma}(y)) - \theta_{t,u}(y)|$$
$$\leqslant |X_{\sigma,u}(X_\sigma(x)) - \theta_{\sigma,u}(X_\sigma(x))| + C_0\eta.$$

Let $K \geqslant C_0 + 1$. Then

$$\left\{ |X_{\sigma,u}(X_\sigma(x)) - \theta_{\sigma,u}(X_\sigma(x))| < \eta \right\} \subset \left\{ |X_u(x) - \theta_{t,u}(y)| \leqslant K\eta \right\}.$$

Thus, by the strong Markov property, we have

$$\mathbb{P}(|X_u(x) - \theta_{t,u}(y)| \leqslant K\eta) \geqslant \mathbb{P}(\sigma \leqslant u; |X_u(x) - \theta_{t,u}(y)| \leqslant K\eta)$$
$$\geqslant \mathbb{P}(\sigma \leqslant u; |X_{\sigma,u}(X_\sigma(x)) - \theta_{\sigma,u}(X_\sigma(x))| < \eta)$$
$$\geqslant \mathbb{P}(\sigma \leqslant u) \inf_{(s,z)\in[0,u]\times\mathbb{R}^d} \mathbb{P}(|X_{s,u}(z) - \theta_{s,u}(z)| < \eta).$$

Let $\tau_{s,z}^\eta$ be defined by (2.11). By Lemma 2.5 we have

$$(2.12) \qquad \mathbb{P}\left(|X_{s,u}(z) - \theta_{s,u}(z)| \geqslant \eta\right) \leqslant \mathbb{P}(\tau_{s,z}^\eta \leqslant u) \leqslant \mathbb{P}(\tau_{s,z}^\eta \leqslant s + \varepsilon\eta^\alpha) \leqslant 1/2,$$

which implies that

$$\inf_{(s,z)\in[0,u]\times\mathbb{R}^d} \mathbb{P}\left(|X_{s,u}(z) - \theta_{s,u}(z)| < \eta\right) \geqslant 1/2$$

and

$$(2.13) \qquad \mathbb{P}(|X_u(x) - \theta_{t,u}(y)| \leqslant K\eta) \geqslant \mathbb{P}(\sigma \leqslant u)/2.$$

Next we need to obtain a lower bound estimate for $\mathbb{P}(\sigma \leqslant u)$. Let $\tau := \tau_{0,x}^\eta$. For $r < u \wedge \tau$, by (2.7), there are constants $c_0, C_1 > 0$ such that

$$|X_r(x) - \theta_{t,r}(y)| \geqslant |\theta_{0,r}(x) - \theta_{t,r}(y)| - |X_r(x) - \theta_{0,r}(x)|$$
$$\geqslant c_0|x - \theta_{t,0}(y)| - C_1 t^{1/\alpha} - \eta.$$

In particular, if we choose $K \geqslant (C_0 + 1) \vee ((C_1 + 2)/c_0)$, then since by assumption

$$|x - \theta_{t,0}(y)| \geqslant K\eta,$$

it holds that for $r < u \wedge \tau$,

$$|X_r(x) - \theta_{t,r}(y)| \geqslant c_0 K\eta - (C_1 + 1)\eta > \eta.$$

Thus we have

$$\mathbf{1}_{\{|X_{u\wedge\tau}(x) - \theta_{t,u\wedge\tau}(y)|\leqslant\eta\}} = \sum_{r\in(0,u\wedge\tau]} \mathbf{1}_{\{|X_r(x)-\theta_{t,r}(y)|\leqslant\eta\}},$$

i.e., we have at most one term in the above summand. We then derive from Lemma 2.4,

$$\mathbb{P}\{|X_{u\wedge\tau}(x) - \theta_{t,u\wedge\tau}(y)| \leqslant \eta\} = \mathbb{E} \sum_{r\in(0,u\wedge\tau]} \mathbf{1}_{\{|X_r(x)-\theta_{t,r}(y)|\leqslant\eta\}}$$
$$= \mathbb{E} \int_0^{u\wedge\tau} \int_{|z-\theta_{t,r}(y)|\leqslant\eta} \frac{\kappa(r, X_{r-}(x), z - X_{r-}(x))}{|z - X_{r-}(x)|^{d+\alpha}} \, \mathrm{d}z \, \mathrm{d}r.$$

On the other hand, noting that for $r < u \wedge \tau$ and $|z - \theta_{t,r}(y)| \leqslant \eta$, we get

$$|z - X_{r-}(x)| \leqslant |z - \theta_{t,r}(y)| + |\theta_{t,r}(y) - \theta_{0,r}(x)| + |\theta_{0,r}(x) - X_{r-}(x)|$$
$$\leqslant \eta + C_2|x - \theta_{t,0}(y)| + C_2 t^{1/\alpha} + \eta \leqslant C_2|x - \theta_{t,0}(y)| + C_3\eta,$$

using as well (2.7) for the last but one inequality. Since

$$\{|X_{u \wedge \tau}(x) - \theta_{t,u \wedge \tau}(y)| \leqslant \eta\} \subset \{\sigma \leqslant u\},$$

we further have

$$\mathbb{P}\{\sigma \leqslant u\} \geqslant \mathbb{P}\{|X_{u \wedge \tau}(x) - \theta_{t,u \wedge \tau}(y)| \leqslant \eta\}$$

$$\text{(2.14)} \qquad \geqslant \mathbb{E} \int_0^{u \wedge \tau} \int_{|z - \theta_{t,r}(y)| \leqslant \eta} \frac{\kappa_1^{-1}}{(C_2|x - \theta_{t,0}(y)| + C_3\eta)^{d+\alpha}} \, dz \, dr$$

$$= \frac{\mathbb{E}(u \wedge \tau)\kappa_1^{-1}\eta^d \cdot \mathrm{Vol}(B_1)}{(C_2|x - \theta_{t,0}(y)| + C_3\eta)^{d+\alpha}} \geqslant \frac{c_0 t^{1+d/\alpha}}{|x - \theta_{t,0}(y)|^{d+\alpha}},$$

where the last step is due to $|x - \theta_{t,0}(y)| \geqslant K\eta$ and

$$\mathbb{E}(u \wedge \tau) \geqslant u\mathbb{P}(\tau > u) \overset{(2.12)}{\geqslant} u/2 = \varepsilon t/2.$$

Combining (2.13) and (2.14), we obtain the desired estimate. $\qquad \square$

2.3. Convolution inequalities. — This Section is dedicated to some useful convolution controls associated with functions that are known to be upper-bounds of the isotropic stable density and its gradient, see e.g. [22], [4]. Though a bit technical, these results will turn out to be crucial in order to control the parametrix series representation of the density and its gradient (see e.g. equation (3.12), Lemma 3.3 and Theorem 3.5 below).

For $\eta \in (0, 2)$ and $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^d$, let

$$\varrho^{(\eta)}(x) := (1 + |x|)^{-d-\eta}, \quad \varrho^{(\eta)}(t, x) := t^{-d/\alpha}\varrho^{(\eta)}(t^{-1/\alpha}x).$$

For $\beta \geqslant 0$ and $\gamma \in \mathbb{R}$, we introduce the following functions for later use

$$\text{(2.15)} \qquad \varrho_{\beta,\gamma}^{(\eta)}(t, x) := (1 \wedge (t^{1/\alpha} + |x|))^\beta t^{(\gamma-\eta)/\alpha}\varrho^{(\eta)}(t, x)$$

and for $0 \leqslant s < t$,

$$\text{(2.16)} \qquad \phi_{\beta,\gamma}^{(\eta)}(s, x, t, y) := \varrho_{\beta,\gamma}^{(\eta)}(t - s, x - \theta_{t,s}(y)).$$

Note that

$$\varrho_{\beta,\gamma}^{(\eta)}(t, x) = \frac{(1 \wedge (t^{1/\alpha} + |x|))^\beta t^{\gamma/\alpha}}{(t^{1/\alpha} + |x|)^{d+\eta}}.$$

For $T > 0$, by (2.7) we have for $(s, x, t, y) \in \mathbb{D}_T$,

$$\text{(2.17)} \qquad \phi_{\beta,\gamma}^{(\eta)}(s, x, t, y) \asymp \varrho_{\beta,\gamma}^{(\eta)}(t - s, \theta_{s,t}(x) - y),$$

and for $\beta \in [0, \eta]$,

$$\text{(2.18)} \qquad \int_{\mathbb{R}^d} \phi_{\beta,\gamma}^{(\eta)}(s, x, t, y) \, dy \lesssim \int_{\mathbb{R}^d} \varrho_{\beta,\gamma}^{(\eta)}(t - s, y) \, dy \lesssim (t - s)^{(\beta+\gamma-\eta)/\alpha}.$$

Observe as well importantly that, from (2.15), for $0 \leqslant \beta' \leqslant \beta$, $0 \leqslant \gamma' \leqslant \gamma$ and $(s, x, t, y) \in \mathbb{D}_T$,

$$(2.19) \qquad \phi_{\beta,\gamma}^{(\eta)}(s, x, t, y) \lesssim \phi_{\beta',\gamma'}^{(\eta)}(s, x, t, y).$$

For two functions $f, g$ on $\mathbb{D}_\infty$, we write

$$(f \odot g)_r(s, x, t, y) := \int_{\mathbb{R}^d} f(s, x, r, z) g(r, z, t, y) \, \mathrm{d}z$$

and

$$(f \otimes g)(s, x, t, y) := \int_s^t (f \odot g)_r(s, x, t, y) \, \mathrm{d}r.$$

The following lemma is the same as in [9, Lem. 2.1].

LEMMA 2.7. — *Fix $\alpha \in (0, 2)$. For any $\beta_1, \beta_2 \in [0, \alpha/4]$ and $T > 0$, there is a $C = C(T, \Theta, \beta_1, \beta_2) > 0$ such that for all $\gamma_1 > -\beta_1$ and $\gamma_2 > -\beta_2$, $r \in [s, t]$ and $x, y \in \mathbb{R}^d$,*

$$(2.20) \quad \left(\phi_{\beta_1,0}^{(\alpha)} \odot \phi_{\beta_2,0}^{(\alpha)}\right)_r(s, x, t, y) \lesssim_C \left((r-s)^{(\beta_1-\alpha)/\alpha} + (t-r)^{(\beta_2-\alpha)/\alpha}\right) \phi_{\beta_1 \wedge \beta_2, 0}^{(\alpha)}(s, x, t, y)$$

*and*

$$(2.21) \quad \phi_{\beta_1,\gamma_1}^{(\alpha)} \otimes \phi_{\beta_2,\gamma_2}^{(\alpha)}(s, x, t, y) \lesssim_C \mathcal{B}\left(\frac{\beta_1+\gamma_1}{\alpha}, \frac{\beta_2+\gamma_2}{\alpha}\right) \phi_{\beta_1 \wedge \beta_2, \beta_1+\beta_2+\gamma_1+\gamma_2}^{(\alpha)}(s, x, t, y),$$

*where $\mathcal{B}(\gamma, \beta)$ is the usual Beta function defined by*

$$\mathcal{B}(\gamma, \beta) := \int_0^1 (1-s)^{\gamma-1} s^{\beta-1} \, \mathrm{d}s, \quad \gamma, \beta > 0.$$

*Proof.* — We follow the proof in [10]. Let $\ell(u) := u^{d+\alpha}/1 \wedge u^\beta$. It is easy to see that, as soon as $d + \alpha > \beta$, $\ell$ is increasing on $\mathbb{R}_+$ and for any $\lambda \geqslant 1$,

$$(2.22) \qquad \ell(\lambda u) \leqslant \lambda^{d+\alpha} \ell(u).$$

Hence,

$$(2.23) \qquad \ell(u + w) \leqslant \ell(2(u \vee w)) \leqslant 2^{d+\alpha} \ell(u \vee w) \leqslant 2^{d+\alpha}(\ell(u) + \ell(w)).$$

Now for $r \in [s, t]$ and $x, y \in \mathbb{R}^d$, since

$$|t + s|^{1/\alpha} + |x + y| \leqslant 2^{1/\alpha}\left(|s|^{1/\alpha} + |x| + |t|^{1/\alpha} + |y|\right),$$

by (2.22) and (2.23), we have

$$\ell(|t + s|^{1/\alpha} + |x + y|) \lesssim_C \ell(|s|^{1/\alpha} + |x|) + \ell(|t|^{1/\alpha} + |y|).$$

In particular,

$$\frac{((t+s)^{1/\alpha} + |x+y|)^{d+\alpha}}{1 \wedge ((t+s)^{1/\alpha} + |x+y|)^{\beta_1 \wedge \beta_2}} \lesssim \frac{(s^{1/\alpha} + |x|)^{d+\alpha}}{1 \wedge (s^{1/\alpha} + |x|)^{\beta_1}} + \frac{(t^{1/\alpha} + |y|)^{d+\alpha}}{1 \wedge (t^{1/\alpha} + |y|)^{\beta_2}}.$$

Hence,

$$\frac{1 \wedge (s^{1/\alpha} + |x|)^{\beta_1}}{(s^{1/\alpha} + |x|)^{d+\alpha}} \times \frac{1 \wedge (t^{1/\alpha} + |y|)^{\beta_2}}{(t^{1/\alpha} + |y|)^{d+\alpha}}$$

$$\lesssim_C \left[\frac{1 \wedge (s^{1/\alpha} + |x|)^{\beta_1}}{(s^{1/\alpha} + |x|)^{d+\alpha}} + \frac{1 \wedge (t^{1/\alpha} + |y|)^{\beta_2}}{(t^{1/\alpha} + |y|)^{d+\alpha}}\right] \cdot \frac{1 \wedge ((t+s)^{1/\alpha} + |x+y|)^{\beta_1 \wedge \beta_2}}{((t+s)^{1/\alpha} + |x+y|)^{d+\alpha}}.$$

By this, the desired estimates follow by (2.17), (2.18) and (2.8). $\qquad \square$

2.4. Density estimate. — Let $a : \mathbb{R}_+ \to \mathbb{R}^d \otimes \mathbb{R}^d$ be a measurable $d \times d$-matrix-valued function satisfying the non-degeneracy condition

$$(2.24) \qquad \kappa_1^{-1} |\xi|^2 \leqslant |a(s)\xi|^2 \leqslant \kappa_1 |\xi|^2.$$

Fix $\alpha \in (0,2)$ and consider the following jump process

$$(2.25) \qquad X_{s,t}^a := \int_s^t a(r) \, dW_{S_r},$$

where $W$ is a $d$-dimensional Brownian motion and $S$ is an $\alpha/2$-stable subordinator independent of $W$, both defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Note that

$$X_{s,t}^a \overset{\text{(law)}}{=} (t-s)^{1/\alpha} X_{0,1}^{\widetilde{a}},$$

where

$$\widetilde{a}(r) := a(s + r(t-s)).$$

We have the following lemma that can be derived from the approach initially used in [3] (see also [4]). We provide below a proof for completeness.

Lemma 2.8. — *For any $0 \leqslant s < t < \infty$, $X_{s,t}^a$ has a smooth density $p_{s,t}^a(x)$ with the scaling property*

$$(2.26) \qquad p_{s,t}^a(x) = (t-s)^{-d/\alpha} p_{0,1}^{\widetilde{a}}((t-s)^{-1/\alpha} x),$$

*which enjoys the following estimates:*

$$(2.27) \qquad p_{s,t}^a(x) \asymp_{C_0} \varrho_{0,\alpha}^{(\alpha)}(t-s, x),$$

*and for any $j \in \mathbb{N}$,*

$$(2.28) \qquad |\nabla^j p_{s,t}^a(x)| \lesssim_{C_j} \varrho_{0,\alpha}^{(\alpha+j)}(t-s, x),$$

*where the constants only depend on $\kappa_1, d, \alpha$. Moreover, suppose that the integrand in (2.25) writes as $a_\xi(r)$ and smoothly depends on the parameter $\xi \in \mathbb{R}^d$ so that (2.24) holds uniformly and $\sup_{r,\xi} |\nabla_\xi^k a_\xi(r)| < \infty$ for any $k \in \mathbb{N}$. Let $p_{s,t}^{a_\xi}$ be the density of the integral in (2.25) associated with $a_\xi$. Then we have for $k \in \mathbb{N}$ and $j \in \mathbb{N}_0$,*

$$(2.29) \qquad |\nabla_\xi^k \nabla_x^j p_{s,t}^{a_\xi}(x)| \lesssim_{C_{j,k}} \varrho_{0,\alpha}^{(\alpha+j)}(t-s, x).$$

*Importantly, this last bound means that, the differentiation with respect to the parameter $\xi$ appearing in the diffusion coefficient $a_\xi$ does not yield an additional time singularity.*

*Proof.* — The two sided estimate (2.27) is well known (see e.g. [11]). We show (2.28). Without loss of generality, we assume $s = 0$ and write

$$X_t := \int_0^t a(r) \, dW_{S_r}.$$

Fix a càdlàg path $\ell_s$. Consider the following Gaussian random variable:

$$X_t^\ell := \int_0^t a(r) \, dW_{\ell_r}.$$

It has a density

$$g_t^{a,\ell}(x) = (2\pi)^{-d/2}\sqrt{\det\big((\mathbb{C}_t^{a,\ell})^{-1}\big)}\,\exp\{-\langle(\mathbb{C}_t^{a,\ell})^{-1}x,x\rangle/2\},$$

where

$$\mathbb{C}_t^{a,\ell} := \int_0^t (aa^*)(r)\,d\ell_r.$$

From the non-degeneracy assumption (2.24), we have

$$\langle(\mathbb{C}_t^{a,\ell})^{-1}x,x\rangle \asymp |x|^2/\ell_t, \quad \det\big((\mathbb{C}_t^{a,\ell})^{-1}\big) \asymp \ell_t^{-d},$$

and

$$|\nabla g_t^{a,\ell}(x)| \lesssim |x|/\ell_t \exp\{-\lambda|x|^2/\ell_t\}.$$

The density $p_{0,t}^a(x) =: p_t^a(x)$ of $X_t$ is given by

$$(2.30) \qquad p_t^a(x) = \mathbb{E}g_t^{a,S}(x).$$

The bound of (2.28) is direct from the Fourier representation of the density when $|x| \leqslant t^{1/\alpha}$. On the other hand, for $|x| > t^{1/\alpha}$, from the global bound on the law of the subordinator

$$\mu_{S_t}(dr) := \mathbb{P}\circ S_t^{-1}(dr) \lesssim t\,r^{-\alpha/2-1}\,dr,$$

it readily follows that

$$|\nabla p_t^a(x)| \leqslant \mathbb{E}|\nabla g_t^{a,S}(x)| \lesssim |x|\mathbb{E}(S_t^{-d/2-1}\exp\{-\lambda|x|^2/S_t\}) < +\infty.$$

Hence, from the bounded convergence theorem it holds that

$$|\nabla p_t^a(x)| \lesssim |x|\int_0^\infty r^{-(d+2)/2}e^{-\lambda|x|^2/r}\mu_{S_t}(dr),$$

and the integral expression in the r.h.s. precisely corresponds to the stable heat kernel in dimension $d+2$ at time $t$ and point $\widetilde{x} \in \mathbb{R}^{d+2}$ s.t. $|\widetilde{x}| = \sqrt{\lambda}|x|$. Thus, from (2.27),

$$|\nabla p_t^a(x)| \lesssim |x|t^{-(d+2)/\alpha}\frac{1}{(1+t^{-1/\alpha}|\widetilde{x}|)^{d+2+\alpha}}$$

$$\lesssim t(t^{1/\alpha}+|x|)^{-d-\alpha-1} = \varrho_{0,\alpha}^{(\alpha+1)}(t,x).$$

The approach is similar for higher order derivatives. This is also the case for (2.29) recalling that differentiating a Gaussian density with respect to the variance does not induce additional singularities. The proof is complete. $\qquad\square$

Remark 2.9. — We would like to emphasize that the gradient estimate (2.28) plays a crucial role for two-sided estimates due to the fact that for any $\beta \in [0,1]$,

$$|x|^\beta \varrho_{0,\alpha}^{(\alpha+1)}(t,x) = \frac{t|x|^\beta}{(t^{1/\alpha}+|x|)^{d+\alpha+1}} \leqslant \frac{t^{(\alpha+\beta-1)/\alpha}}{(t^{1/\alpha}+|x|)^{d+\alpha}} = \varrho_{0,\beta+\alpha-1}^{(\alpha)}(t,x).$$

In particular, for any $\beta \in [0,1]$,

$$(2.31) \qquad |x-\theta_{t,s}(x)|^\beta \phi_{0,\alpha}^{(\alpha+1)}(s,x,t,y) \leqslant \phi_{0,\beta+\alpha-1}^{(\alpha)}(s,x,t,y).$$

We carefully point out that the Gradient estimate (2.28), which remarkably emphasizes a concentration gain, does not hold for a general $\alpha$-stable like process [15]. This is also why, for the driving process in (1.1), we limit ourselves to the rotationally

invariant, and thus symmetric, $\alpha$-stable process and do not handle general $\alpha$-stable like processes.

The following lemma is taken from [11, Lem. 3.2 & 3.3].

Lemma 2.10. — *Under* (2.24), *there is a constant* $C = C(d, \alpha, \kappa_1) > 0$ *such that*

$$(2.32) \qquad |\nabla p_{s,t}^a - \nabla p_{s,t}^{\overline{a}}|(x) \lesssim_C \|a - \overline{a}\|_\infty \varrho_{0,\alpha}^{(\alpha+1)}(t - s, x).$$

*Also,*

$$(2.33) \qquad |\mathcal{D}^{(\alpha)} p_{s,t}^a|(x) \lesssim_C \varrho_{0,0}^{(\alpha)}(t - s, x),$$

*and*

$$(2.34) \qquad |\mathcal{D}^{(\alpha)}(p_{s,t}^a - p_{s,t}^{\overline{a}})|(x) \lesssim_C \|a - \overline{a}\|_\infty \varrho_{0,0}^{(\alpha)}(t - s, x).$$

*Moreover, we also have*

$$(2.35) \quad \int_{\mathbb{R}^d} |\delta_{p_{s,t}^a}^{(2)}(x_1; z) - \delta_{p_{s,t}^a}^{(2)}(x_2; z)| \frac{\mathrm{d}z}{|z|^{d+\alpha}}$$
$$\lesssim_C \left( \frac{|x_1 - x_2|}{(t - s)^{1/\alpha}} \wedge 1 \right) \left( \sum_{i=1,2} \varrho_{0,0}^{(\alpha)}(t - s, x_i) \right).$$

*Proof.* — From the scaling property (2.26), it suffices to consider $s = 0$ and $t = 1$. Note that

$$|\delta_{p_1^a}^{(2)}(x; z)| = |p_1^a(x + z) + p_1^a(x - z) - 2p_1^a(x)|$$
$$\lesssim (|z|^2 \wedge 1)(\varrho^{(\alpha)}(x + z) + \varrho^{(\alpha)}(x - z) + \varrho^{(\alpha)}(x)).$$

By elementary calculations, one sees that

$$(2.36) \qquad \int_{\mathbb{R}^d} \varrho^{(\alpha)}(x + z) \frac{(|z|^2 \wedge 1) \, \mathrm{d}z}{|z|^{d+\alpha}} \lesssim_C \varrho^{(\alpha)}(x).$$

Thus (2.33) follows. As for (2.34) and (2.35), they can be derived similarly to [10, Lem. 2.7 & 2.8]. The statement (2.32) can also be derived from the arguments developed therein. We omit the details. $\qquad \square$

## 3. Heat kernel of nonlocal operators with smooth coefficients

In this section we assume that $(\mathbf{H}_b^\beta)$ and $(\mathbf{H}_a^\gamma)$ hold and additionally that

$$(3.1) \qquad \|\nabla^j b\|_\infty + \|\nabla^j a\|_\infty < \infty, \quad j \in \{0\} \cup \mathbb{N}.$$

We again emphasize that we here assume that the coefficients are smooth and the drift is bounded. This last point precisely allows to derive that Duhamel like expansions holds for the semigroup (see equations (3.2) and (3.3) below). We will then first remove the smoothness assumption in Section 5.1 and the boundedness assumption on the drift in Section 5.2.

We shall denote

$$\mathscr{C} := \big\{(b, a) \text{ satisfying } (\mathbf{H}_b^\beta), (\mathbf{H}_a^\gamma) \text{ with common bounds } \kappa_0, \kappa_1 \text{ and } (3.1)\big\}.$$

Under $(\mathbf{H}_a^\gamma)$ and (3.1), for each $(s,x) \in \mathbb{R}_+ \times \mathbb{R}^d$, it is well known that there is a unique solution $(X_{s,t}(x))_{t\geqslant s}$ to SDE (1.1), and $X_{s,t}(x)$ has for $t > s$ a density $p(s,x,t,y)$ so that (cf. [13, 7])

$$P_{s,t}f(x) := \mathbb{E}f(X_{s,t}(x)) = \int_{\mathbb{R}^d} f(y)p(s,x,t,y)\,\mathrm{d}y, \quad f \in L^\infty(\mathbb{R}^d).$$

The density is also a mild solution of the Kolmogorov equation in the sense that for all $\varphi \in C_0^2(\mathbb{R}^d)$

$$(3.2) \qquad P_{s,t}\varphi(x) = \varphi(x) + \int_s^t P_{s,r}\mathscr{L}_r\varphi(x)\,\mathrm{d}r.$$

Moreover, by Schauder's estimate for nonlocal parabolic equations [39, Th. 3.5], we also have $P_{s,t}\varphi \in C_b^\infty(\mathbb{R}^d)$, and $P_{s,t}\varphi$ solves the following backward Kolmogorov equation

$$(3.3) \qquad P_{s,t}\varphi(x) = \varphi(x) + \int_s^t \mathscr{L}_r P_{r,t}\varphi(x)\,\mathrm{d}r.$$

Fix $(\tau,\xi) \in [s,t] \times \mathbb{R}^d$. Consider the following freezing process

$$X_{s,t}^{(\tau,\xi)} := x + \int_s^t b_{|r-\tau|^{1/\alpha}}(r, \theta_{\tau,r}(\xi))\,\mathrm{d}r + \int_s^t a(r, \theta_{\tau,r}(\xi))\,\mathrm{d}L_r^{(\alpha)}.$$

By Lemma 2.8, the density of $X_{s,t}^{(\tau,\xi)}$ is given by

$$(3.4) \qquad \widetilde{p}^{(\tau,\xi)}(s,x,t,y) = p_{s,t}^{a^{(\tau,\xi)}}\left(x - y + \int_s^t b_{|r-\tau|^{1/\alpha}}(r, \theta_{\tau,r}(\xi))\,\mathrm{d}r\right),$$

where $a^{(\tau,\xi)}(r) := a(r, \theta_{\tau,r}(\xi))$ and $p_{s,t}^{a^{(\tau,\xi)}}$ is the density of $\int_s^t a^{(\tau,\xi)}(r)\,\mathrm{d}L_r^{(\alpha)}$ given in Lemma 2.8. In particular,

$$(3.5) \qquad \partial_s \widetilde{p}^{(\tau,\xi)}(s,x,t,y) + \widetilde{\mathscr{L}}_s^{(\tau,\xi)}\widetilde{p}^{(\tau,\xi)}(s,\cdot,t,y)(x) = 0,$$

where

$$\widetilde{\mathscr{L}}_s^{(\tau,\xi)}f(x) := \tfrac{1}{2}\widetilde{\mathcal{L}}_s^{(\tau,\xi)}f(x) + b_{|s-\tau|^{1/\alpha}}(s, \theta_{\tau,s}(\xi)) \cdot \nabla f(x)$$

and

$$\widetilde{\mathcal{L}}_s^{(\tau,\xi)}f(x) = \int_{\mathbb{R}^d} \delta_f^{(2)}(x;z)\frac{\kappa(s, \theta_{\tau,s}(\xi), z)}{|z|^{d+\alpha}}\,\mathrm{d}z,$$

with

$$\kappa(s, \theta_{s,\tau}(\xi), z) := \frac{\det(a^{-1}(s, \theta_{\tau,s}(\xi)))|z|^{d+\alpha}}{|a^{-1}(s, \theta_{\tau,s}(\xi))z|^{d+\alpha}}.$$

For simplicity, we shall write

$$\mathscr{A}_s^{(\tau,\xi)}f(x) := (\mathscr{L}_s - \widetilde{\mathscr{L}}_s^{(\tau,\xi)})f(x) = \mathscr{K}_s^{(\tau,\xi)}f(x) + \mathscr{B}_s^{(\tau,\xi)}f(x),$$

where

$$(3.6) \qquad \mathscr{K}_s^{(\tau,\xi)}f(x) := \tfrac{1}{2}(\mathcal{L}_s - \widetilde{\mathcal{L}}_s^{(\tau,\xi)})f(x),$$

and

$$(3.7) \qquad \mathscr{B}_s^{(\tau,\xi)}f(x) := \big(b(s,x) - b_{|s-\tau|^{1/\alpha}}(s, \theta_{\tau,s}(\xi))\big) \cdot \nabla f(x).$$

Let us introduce the corresponding frozen semi-group:

$$(3.8) \qquad \widetilde{P}_{s,t}^{(\tau,\xi)}f(x) := \mathbb{E}f(X_{s,t}^{(\tau,\xi)}(x)).$$

We have the following Duhamel type representation formula:

Lemma 3.1. — *For any $f \in C_b^\infty(\mathbb{R}^d)$ and $(\tau, \xi) \in [s,t] \times \mathbb{R}^d$, it holds that*

$$P_{s,t}f = \widetilde{P}_{s,t}^{(\tau,\xi)}f + \int_s^t P_{s,r}\mathscr{A}_r^{(\tau,\xi)}\widetilde{P}_{r,t}^{(\tau,\xi)}f \, \mathrm{d}r = \widetilde{P}_{s,t}^{(\tau,\xi)}f + \int_s^t \widetilde{P}_{s,r}^{(\tau,\xi)}\mathscr{A}_r^{(\tau,\xi)}P_{r,t}f \, \mathrm{d}r.$$

*Proof.* — We drop for the proof the superscript $(\tau, \xi)$ for notational simplicity. From (3.2) and (3.5),

$$\partial_t P_{s,t}f = P_{s,t}\mathscr{L}_t f, \quad \partial_s \widetilde{P}_{s,t}f = -\widetilde{\mathscr{L}_s}\widetilde{P}_{s,t}f.$$

Note that $f \in C_b^\infty$, $\partial_r \widetilde{P}_{r,t}f \in C_b^\infty$. This indeed follows from the smoothness of $f$ and equation (3.5) through integration by parts and using the self-adjoint property of the frozen non-local operator. We thus have by the chain rule,

$$\partial_r(P_{s,r}\widetilde{P}_{r,t}f) = P_{s,r}\mathscr{L}_r\widetilde{P}_{r,t}f - P_{s,r}\widetilde{\mathscr{L}_r}\widetilde{P}_{r,t}f = P_{s,r}\mathscr{A}_r\widetilde{P}_{r,t}f.$$

Integrating both sides from $s$ to $t$ with respect to $r$ yields

$$P_{s,t}f = \widetilde{P}_{s,t}f + \int_s^t P_{s,r}\mathscr{A}_r\widetilde{P}_{r,t}f \, \mathrm{d}r.$$

Similarly, by (3.3), one can show that

$$\widetilde{P}_{s,t}f = P_{s,t}f - \int_s^t \widetilde{P}_{s,r}\mathscr{A}_r P_{r,t}f \, \mathrm{d}r.$$

The proof is complete.                                                          □

By Lemma 3.1, we have for each $(\tau, \xi) \in [s,t] \times \mathbb{R}^d$ and $x, y \in \mathbb{R}^d$,

$$p(s,x,t,y) = \widetilde{p}^{(\tau,\xi)}(s,x,t,y) + \int_s^t \int_{\mathbb{R}^d} p(s,x,r,z)\mathscr{A}_r^{(\tau,\xi)}\widetilde{p}^{(\tau,\xi)}(r,\cdot,t,y)(z) \, \mathrm{d}z \, \mathrm{d}r.$$

In particular, if we take $(\tau, \xi) = (t, y)$ and define

$$(3.9) \qquad p_0(s,x,t,y) := \widetilde{p}^{(t,y)}(s,x,t,y) = p_{s,t}^{a^{(t,y)}}(x - \theta_{t,s}(y)),$$

then we obtain the *forward* representation,

$$(3.10) \qquad p(s,x,t,y) = p_0(s,x,t,y) + \int_s^t \int_{\mathbb{R}^d} p(s,x,r,z)\mathscr{A}_r^{(t,y)}p_0(r,\cdot,t,y)(z) \, \mathrm{d}z \, \mathrm{d}r.$$

Let

$$q_0(s,x,t,y) := \mathscr{A}_s^{(t,y)}p_0(s,\cdot,t,y)(x),$$

and define recursively for $n \geqslant 1$,

$$(3.11) \qquad q_n := q_0 \otimes q_{n-1}, \quad q = \sum_{n=0}^\infty q_n.$$

By iteration, we *formally* obtain from (3.10) and (3.11),

$$(3.12) \qquad p = p_0 + p \otimes q_0 = p_0 + \sum_{n=0}^\infty p_0 \otimes q_n = p_0 + p_0 \otimes q.$$

The following lemma is a direct consequence of (3.9), (2.33) and (2.28).

LEMMA 3.2. — *For any $\alpha \in (0,2)$ and $j = 0, 1, \cdots$, we have*

$$(3.13) \qquad |\nabla^j p_0(s, \cdot, t, y)|(x) \lesssim \phi_{0,\alpha}^{(\alpha+j)}(s, x, t, y)$$

*and*

$$(3.14) \qquad |\mathcal{D}^{(\alpha)} p_0(s, \cdot, t, y)|(x) \lesssim \phi_{0,0}^{(\alpha)}(s, x, t, y).$$

The following lemma corresponds to [9, Th. 3.1].

LEMMA 3.3. — *The series $q = \sum_{n=0}^{\infty} q_n$ is absolutely convergent, and for each $s < t$, $(x,y) \mapsto q(s,x,t,y)$ is equi-continuous in $(b,a) \in \mathscr{C}$. Moreover, for any $T > 0$, there is a constant $C = C(T, \Theta) > 0$ such that for all $(s,x,t,y) \in \mathbb{D}_T$,*

$$(3.15) \qquad |q(s,x,t,y)| \lesssim_C \left( \phi_{\gamma_0,0}^{(\alpha)} + \phi_{0,\gamma_0}^{(\alpha)} \right)(s,x,t,y),$$

*where $\gamma_0 := (\alpha + \beta - 1) \wedge \gamma$, and for any $\gamma_1 \in (0, \gamma_0)$,*

$$(3.16) \quad |q(s,x,t,y) - q(s,x',t,y)| \lesssim_C (|x - x'|^{\gamma_1} \wedge 1)$$
$$\times \left( \left( \phi_{\gamma_0,-\gamma_1}^{(\alpha)} + \phi_{0,\gamma_0-\gamma_1}^{(\alpha)} \right)(s,x,t,y) + \left( \phi_{\gamma_0,-\gamma_1}^{(\alpha)} + \phi_{0,\gamma_0-\gamma_1}^{(\alpha)} \right)(s,x',t,y) \right).$$

*Proof*

(i) First of all, note that by (1.5),

$$|\kappa(s,x,z) - \kappa(s, \theta_{t,s}(y), z)| \lesssim (|x - \theta_{t,s}(y)|^{\gamma} \wedge 1)$$

and by $(\mathbf{H}_b^\beta)$,

$$|b(s,x) - b_{|s-t|^{1/\alpha}}(s, \theta_{t,s}(y))| \lesssim |x - \theta_{t,s}(y)|^{\beta} + |x - \theta_{t,s}(y)| + |t - s|^{\beta/\alpha}.$$

Thus, we have by (3.14),

$$|\mathscr{K}_s^{(t,y)} p_0(s, \cdot, t, y)(x)| \lesssim \phi_{\gamma,0}^{(\alpha)}(s,x,t,y),$$

and by (3.13) and (2.31),

$$|\mathscr{B}_s^{(t,y)} p_0(s, \cdot, t, y)(x)| \lesssim \phi_{0,\alpha+\beta-1}^{(\alpha)}(s,x,t,y).$$

So, for $\gamma_0 = \gamma \wedge (\alpha + \beta - 1)$,

$$|q_0(s,x,t,y)| \lesssim \left( \phi_{\gamma,0}^{(\alpha)} + \phi_{0,\alpha+\beta-1}^{(\alpha)} \right)(s,x,t,y) \lesssim \left( \phi_{\gamma_0,0}^{(\alpha)} + \phi_{0,\gamma_0}^{(\alpha)} \right)(s,x,t,y),$$

recalling (2.19) for the last inequality. Suppose now that for some $k \in \mathbb{N}$,

$$|q_{k-1}(s,x,t,y)| \leqslant C_k \left( \phi_{\gamma_0,(k-1)\gamma_0}^{(\alpha)} + \phi_{0,k\gamma_0}^{(\alpha)} \right)(s,x,t,y).$$

By Lemma 2.7, we have

$$|q_k(s,x,t,y)| \leqslant CC_k \left( \phi_{\gamma_0,0}^{(\alpha)} + \phi_{0,\gamma_0}^{(\alpha)} \right) \otimes \left( \phi_{\gamma_0,(k-1)\gamma_0}^{(\alpha)} + \phi_{0,k\gamma_0}^{(\alpha)} \right)(s,x,t,y)$$
$$\leqslant C_0 C_k \mathcal{B} \left( \tfrac{\gamma_0}{\alpha}, \tfrac{k\gamma_0}{\alpha} \right) \left( \phi_{\gamma_0,k\gamma_0}^{(\alpha)} + \phi_{0,(k+1)\gamma_0}^{(\alpha)} \right)(s,x,t,y).$$

Hence,

$$C_{k+1} = C_0 C_k \mathcal{B} \left( \tfrac{\gamma_0}{\alpha}, \tfrac{k\gamma_0}{\alpha} \right).$$

From the relation $\mathcal{B}(\gamma, \beta) = \Gamma(\gamma)\Gamma(\beta)/\Gamma(\gamma + \beta)$, where $\Gamma$ is the usual Gamma function, we obtain

$$C_k = C_0^k \prod_{i=1}^{k-1} \mathcal{B}(\tfrac{\gamma_0}{\alpha}, \tfrac{(k-1)\gamma_0}{\alpha}) = \frac{(C_0 \Gamma(\gamma_0/\alpha))^k}{\Gamma(k\gamma_0/\alpha)},$$

with the usual convention that $\prod_{i=1}^0 = 1$. Thus

$$\sum_{k=0}^{\infty} |q_k(s,x,t,y)| \leqslant \sum_{k=0}^{\infty} \frac{(C_0 \Gamma(\gamma_0/\alpha))^k}{\Gamma(k\gamma_0/\alpha)} \left( \phi_{\gamma_0, k\gamma_0}^{(\alpha)} + \phi_{0,(k+1)\gamma_0}^{(\alpha)} \right)(s,x,t,y)$$

$$\leqslant \sum_{k=0}^{\infty} \frac{(C_0 \Gamma(\gamma_0/\alpha))^k}{\Gamma(k\gamma_0/\alpha)} \left( \phi_{\gamma_0, 0}^{(\alpha)} + \phi_{0,\gamma_0}^{(\alpha)} \right)(s,x,t,y).$$

This gives (3.15).

(ii) For fixed $s < t$, by Lemma 2.8 and the definition of $q_0$, one sees that $(x,y) \mapsto q_0(s,x,t,y)$ is equi-continuous in $(b,a) \in \mathscr{C}$. Furthermore, it follows by induction that, for each $k \in \mathbb{N}$, $(x,y) \mapsto q_k(s,x,t,y)$ is also equi-continuous in $(b,a) \in \mathscr{C}$. Hence, $(x,y) \mapsto q(s,x,t,y)$ is equi-continuous in $(b,a) \in \mathscr{C}$.

(iii) If $|x - x'| \geqslant (t-s)^{1/\alpha}$, then we have

$$|q_0(s,x,t,y)| \lesssim (|x-x'|^{\gamma_1} \wedge 1)(t-s)^{-\gamma_1/\alpha} \big( \phi_{\gamma_0,0}^{(\alpha)} + \phi_{0,\gamma_0}^{(\alpha)} \big)(s,x,t,y)$$

$$= (|x-x'|^{\gamma_1} \wedge 1) \big( \phi_{\gamma_0,-\gamma_1}^{(\alpha)} + \phi_{0,\gamma_0-\gamma_1}^{(\alpha)} \big)(s,x,t,y).$$

Next we assume $|x - x'| \leqslant (t-s)^{1/\alpha}$. In this case, it is easy to see from (2.15)–(2.16), that

$$(3.17) \qquad\qquad \phi_{0,0}^{(\eta)}(s,x,t,y) \asymp \phi_{0,0}^{(\eta)}(s,x',t,y), \quad \eta \geqslant 0.$$

By (2.33), (2.35) and (3.17), we have

$$|\mathscr{K}_s^{(t,y)} p_0(s,\cdot,t,y)(x) - \mathscr{K}_s^{(t,y)} p_0(s,\cdot,t,y)(x')|$$

$$\leqslant \|\kappa(\cdot,x,\cdot) - \kappa(\cdot,x',\cdot)\|_\infty \int_{\mathbb{R}^d} |\delta_{p_0(s,\cdot,t,y)}^{(2)}(x;z)| \frac{\mathrm{d}z}{|z|^{d+\alpha}}$$

$$+ \|\kappa(\cdot,x,\cdot) - \kappa(\cdot,\theta_{t,s}(y),\cdot)\|_\infty$$

$$\times \int_{\mathbb{R}^d} |\delta_{p_0(s,\cdot,t,y)}^{(2)}(x;z) - \delta_{p_0(s,\cdot,t,y)}^{(2)}(x';z)| \frac{\mathrm{d}z}{|z|^{d+\alpha}}$$

$$\leqslant (|x-x'|^\gamma \wedge 1) \phi_{0,0}^{(\alpha)}(s,x,t,y) + (|x - \theta_{t,s}(y)|^\gamma \wedge 1)$$

$$\times \left( \frac{|x-x'|}{(t-s)^{1/\alpha}} \wedge 1 \right) \big( \phi_{0,0}^{(\alpha)}(s,x,t,y) + \phi_{0,0}^{(\alpha)}(s,x',t,y) \big)$$

$$\lesssim (|x-x'|^{\gamma_1} \wedge 1) \big( \phi_{0,\gamma-\gamma_1}^{(\alpha)}(s,x,t,y) + \phi_{\gamma,-\gamma_1}^{(\alpha)}(s,x,t,y) \big).$$

Moreover, by (3.13), (3.17) and (2.31), we also have

$$|\mathscr{B}_s^{(t,y)}p_0(s,\cdot,t,y)(x) - \mathscr{B}_s^{(t,y)}p_0(s,\cdot,t,y)(x')|$$

$$\leqslant |b(s,x) - b(s,x')| \cdot |\nabla p_0(s,\cdot,t,y)|(x')$$

$$+ \left|b(s,x) - b_{|s-t|^{1/\alpha}}(s,\theta_{t,s}(y))\right| \cdot |\nabla p_0(s,\cdot,t,y)(x') - \nabla p_0(s,\cdot,t,y)(x)|$$

$$\lesssim |x-x'|^\beta \phi_{0,\alpha}^{(\alpha+1)}(s,x,t,y) + (|x-\theta_{t,s}(y)|^\beta + |t-s|^{\beta/\alpha})|x-x'|\phi_{0,\alpha}^{(\alpha+2)}(s,x,t,y)$$

$$\lesssim (|x-x'|^{\gamma_1}\wedge 1)\phi_{0,\alpha+\beta-1-\gamma_1}^{(\alpha)}(s,x,t,y).$$

Combining the above calculations and recalling $\gamma_0 = \gamma \wedge (\alpha+\beta-1)$, we obtain

$$|q_0(s,x,t,y) - q_0(s,x',t,y)| \lesssim_C (|x-x'|^{\gamma_1}\wedge 1)$$

$$\times \left( \left(\phi_{\gamma_0,-\gamma_1}^{(\alpha)} + \phi_{0,\gamma_0-\gamma_1}^{(\alpha)}\right)(s,x,t,y) + \left(\phi_{\gamma_0,-\gamma_1}^{(\alpha)} + \phi_{0,\gamma_0-\gamma_1}^{(\alpha)}\right)(s,x',t,y)\right).$$

Using this last estimate, equation (3.16) follows from the same iterative argument as in (i). □

REMARK 3.4. — This lemma allows to iterate the representation (3.10) which leads to the representation (3.12) of the density.

We now aim at proving the following a priori estimate about $p(s,x,t,y)$.

THEOREM 3.5. — *Under* $(\mathbf{H}_a^\gamma)$, $(\mathbf{H}_b^\beta)$ *and* (3.1), *for each* $0 \leqslant s < t < \infty$, $X_{s,t}(x)$ *admits a density* $p(s,x,t,y)$ *that is equi-continuous in* $(b,a) \in \mathscr{C}$ *as a function of* $x,y \in \mathbb{R}^d$, *and there is a constant* $C = C(T,\Theta) > 0$ *so that for all* $(s,x,t,y) \in \mathbb{D}_T$,

$$p(s,x,t,y) \asymp_C \phi_{0,\alpha}^{(\alpha)}(s,x,t,y).$$

*Proof.* — Note that by (3.9), (2.27) and (2.7),

$$p_0(s,x,t,y) \asymp_C \phi_{0,\alpha}^{(\alpha)}(s,x,t,y).$$

By Lemma 2.7, we have

$$|p_0 \otimes q|(s,x,t,y) \lesssim_C (\phi_{0,\alpha+\gamma_0}^{(\alpha)} + \phi_{\gamma_0,\alpha}^{(\alpha)})(s,x,t,y).$$

The upper bound follows from (3.12).

Next we use Lemma 2.6 to show the lower bound estimate. Let $K$ be as in Lemma 2.6. Suppose that $|x - \theta_{t,s}(y)| \leqslant 2K(t-s)^{1/\alpha}$ (diagonal regime). Then we have

$$p(s,x,t,y) \geqslant p_0(s,x,t,y) - |p_0 \otimes q(s,x,t,y)|$$

$$\geqslant c_0 \phi_{0,\alpha}^{(\alpha)}(s,x,t,y) - (\phi_{0,\alpha+\gamma_0}^{(\alpha)} + \phi_{\gamma_0,\alpha}^{(\alpha)})(s,x,t,y)$$

$$\geqslant (c_0 - C_1(t-s)^{\gamma_0/\alpha})\phi_{0,\alpha}^{(\alpha)}(s,x,t,y),$$

recalling from (2.16) and (2.15) that, in the current diagonal regime,

$$(\phi_{0,\alpha+\gamma_0}^{(\alpha)} + \phi_{\gamma_0,\alpha}^{(\alpha)})(s,x,t,y) \leqslant C_1(t-s)^{\gamma_0/\alpha}\phi_{0,\alpha}^{(\alpha)}(s,x,t,y)$$

for the last inequality. In particular, if $t - s \leqslant \ell$ with $\ell$ small enough and $|x - \theta_{t,s}(y)| \leqslant 2K(t-s)^{1/\alpha}$, then

$$(3.18) \qquad p(s,x,t,y) \geqslant \frac{c_0}{2} \phi_{0,\alpha}^{(\alpha)}(s,x,t,y) \geqslant c_1(t-s)^{-d/\alpha}.$$

Next we prove the above estimate still holds for

$$|x - \theta_{t,s}(y)| \geqslant 2K(t-s)^{1/\alpha}.$$

Let $\varepsilon \in (0, 1/2)$ be as in Lemma 2.6 and small enough so that $2(1-\varepsilon)^{1/\alpha} \geqslant 1$. Let

$$r := s + \varepsilon(t-s), \quad B := \{z : |z - \theta_{t,r}(y)| \leqslant 2K(t-r)^{1/\alpha}\}.$$

Since $2(1-\varepsilon)^{1/\alpha} \geqslant 1$, we clearly have

$$B \supset \{z : |z - \theta_{t,r}(y)| \leqslant K(t-s)^{1/\alpha}\} =: B'.$$

Now from the Chapman-Kolmogorov equation, we have for $t - s \leqslant \ell$,

$$p(s,x,t,y) = \int_{\mathbb{R}^d} p(s,x,r,z) p(r,z,t,y) \, \mathrm{d}z \geqslant \int_B p(s,x,r,z) p(r,z,t,y) \, \mathrm{d}z$$

$$\geqslant \inf_{z \in B} p(r,z,t,y) \int_B p(s,x,r,z) \, \mathrm{d}z \overset{(3.18)}{\geqslant} c_1(t-r)^{-d/\alpha} \mathbb{P}(X_{s,r}(x) \in B)$$

$$\geqslant c_2(t-s)^{-d/\alpha} \mathbb{P}(X_{s,r}(x) \in B') \geqslant c_3(t-s)|x - \theta_{t,s}(y)|^{-d-\alpha},$$

where the last step is due to Lemma 2.6. Thus we obtain that for some $c_4 > 0$ and all $s, t \in [0, T]$,

$$p(s,x,t,y) \geqslant c_4 \phi_{0,\alpha}^{(\alpha)}(s,x,t,y), \quad t - s \leqslant \ell, \ x, y \in \mathbb{R}^d.$$

For $t - s > \ell$, the bound follows iteratively from the Chapman-Kolmogorov equation. The proof is complete. □

For the fractional derivative estimates, we need the following lemma.

Lemma 3.6. — *For $s < t$, let $h_{s,t}(x) := \int_{\mathbb{R}^d} p_0(s,x,t,y) \, \mathrm{d}y$. We have for some $C > 0$,*

$$|\mathscr{D}^{(\alpha)} h_{s,t}|(x) \lesssim_C (t-s)^{\gamma_0/\alpha - 1}, \ \gamma_0 := \gamma \wedge (\alpha + \beta - 1).$$

*Proof.* — By definition we have

$$|\mathscr{D}^{(\alpha)} h_{s,t}|(x) = \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \delta_{\widetilde{p}^{(t,y)}(s,\cdot,t,y)}^{(2)}(x;z) \, \mathrm{d}y \right| \frac{\mathrm{d}z}{|z|^{d+\alpha}}$$

$$= \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \delta_{p_{s,t}^{a(t,y)}}^{(2)}(x - \theta_{t,s}(y); z) \, \mathrm{d}y \right| \frac{\mathrm{d}z}{|z|^{d+\alpha}}$$

$$\leqslant \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \delta_{p_{s,t}^{a(t,y)} - p_{s,t}^{a(s,x)}}^{(2)}(x - \theta_{t,s}(y); z) \, \mathrm{d}y \right| \frac{\mathrm{d}z}{|z|^{d+\alpha}}$$

$$+ \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \delta_{p_{s,t}^{a(s,x)}}^{(2)}(x - \theta_{t,s}(y); z) \, \mathrm{d}y \right| \frac{\mathrm{d}z}{|z|^{d+\alpha}} =: I_1 + I_2.$$

For $I_1$, noting that by $(\mathbf{H}_a^\gamma)$ and Lemma 2.2,

$$|a(r, \theta_{s,r}(x)) - a(r, \theta_{t,r}(y))| \lesssim 1 \wedge |x - \theta_{t,s}(y)|^\gamma + |t-s|^{\gamma/\alpha},$$

we have

$$I_1 \leqslant \int_{\mathbb{R}^d} |\mathcal{D}^{(\alpha)}(p_{s,t}^{a^{(t,y)}} - p_{s,t}^{a^{(s,x)}})|(x - \theta_{t,s}(y))\,\mathrm{d}y$$

$$\overset{(2.34)}{\lesssim} \int_{\mathbb{R}^d} \left(\phi_{\gamma,0}^{(\alpha)} + \phi_{0,\gamma}^{(\alpha)}\right)(s,x,t,y)\,\mathrm{d}y \overset{(2.18)}{\lesssim} (t-s)^{\gamma/\alpha - 1}.$$

For $I_2$, by the change of variable we have

$$I_2 = \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \delta^{(2)}_{p_{s,t}^{a^{(s,x)}}}(x - y; z) \det(\nabla \theta_{s,t}^{-1}(y))\,\mathrm{d}y \right| \frac{\mathrm{d}z}{|z|^{d+\alpha}}$$

$$= \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \delta^{(2)}_{p_{s,t}^{a^{(s,x)}}}(x - y; z) \Big(\det(\nabla \theta_{s,t}^{-1}(y)) - 1\Big)\,\mathrm{d}y \right| \frac{\mathrm{d}z}{|z|^{d+\alpha}},$$

where we have used that

$$\int_{\mathbb{R}^d} p_{s,t}^{a^{(s,x)}}(x - y)\,\mathrm{d}y = 1 \implies \int_{\mathbb{R}^d} \delta^{(2)}_{p_{s,t}^{a^{(s,x)}}}(x - y)\,\mathrm{d}y = 0.$$

Thus by (i) of Lemma 2.2 and (2.33), we have

$$I_2 \lesssim (t-s)^{(\beta+\alpha-1)/\alpha} \int_{\mathbb{R}^d} |\mathcal{D}^{(\alpha)} p_{s,t}^{a^{(s,x)}}|(x - y)\,\mathrm{d}y$$

$$\lesssim (t-s)^{(\beta+\alpha-1)/\alpha} \int_{\mathbb{R}^d} \varrho_{0,0}^{(\alpha)}(t-s, x-y)\,\mathrm{d}y$$

$$\lesssim (t-s)^{(\beta-1)/\alpha} = (t-s)^{-1+(\alpha+\beta-1)/\alpha}.$$

The proof is complete. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Lemma 3.7 (Fractional derivative estimate). — *For any $T > 0$, we have for some $C = C(T, \Theta) > 0$,*

$$|\mathcal{D}^{(\alpha)}p(s, \cdot, t, y)|(x) \lesssim_C \phi_{0,0}^{(\alpha)}(s, x, t, y).$$

*Proof.* — Let $u = (s + t)/2$. By (3.12) and the definition of $\delta^{(2)}$, we have

$$\delta^{(2)}_{p(s,\cdot,t,y)}(x; \overline{z}) = \delta^{(2)}_{p_0(s,\cdot,t,y)}(x; \overline{z}) + \int_s^t \int_{\mathbb{R}^d} \delta^{(2)}_{p_0(s,\cdot,r,z)}(x; \overline{z})q(r, z, t, y)\,\mathrm{d}z\,\mathrm{d}r$$

$$= \delta^{(2)}_{p_0(s,\cdot,t,y)}(x; \overline{z}) + \int_s^u \left(\int_{\mathbb{R}^d} \delta^{(2)}_{p_0(s,\cdot,r,z)}(x; \overline{z})\,\mathrm{d}z\right) q(r, \theta_{s,r}(x), t, y)\,\mathrm{d}r$$

$$+ \int_s^u \int_{\mathbb{R}^d} \delta^{(2)}_{p_0(s,\cdot,r,z)}(x; \overline{z})(q(r, z, t, y) - q(r, \theta_{s,r}(x), t, y))\,\mathrm{d}z\,\mathrm{d}r$$

$$+ \int_u^t \int_{\mathbb{R}^d} \delta^{(2)}_{p_0(s,\cdot,r,z)}(x; \overline{z})q(r, z, t, y)\,\mathrm{d}z\,\mathrm{d}r.$$

With the notations of Lemma 3.6, set $h_{s,r}(x) = \int_{\mathbb{R}^d} p_0(s, x, r, z)\,\mathrm{d}z$. By (1.23) and the Fubini theorem, we have

$$
|\mathcal{D}^{(\alpha)} p(s, \cdot, t, y)|(x) \leqslant |\mathcal{D}^{(\alpha)} p_0(s, \cdot, t, y)|(x) + \int_s^u |\mathcal{D}^{(\alpha)} h_{s,r}|(x)|q(r, \theta_{s,r}(x), t, y)|\,\mathrm{d}r
$$

$$
+ \int_s^u \int_{\mathbb{R}^d} |\mathcal{D}^{(\alpha)} p_0(s, \cdot, r, z)|(x)|q(r, z, t, y) - q(r, \theta_{s,r}(x), t, y)|\,\mathrm{d}z\,\mathrm{d}r
$$

$$
+ \int_u^t \int_{\mathbb{R}^d} |\mathcal{D}^{(\alpha)} p_0(s, \cdot, r, z)|(x)|q(r, z, t, y)|\,\mathrm{d}z\,\mathrm{d}r
$$

$$
=: I_1(x) + I_2(x) + I_3(x) + I_4(x).
$$

For $I_1$, by (3.14) we have

$$
I_1(x) \lesssim \phi_{0,0}^{(\alpha)}(s, x, t, y).
$$

Recall

$$
\gamma_0 = (\alpha + \beta - 1) \wedge \gamma, \quad \gamma_1 \in (0, \gamma_0).
$$

For $I_2$, by Lemma 3.6, (3.15), (2.17) and (2.8), we have

$$
I_2(x) \lesssim \int_s^u (r - s)^{(\gamma_0/\alpha) - 1} \big(\phi_{\gamma_0,0}^{(\alpha)} + \phi_{0,\gamma_0}^{(\alpha)}\big)(r, \theta_{s,r}(x), t, y)\,\mathrm{d}r
$$

$$
\lesssim \left( \int_s^u (r - s)^{(\gamma_0/\alpha) - 1}\,\mathrm{d}r \right) \big(\phi_{\gamma_0,0}^{(\alpha)} + \phi_{0,\gamma_0}^{(\alpha)}\big)(s, x, t, y)
$$

$$
\lesssim \big(\phi_{\gamma_0,\gamma_0}^{(\alpha)} + \phi_{0,2\gamma_0}^{(\alpha)}\big)(s, x, t, y).
$$

For $I_3$, by (3.14), (3.16) and (2.21), we have

$$
I_3(x) \lesssim \int_s^u \int_{\mathbb{R}^d} \phi_{\gamma_1,0}^{(\alpha)}(s, x, r, z)\big(\phi_{\gamma_0,-\gamma_1}^{(\alpha)} + \phi_{0,\gamma_0-\gamma_1}^{(\alpha)}\big)(r, z, t, y)\,\mathrm{d}z\,\mathrm{d}r
$$

$$
+ \int_s^u \int_{\mathbb{R}^d} \phi_{\gamma_1,0}^{(\alpha)}(s, x, r, z)\,\mathrm{d}z\big(\phi_{\gamma_0,-\gamma_1}^{(\alpha)} + \phi_{0,\gamma_0-\gamma_1}^{(\alpha)}\big)(s, x, t, y)\,\mathrm{d}r
$$

$$
\lesssim \big(\phi_{\gamma_0,0}^{(\alpha)} + \phi_{0,\gamma_0}^{(\alpha)}\big)(s, x, t, y).
$$

For $I_4$, by (3.14), (3.15) and (2.20), we have

$$
I_4(x) \lesssim \int_u^t \big(\phi_{0,0}^{(\alpha)} \odot (\phi_{\gamma_0,0}^{(\alpha)} + \phi_{0,\gamma_0}^{(\alpha)})\big)_r(s, x, t, y)\,\mathrm{d}r \lesssim \phi_{0,0}^{(\alpha)}(s, x, t, y).
$$

Combining the above estimates, we complete the proof. $\qquad\square$

## 4. A priori gradient estimates

The aim of this section is to show the following a priori gradient estimate.

Theorem 4.1. — *Under* $(\mathbf{H}_b^\beta)$, $(\mathbf{H}_a^\gamma)$ *and* (3.1), *for any* $T > 0$, *there is a constant* $C = C(T, \Theta) > 0$ *such that for all* $f \in \mathcal{B}_b(\mathbb{R}^d)$, $0 \leqslant s < t \leqslant T$ *and* $x \in \mathbb{R}^d$,

$$
(4.1) \qquad\qquad |\nabla P_{s,t} f(x)| \lesssim_C (t - s)^{-1/\alpha} P_{s,t}|f|(x).
$$

*Moreover,* $x \mapsto \nabla P_{s,t} f(x)$ *is equi-continuous in* $(b, a) \in \mathscr{C}$.

We again emphasize that Theorem 4.1 gives that the constants in the gradient estimates actually *do not* depend on the smoothness of the coefficients, neither on the boundedness of the drift assumed in (3.1).

We shall prove this theorem for $\alpha \in [1, 2)$ and $\alpha \in (0, 1)$ separately by different methods.

4.1. CRITICAL AND SUBCRITICAL CASES: $\alpha \in [1, 2)$. — In this subsection we start from the series expansion (3.12) for the density to derive the estimate

$$(4.2) \qquad |\nabla_x p(s, x, t, y)| \lesssim_{C_3} \phi_{0,\alpha-1}^{(\alpha)}(s, x, t, y),$$

when $(\mathbf{H}_b^\beta)$, $(\mathbf{H}_a^\gamma)$ and (3.1) are in force and $\alpha \in [1, 2)$. This precisely gives (4.1). We recall that, with the notations of Section 3:

$$p(s, x, t, y) = p_0(s, x, t, y) + (p_0 \otimes q)(s, x, t, y).$$

We will first assume that $\alpha \in (1, 2)$ and handle the critical case $\alpha = 1$ afterward through a domination argument. Now, for $\alpha \in (1, 2)$, $u = (s + t)/2$ and $\xi = \theta_{s,r}(x)$,

$$
\begin{aligned}
(4.3) \qquad \nabla_x p(s, x, t, y) = {} & \nabla_x p_0(s, x, t, y) + \int_u^t \int_{\mathbb{R}^d} \nabla_x p_0(s, x, r, z) q(r, z, t, y) \, \mathrm{d}z \, \mathrm{d}r \\
& + \int_s^u \int_{\mathbb{R}^d} (\nabla_x p_0 - \nabla_x \widetilde{p}^{(r,\xi)})(s, x, r, z) q(r, z, t, y) \, \mathrm{d}z \, \mathrm{d}r \\
& + \int_s^u \int_{\mathbb{R}^d} \nabla_x \widetilde{p}^{(r,\xi)}(s, x, r, z)(q(r, z, t, y) - q(r, \xi, t, y)) \, \mathrm{d}z \, \mathrm{d}r \\
=: {} & G_1(s, x, t, y) + G_2(s, x, t, y) + G_3(s, x, t, y) + G_4(s, x, t, y),
\end{aligned}
$$

where for the last term, we have used precisely the cancellation property

$$\int_{\mathbb{R}^d} \nabla_x \widetilde{p}^{(r,\xi)}(s, x, r, z) \, \mathrm{d}z = 0.$$

For $G_1$, by (3.13) we clearly have

$$|G_1(s, x, t, y)| \lesssim \phi_{0,\alpha}^{(\alpha+1)}(s, x, t, y) \leqslant \phi_{0,\alpha-1}^{(\alpha)}(s, x, t, y),$$

using Remark 2.9, equation (2.31), for the last inequality. For $G_2$, by (3.13), (3.15) and (2.21), we have

$$
\begin{aligned}
|G_2(s, x, t, y)| & \leqslant \int_u^t \int_{\mathbb{R}^d} \phi_{0,\alpha}^{(\alpha+1)}(s, x, r, z) |q(r, z, t, y)| \, \mathrm{d}z \, \mathrm{d}r \\
& \lesssim (t - s)^{-1/\alpha} \phi_{0,\alpha}^{(\alpha)} \otimes \left( \phi_{\gamma_0,0}^{(\alpha)} + \phi_{0,\gamma_0}^{(\alpha)} \right)(s, x, t, y) \\
& \lesssim (t - s)^{-1/\alpha} \phi_{0,\alpha+\gamma_0}^{(\alpha)}(s, x, t, y) = \phi_{0,\alpha+\gamma_0-1}^{(\alpha)}(s, x, t, y).
\end{aligned}
$$

For $G_3$, noting that by (3.4),

$$\nabla_x p_0(s, x, r, z) = \nabla_x p_{s,r}^{a^{(r,z)}}\left( x - z + \int_s^r b_{|r'-r|^{1/\alpha}}(r', \theta_{r,r'}(z)) \, \mathrm{d}r' \right),$$

$$\nabla_x \widetilde{p}^{(r,\xi)}(s, x, r, z) = \nabla_x p_{s,r}^{a^{(r,\xi)}}\left( x - z + \int_s^r b_{|r'-r|^{1/\alpha}}(r', \theta_{r,r'}(\xi)) \, \mathrm{d}r' \right),$$

by (2.32), (2.28), (1.20) and (2.5), one finds that

$$
\begin{aligned}
|\nabla_x p_0 - \nabla_x \widetilde{p}^{(r,\xi)}|(s,x,r,z) &\lesssim \phi_{0,\alpha}^{(\alpha+1)}(s,x,r,z)(1 \wedge |z - \theta_{s,r}(x)|^\gamma) \\
&\quad + \phi_{0,\alpha}^{(\alpha+2)}(s,x,r,z)\big(|z - \theta_{s,r}(x)|^\beta + (r-s)^{\beta/\alpha}\big)(r-s) \\
&\underset{(2.31)}{\lesssim} \big(\phi_{0,\alpha+\gamma-1}^{(\alpha)} + \phi_{0,2\alpha+\beta-2}^{(\alpha)}\big)(s,x,r,z) \\
&\lesssim \phi_{0,\alpha-1+\gamma_0}^{(\alpha)}(s,x,r,z),
\end{aligned}
$$

where $\gamma_0 = \gamma \wedge (\alpha + \beta - 1)$. Therefore, due to $\alpha \in [1,2)$, by (2.21),

$$
\begin{aligned}
\text{(4.4)} \qquad |G_3(s,x,t,y)| &\lesssim \phi_{0,\alpha-1+\gamma_0}^{(\alpha)} \otimes \big(\phi_{\gamma_0,0}^{(\alpha)} + \phi_{0,\gamma_0}^{(\alpha)}\big)(s,x,t,y) \\
&\lesssim \phi_{0,\alpha-1+2\gamma_0}^{(\alpha)}(s,x,t,y) \leqslant \phi_{0,\alpha-1}^{(\alpha)}(s,x,t,y).
\end{aligned}
$$

For $G_4$, by (3.4), (2.28) and (3.16) we have for $\gamma_1 \in (0,\gamma_0)$,

$$
\begin{aligned}
|G_4(s,x,t,y)| &\leqslant \int_s^u \int_{\mathbb{R}^d} |\nabla_x \widetilde{p}^{(r,\xi)}(s,x,r,z)|\, |(q(r,z,t,y) - q(r,\xi,t,y))|\, \mathrm{d}z\, \mathrm{d}r \\
&\lesssim \int_s^u \mathrm{d}r \int_{\mathbb{R}^d} \mathrm{d}z\, \phi_{0,\alpha-1}^{(\alpha)}(s,x,r,z)(1 \wedge |z-\xi|^{\gamma_1}) \\
&\quad \times \Big[\big(\phi_{\gamma_0,-\gamma_1}^{(\alpha)} + \phi_{0,\gamma_0-\gamma_1}^{(\alpha)}\big)(r,z,t,y) + \big(\phi_{\gamma_0,-\gamma_1}^{(\alpha)} + \phi_{0,\gamma_0-\gamma_1}^{(\alpha)}\big)(r,\xi,t,y)\Big].
\end{aligned}
$$

Since $t - r \asymp t - s$ for $r \in [s,u]$ and $\xi = \theta_{s,r}(x)$, from (2.7) in Lemma 2.2, it holds

$$
\big(\phi_{\gamma_0,-\gamma_1}^{(\alpha)} + \phi_{0,\gamma_0-\gamma_1}^{(\alpha)}\big)(r,\xi,t,y) \lesssim \big(\phi_{\gamma_0,-\gamma_1}^{(\alpha)} + \phi_{0,\gamma_0-\gamma_1}^{(\alpha)}\big)(s,x,t,y).
$$

Thus by (2.21), we eventually have

$$
\begin{aligned}
|G_4(s,x,t,y)| &\lesssim \int_s^u \mathrm{d}r \int_{\mathbb{R}^d} \mathrm{d}z\, \phi_{0,\alpha+\gamma_1-1}^{(\alpha)}(s,x,r,z) \\
&\quad \times \Big[\big(\phi_{\gamma_0,-\gamma_1}^{(\alpha)} + \phi_{0,\gamma_0-\gamma_1}^{(\alpha)}\big)(r,z,t,y) + \big(\phi_{\gamma_0,-\gamma_1}^{(\alpha)} + \phi_{0,\gamma_0-\gamma_1}^{(\alpha)}\big)(s,x,t,y)\Big] \\
&\lesssim \big(\phi_{0,\gamma_0-\gamma_1}^{(\alpha)} + \phi_{\gamma_0,\alpha-1}^{(\alpha)}\big)(s,x,t,y).
\end{aligned}
$$

Combining the above calculations, we obtain (4.2) in the case $\alpha \in (1,2)$.

We mention that in the critical case $\alpha = 1$, the previous terms $G_1$ and $G_2$ in (4.3) are well-defined (no time singularity) and controlled similarly. However, some care is needed to justify that

$$
\text{(4.5)} \quad \nabla_x \int_s^u \iint_{\mathbb{R}^d} p_0(s,x,r,z)q(r,z,t,y)\, \mathrm{d}z\, \mathrm{d}r = \int_s^u \iint_{\mathbb{R}^d} \nabla_x p_0(s,x,r,z)q(r,z,t,y)\, \mathrm{d}z\, \mathrm{d}r.
$$

The previous controls on $G_3, G_4$ can be used to prove that setting

$$
\int_{s+\varepsilon}^u \int_{\mathbb{R}^d} \nabla_x p_0(s,x,r,z)q(r,z,t,y)\, \mathrm{d}z\, \mathrm{d}r =: \int_{s+\varepsilon}^u \Gamma_{s,x,t,y}(r)\, \mathrm{d}r,
$$

the function $r \mapsto \Gamma_{s,x,t,y}(r)$ is integrable. Hence, a domination argument yields (4.5) which in turns establishes that (4.3) still holds, as well as the associated estimates, in the critical case $\alpha = 1$.

Let us now turn to the equicontinuity part of the theorem. From the dominated convergence theorem and the above calculations, it is easy to see that

$$\lim_{x \to x_0} \sup_{(b,a) \in \mathscr{C}} |G_i^{b,a}(s,x,t,y) - G_i^{b,a}(s,x_0,t,y)| = 0, \quad i = 1,2,3,4,$$

where $G_i^{b,a}$ are defined as above through the coefficients $b, a$. For instance,

$$\lim_{x \to x_0} \sup_{(b,a) \in \mathscr{C}} |G_2^{b,a}(s,x,t,y) - G_2^{b,a}(s,x_0,t,y)|$$

$$\leqslant \int_u^t \lim_{x \to x_0} \sup_{(b,a) \in \mathscr{C}} \int_{\mathbb{R}^d} |\nabla_x p_0(s,x,r,z) - \nabla_x p_0(s,x_0,r,z)| \, |q(r,z,t,y)| \, \mathrm{d}z \, \mathrm{d}r,$$

and for each $r \in (u,t)$, by (3.15) and (3.13),

$$\lim_{x \to x_0} \sup_{(b,a) \in \mathscr{C}} \int_{\mathbb{R}^d} |\nabla_x p_0(s,x,r,z) - \nabla_x p_0(s,x_0,r,z)| \, |q(r,z,t,y)| \, \mathrm{d}z$$

$$\lesssim \int_{\mathbb{R}^d} \lim_{x \to x_0} \sup_{(b,a) \in \mathscr{C}} |\nabla_x p_0(s,x,r,z) - \nabla_x p_0(s,x_0,r,z)| \phi_{0,0}^{(\alpha)}(r,z,t,y) \, \mathrm{d}z = 0.$$

In particular, Theorem 4.1 holds for $\alpha \in [1,2)$.

REMARK 4.2. — We remark that for $\alpha \in (0,1)$, under $\alpha + \beta > 1$, the second inequality in (4.4) may not hold since $\alpha + \gamma_0 - 1$ may be less than zero. This is also the reason that we have to make a different treatment for supercritical case. Let us mention that this proof anyhow works even in the super-critical case under the most stringent condition $\alpha + \beta/2 > 1$. Eventually, we also point out that the previous arguments can be simplified if $\alpha \in (1,2)$ for which the full parametrix expansion (3.12) of the density can actually be directly differentiated since the induced singularity in time remains integrable.

4.2. SUPERCRITICAL CASE $\alpha \in (0,1)$. — The following gradient estimate comes in [37].

THEOREM 4.3. — (Gradient estimate) Under $(\mathbf{H}_b^\beta)$, $(\mathbf{H}_a^\gamma)$ and (3.1), for any $T > 0$, there is a constant $C > 0$ such that for all $f \in C_b(\mathbb{R}^d)$ and $0 \leqslant s < t \leqslant T$,

$$|\nabla P_{s,t} f(x)| \lesssim_C (t-s)^{-1/\alpha} \|f\|_\infty,$$

where the constant $C$ may depend on $\|\nabla b\|_\infty$ and $\|\nabla a\|_\infty$.

Theorem 4.3 is important since it will precisely allow to justify that the Gronwall-Volterra lemma applies in the procedure below (see Step 5). It is used as a *prior* estimate. The analysis we now develop actually aims at proving that the constants in the gradient estimate are indeed *independent* of the smoothness of the coefficients and the boundedness of the drift. This is the main achievement of Theorem 4.1 which we here prove for the supercritical case $\alpha \in (0,1)$.

Below we fix $s < t$ and $x \in \mathbb{R}^d$ and assume $f \in C_b^\infty(\mathbb{R}^d)$. We divide the proof into six steps.

*Step 1.* — For notational simplicity, we shall write for $r \in [s, t]$,
$$\widetilde{\mathscr{A}_r} := \mathscr{A}_r^{(s,x)} = \mathscr{K}_r^{(s,x)} + \mathscr{B}_r^{(s,x)} =: \widetilde{\mathscr{K}_r} + \widetilde{\mathscr{B}_r},$$

with $\mathscr{K}_r^{(s,x)}$, $\mathscr{B}_r^{(s,x)}$ introduced in (3.6) and (3.7), and

$$(4.6) \qquad h(s, x, t, y) := \big(\nabla \widetilde{p}^{(\tau,\xi)}(s, \cdot, t, y)(x)\big)_{(\tau,\xi)=(s,x)} \overset{(3.4)}{=} -\nabla_y g_{s,t}^{(s,x)}(\theta_{s,t}(x) - y),$$

and for a function $f$,

$$H_{s,t} f(x) := \int_{\mathbb{R}^d} h(s, x, t, y) f(y) \, \mathrm{d}y.$$

By Lemma 3.1 we have

$$\nabla P_{s,t} f(x) = \nabla \widetilde{P}_{s,t}^{(\tau,\xi)} f(x) + \int_s^t \nabla \widetilde{P}_{s,r}^{(\tau,\xi)} \mathscr{A}_r^{(\tau,\xi)} P_{r,t} f(x) \, \mathrm{d}r.$$

Taking $(\tau, \xi) = (s, x)$ and using the above notations, we can write

$$(4.7) \qquad \nabla P_{s,t} f(x) = H_{s,t} f(x) + \int_s^t H_{s,r} \widetilde{\mathscr{A}_r} P_{r,t} f(x) \, \mathrm{d}r = H_{s,t} f(x) + \sum_{i=1}^4 I_{s,t}^{(i)} f(x),$$

where for $u := (s + t)/2$,

$$I_{s,t}^{(1)} f(x) := \int_s^u H_{s,r} \widetilde{\mathscr{K}_r} P_{r,t} f(x) \, \mathrm{d}r, \qquad I_{s,t}^{(2)} f(x) := \int_s^u H_{s,r} \widetilde{\mathscr{B}_r} P_{r,t} f(x) \, \mathrm{d}r,$$

$$I_{s,t}^{(3)} f(x) := \int_u^t H_{s,r} \widetilde{\mathscr{K}_r} P_{r,t} f(x) \, \mathrm{d}r, \qquad I_{s,t}^{(4)} f(x) := \int_u^t H_{s,r} \widetilde{\mathscr{B}_r} P_{r,t} f(x) \, \mathrm{d}r.$$

*Step 2.* — Note that for $j \in \mathbb{N}$,

$$(4.8) \qquad |\nabla_y^j h(s, x, t, y)| \overset{(4.6)}{=} |\nabla_y^{j+1} g_{s,t}^{(s,x)}(\theta_{s,t}(x) - y)| \overset{(2.28)}{\lesssim} \phi_{0,\alpha}^{(\alpha+j+1)}(s, x, t, y).$$

Thus we have

$$|H_{s,t} f(x)| \lesssim \int_{\mathbb{R}^d} \phi_{0,\alpha}^{(\alpha+1)}(s, x, t, y)|f(y)| \, \mathrm{d}y$$

$$\leqslant (t - s)^{-1/\alpha} \int_{\mathbb{R}^d} \phi_{0,\alpha}^{(\alpha)}(s, x, t, y)|f(y)| \, \mathrm{d}y.$$

For $I_{s,t}^{(1)} f(x)$, noting that by Lemma 3.7,

$$|\widetilde{\mathscr{K}_r} P_{r,t} f(z)| \lesssim (1 \wedge |z - \theta_{s,r}(x)|^\gamma)|\mathcal{D}^{(\alpha)} P_{r,t} f|(z)$$

$$\lesssim (1 \wedge |z - \theta_{s,r}(x)|^\gamma) \int_{\mathbb{R}^d} \phi_{0,0}^{(\alpha)}(r, z, t, y)|f(y)| \, \mathrm{d}y,$$

and using (2.31) and Lemma 2.7, we have

$$|I_{s,t}^{(1)} f(x)| \overset{(4.8)}{\lesssim} \int_s^u \int_{\mathbb{R}^d} \phi_{\gamma,\alpha}^{(\alpha+1)}(s, x, r, z) \int_{\mathbb{R}^d} \phi_{0,0}^{(\alpha)}(r, z, t, y)|f(y)| \, \mathrm{d}y \, \mathrm{d}z \, \mathrm{d}r$$

$$\overset{(2.31)}{\lesssim} \int_s^u \int_{\mathbb{R}^d} \big(\phi_{0,\alpha+\gamma-1}^{(\alpha)} \odot \phi_{0,0}^{(\alpha)}\big)_r(s, x, t, y)|f(y)| \, \mathrm{d}y \, \mathrm{d}r$$

$$\overset{(2.20)}{\lesssim} \int_{\mathbb{R}^d} \phi_{0,\alpha+\gamma-1}^{(\alpha)}(s, x, t, y)|f(y)| \, \mathrm{d}y.$$

For $I_{s,t}^{(2)}f(x)$, noting that

$$|\widetilde{\mathscr{B}}_r P_{r,t}f(z)| \lesssim (|\theta_{s,r}(x) - z|^\beta + |\theta_{s,r}(x) - z| + (r-s)^{\beta/\alpha})|\nabla P_{r,t}f(z)|,$$

using (4.8) and (2.31), we have

$$|I_{s,t}^{(2)}f(x)| \lesssim \int_s^u \int_{\mathbb{R}^d} \phi_{0,\alpha+\beta-1}^{(\alpha)}(s,x,r,z)|\nabla P_{r,t}f(z)|\,\mathrm{d}z\,\mathrm{d}r$$

$$\lesssim (t-s)^{-1/\alpha} \int_s^t \int_{\mathbb{R}^d} \phi_{0,\alpha+\beta-1}^{(\alpha)}(s,x,r,z)(t-r)^{1/\alpha}|\nabla P_{r,t}f(z)|\,\mathrm{d}z\,\mathrm{d}r.$$

*Step 3.* — In this step we treat the hard term $I_{s,t}^{(3)}f(x)$. Let $\varepsilon := (t-r)^{1/\alpha}$ and

$$\kappa_\varepsilon(r,z,z') := \kappa(r,\cdot,z') * \rho_\varepsilon(z), \quad \overline{\kappa}_\varepsilon(r,z,z') := \kappa_\varepsilon(r,z,z') - \kappa_\varepsilon(r,\theta_{r,s}(x),z')$$

and

$$\widetilde{\mathscr{K}}_r^{(\varepsilon)}f(z) = 2\int_{\mathbb{R}^d} \delta_f^{(1)}(z;z') \frac{\overline{\kappa}_\varepsilon(r,z,z')}{|z'|^{d+\alpha}}\,\mathrm{d}z'.$$

Let us write

$$I_{s,t}^{(3)}f(x) = \int_u^t \left(H_{s,r}(\widetilde{\mathscr{K}}_r - \widetilde{\mathscr{K}}_r^{(\varepsilon)})P_{r,t}f(x) + H_{s,r}\widetilde{\mathscr{K}}_r^{(\varepsilon)}P_{r,t}f(x)\right)\mathrm{d}r$$

$$=: \int_u^t \left(J_{1,r}^{(\varepsilon)}(s,x,t) + J_{2,r}^{(\varepsilon)}(s,x,t)\right)\mathrm{d}r.$$

Let $\gamma_1 \in (0,\gamma)$. Noting that

$$|(\kappa - \kappa_\varepsilon)(r,z,z') - (\kappa - \kappa_\varepsilon)(r,\theta_{r,s}(x),z')| \lesssim_C (|z - \theta_{r,s}(x)|^{\gamma_1} \wedge 1)\varepsilon^{\gamma-\gamma_1},$$

by definition and Lemma 3.7, we have

$$|(\widetilde{\mathscr{K}}_r - \widetilde{\mathscr{K}}_r^{(\varepsilon)})P_{r,t}f(z)| \lesssim (|z - \theta_{r,s}(x)|^{\gamma_1} \wedge 1)\varepsilon^{\gamma-\gamma_1}|\mathcal{D}^{(\alpha)}P_{r,t}f|(z)$$

$$\lesssim (|z - \theta_{r,s}(x)|^{\gamma_1} \wedge 1)\varepsilon^{\gamma-\gamma_1}\int_{\mathbb{R}^d}\phi_{0,0}^{(\alpha)}(r,z,t,y)|f(y)|\,\mathrm{d}y.$$

For $J_{1,r}^{(\varepsilon)}$, recalling $\varepsilon = (t-r)^{1/\alpha}$, we have

$$\int_u^t |J_{1,r}^{(\varepsilon)}(s,x,t)|\,\mathrm{d}r \overset{(4.8)}{\lesssim} \int_u^t \int_{\mathbb{R}^d} \left(\phi_{0,\alpha+\gamma_1-1}^{(\alpha)} \odot \phi_{0,\gamma-\gamma_1}^{(\alpha)}\right)_r(s,x,t,y)|f(y)|\,\mathrm{d}y\,\mathrm{d}r$$

$$\overset{(2.20)}{\lesssim} \int_{\mathbb{R}^d} \phi_{0,\alpha+\gamma-1}^{(\alpha)}(s,x,t,y)|f(y)|\,\mathrm{d}y.$$

For $J_{2,r}^{(\varepsilon)}$, by the change of variables and Fubini's theorem, we have

$$J_{2,r}^{(\varepsilon)}(s,x,t) = \int_{\mathbb{R}^d} h(s,x,r,z)\int_{\mathbb{R}^d} \delta_{P_{r,t}f}^{(1)}(z;z') \frac{\overline{\kappa}_\varepsilon(r,z,z')}{|z'|^{d+\alpha}}\,\mathrm{d}z'\,\mathrm{d}z$$

$$= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \delta_{h(s,x,r,\cdot)\overline{\kappa}_\varepsilon(r,\cdot,z')}^{(1)}(z;z') \frac{\mathrm{d}z'}{|z'|^{d+\alpha}}\,P_{r,t}f(z)\,\mathrm{d}z$$

$$= \int_{\mathbb{R}^d} h(s,x,r,z)\int_{\mathbb{R}^d} \delta_{\overline{\kappa}_\varepsilon(r,\cdot,z')}^{(1)}(z;z') \frac{\mathrm{d}z'}{|z'|^{d+\alpha}}\,P_{r,t}f(z)\,\mathrm{d}z$$

$$+ \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \delta_{h(s,x,r,\cdot)}^{(1)}(z;z')\overline{\kappa}_\varepsilon(r,z+z',z') \frac{\mathrm{d}z'}{|z'|^{d+\alpha}}\,P_{r,t}f(z)\,\mathrm{d}z.$$

Noting that by $(\mathbf{H}_a^\gamma)$,

$$\big|\delta^{(1)}_{\overline{\kappa}_\varepsilon(r,\cdot,z')}(z;z')\big| \lesssim (\varepsilon^{\gamma-1}|z'|) \wedge |z'|^\gamma \wedge 1,$$

we have

$$\int_{\mathbb{R}^d}\big|\delta^{(1)}_{\overline{\kappa}_\varepsilon(r,\cdot,z')}(z;z')\big|\frac{\mathrm{d}z'}{|z'|^{d+\alpha}} \lesssim \int_{\mathbb{R}^d}((\varepsilon^{\gamma-1}|z'|)\wedge|z'|^\gamma \wedge 1)\frac{\mathrm{d}z'}{|z'|^{d+\alpha}} \lesssim \varepsilon^{(\gamma-\alpha)\wedge 0}.$$

On the other hand, by (4.6) and (2.28),

$$\big|\delta^{(1)}_{h(s,x,r,\cdot)}(z;z')\big| \lesssim \big(((r-s)^{-1/\alpha}|z'|)\wedge 1\big)\big(\phi^{(\alpha)}_{0,\alpha-1}(s,x,r,z+z')+\phi^{(\alpha)}_{0,\alpha-1}(s,x,r,z)\big).$$

Thus, as in (2.36) we have

$$\int_{\mathbb{R}^d}\frac{\big|\delta^{(1)}_{h(s,x,r,\cdot)}(z;z')\big|\,\mathrm{d}z'}{|z'|^{d+\alpha}} \lesssim \int_{\mathbb{R}^d}\big(((r-s)^{-1/\alpha}|z'|)\wedge 1\big)\phi^{(\alpha)}_{0,\alpha-1}(s,x,r,z+z')\frac{\mathrm{d}z'}{|z'|^{d+\alpha}}$$

$$+ \phi^{(\alpha)}_{0,\alpha-1}(s,x,r,z)\int_{\mathbb{R}^d}\big(((r-s)^{-1/\alpha}|z'|)\wedge 1\big)\frac{\mathrm{d}z'}{|z'|^{d+\alpha}}$$

$$\lesssim \phi^{(\alpha)}_{0,\alpha-1}(s,x,r,z)(r-s)^{-1} = \phi^{(\alpha)}_{0,-1}(s,x,r,z).$$

Therefore,

$$|J^{(\varepsilon)}_{2,r}(s,x,t)| \lesssim \int_{\mathbb{R}^d}\Big[\varepsilon^{(\gamma-\alpha)\wedge 0}\phi^{(\alpha)}_{0,\alpha-1}(s,x,r,z)+\phi^{(\alpha)}_{0,-1}(s,x,r,z)\Big]P_{r,t}|f|(z)\,\mathrm{d}z.$$

Recall $\varepsilon = (t-r)^{1/\alpha}$. By (2.31), we obtain

$$\int_u^t|J^{(\varepsilon)}_{2,r}(s,x,t)|\,\mathrm{d}r \lesssim (t-s)^{-1/\alpha}\int_{\mathbb{R}^d}\phi^{(\alpha)}_{0,\alpha}(s,x,t,y)|f(y)|\,\mathrm{d}y.$$

*Step 4.* — For $\varepsilon = (t-r)^{1/\alpha}$, we define

$$\overline{b}_\varepsilon(r,z) := (b*\rho_\varepsilon)(r,z) - (b*\rho_\varepsilon*\rho_{|r-s|^{1/\alpha}})(r,\theta_{r,s}(x))$$

and

$$\widetilde{\mathscr{B}}^{(\varepsilon)}_r f(z) := \overline{b}_\varepsilon(r,z)\cdot\nabla f(z).$$

For $I^{(4)}_{s,t}$, we similarly write

$$I^{(4)}_{s,t}f(x) = \int_u^t\Big(H_{s,r}(\widetilde{\mathscr{B}}_r - \widetilde{\mathscr{B}}^{(\varepsilon)}_r)P_{r,t}f(x) + H_{s,r}\widetilde{\mathscr{B}}^{(\varepsilon)}_r P_{r,t}f(x)\Big)\,\mathrm{d}r$$

$$=: \int_u^t\Big(J^{(\varepsilon)}_{3,r}(s,x,t) + J^{(\varepsilon)}_{4,r}(s,x,t)\Big)\,\mathrm{d}r.$$

For $J^{(\varepsilon)}_{3,r}$, since

$$|\overline{b}_0 - \overline{b}_\varepsilon|(r,z) \leqslant \kappa_0\varepsilon^\beta = \kappa_0(t-s)^{\beta/\alpha},$$

by (4.8) we have

$$|J^{(\varepsilon)}_{3,r}(s,x,t)| = \left|\int_{\mathbb{R}^d}h(s,x,r,z)(\overline{b}_0(r,z)-\overline{b}_\varepsilon(r,z))\cdot\nabla P_{r,t}f(z)\,\mathrm{d}z\right|$$

$$\lesssim \int_{\mathbb{R}^d}\phi^{(\alpha+1)}_{0,\alpha}(s,x,r,z)(t-r)^{\beta/\alpha}|\nabla P_{r,t}f(z)|\,\mathrm{d}z,$$

and

$$\int_u^t|J^{(\varepsilon)}_{3,r}(s,x,t)|\,\mathrm{d}r \lesssim (t-s)^{-1/\alpha}\int_s^t\int_{\mathbb{R}^d}\phi^{(\alpha)}_{0,\alpha}(s,x,r,z)(t-r)^{\beta/\alpha}|\nabla P_{r,t}f(z)|\,\mathrm{d}z\,\mathrm{d}r.$$

For $J_{4,r}^{(\varepsilon)}$, we derive integrating by parts that

$$|J_{4,r}^{(\varepsilon)}(s,x,t)| = \left| \int_{\mathbb{R}^d} h(s,x,r,z)\, \overline{b}_\varepsilon(r,z) \cdot \nabla_z P_{r,t} f(z) \,\mathrm{d}z \right|$$

$$\leqslant \left| \int_{\mathbb{R}^d} h(s,x,r,z)\, \mathrm{div}\, \overline{b}_\varepsilon(r,z) P_{r,t} f(z) \,\mathrm{d}z \right|$$

$$+ \left| \int_{\mathbb{R}^d} \overline{b}_\varepsilon(r,z) \cdot \nabla_z h(s,x,r,z) P_{r,t} f(z) \,\mathrm{d}z \right|.$$

Since

$$|\,\mathrm{div}\, \overline{b}_\varepsilon(r,z)| = |\,\mathrm{div}\, b_\varepsilon(r,z)| \leqslant \kappa_0 \varepsilon^{\beta-1} = \kappa_0 (t-r)^{(\beta-1)/\alpha}$$

and

$$|\overline{b}_\varepsilon|(r,z) \lesssim |z - \theta_{r,s}(x)|^\beta + (r-s)^{\beta/\alpha},$$

by (4.8) and (2.31) we have

$$|J_{4,r}^{(\varepsilon)}(s,x,t)| \lesssim \int_{\mathbb{R}^d} \phi_{0,\alpha-1}^{(\alpha)}(s,x,r,z)(t-r)^{(\beta-1)/\alpha} |P_{r,t} f(z)| \,\mathrm{d}z$$

$$+ \int_{\mathbb{R}^d} \phi_{0,\alpha+\beta-2}^{(\alpha)}(s,x,r,z) |P_{r,t} f(z)| \,\mathrm{d}z.$$

Thus,

$$\int_u^t |J_{4,r}^{(\varepsilon)}(s,x,t)| \,\mathrm{d}r \lesssim \int_u^t \int_{\mathbb{R}^d} \left( \phi_{0,\alpha-1}^{(\alpha)} \odot \phi_{0,\alpha+\beta-1}^{(\alpha)} \right)_r (s,x,t,y) |f(y)| \,\mathrm{d}y \,\mathrm{d}r$$

$$+ \int_u^t \int_{\mathbb{R}^d} \left( \phi_{0,\alpha+\beta-2}^{(\alpha)} \odot \phi_{0,\alpha}^{(\alpha)} \right)_r (s,x,t,y) |f(y)| \,\mathrm{d}y \,\mathrm{d}r$$

$$\overset{(2.20)}{\lesssim} \int_u^t \left[ (r-s)^{(\alpha-1)/\alpha}(t-r)^{(\alpha+\beta-1)/\alpha} + (r-s)^{(\alpha+\beta-2)/\alpha}(t-r) \right]$$

$$\times \left[ (r-s)^{-1} + (t-r)^{-1} \right] \,\mathrm{d}r \int_{\mathbb{R}^d} \phi_{0,0}^{(\alpha)}(s,x,t,y) |f(y)| \,\mathrm{d}y$$

$$\lesssim \int_{\mathbb{R}^d} \phi_{0,2\alpha+\beta-2}^{(\alpha)}(s,x,t,y) |f(y)| \,\mathrm{d}y$$

$$\lesssim (t-s)^{-1/\alpha} \int_{\mathbb{R}^d} \phi_{0,\alpha}^{(\alpha)}(s,x,t,y) |f(y)| \,\mathrm{d}y,$$

recalling that $\alpha + \beta > 1$ for the last inequality. Hence,

$$|I_{s,t}^{(4)} f(x)| \lesssim (t-s)^{-1/\alpha} \left( \int_s^t \int_{\mathbb{R}^d} \phi_{0,\alpha}^{(\alpha)}(s,x,r,z)(t-r)^{\beta/\alpha} |\nabla P_{r,t} f(z)| \,\mathrm{d}z \,\mathrm{d}r \right.$$

$$\left. + \int_{\mathbb{R}^d} \phi_{0,\alpha}^{(\alpha)}(s,x,t,y) |f(y)| \,\mathrm{d}y \right).$$

*Step 5.* — Combining the above calculations, we obtain

$$|\nabla P_{s,t} f(x)| \lesssim (t-s)^{-1/\alpha} \int_{\mathbb{R}^d} \phi_{0,\alpha}^{(\alpha)}(s,x,t,y) |f(y)| \,\mathrm{d}y$$

$$+ (t-s)^{-1/\alpha} \int_s^t \int_{\mathbb{R}^d} \phi_{0,\alpha}^{(\alpha)}(s,x,r,z)(t-r)^{\beta/\alpha} |\nabla P_{r,t} f(z)| \,\mathrm{d}z \,\mathrm{d}r$$

$$+ (t-s)^{-1/\alpha} \int_s^t \int_{\mathbb{R}^d} \phi_{0,\alpha+\beta-1}^{(\alpha)}(s,x,r,z)(t-r)^{1/\alpha} |\nabla P_{r,t} f(z)| \,\mathrm{d}z \,\mathrm{d}r.$$

By the lower bound estimate, we further have

$$(4.9) \quad (t-s)^{1/\alpha} |\nabla P_{s,t} f(x)| \lesssim P_{s,t}|f|(x) + \int_s^t (t-r)^{\beta/\alpha} P_{s,r} |\nabla P_{r,t} f|(x)\,\mathrm{d}r$$

$$+ \int_s^t (r-s)^{(\beta-1)/\alpha}(t-r)^{1/\alpha} P_{s,r}|\nabla P_{r,t}f|(x)\,\mathrm{d}r.$$

For fixed $0 \leqslant u < t \leqslant T$ and $s \in (u,t)$, we let

$$\Gamma_u^t(s,x) := (t-s)^{1/\alpha} P_{u,s}|\nabla P_{s,t}f|(x).$$

Using $P_{u,s}$ act on both sides of (4.9) and by $P_{u,s}P_{s,r} = P_{u,r}$, we derive that

$$\Gamma_u^t(s,x) \lesssim P_{u,t}|f|(x) + \int_s^t \left[(r-s)^{(\beta-1)/\alpha} + (t-r)^{(\beta-1)/\alpha}\right] \Gamma_u^t(r,x)\,\mathrm{d}r.$$

Note that by Theorem 4.3,

$$\sup_{s\in[u,t]} \|\Gamma_u^t(s,\cdot)\|_\infty < \infty.$$

Since $\alpha + \beta > 1$, from the Volterra-Gronwall inequality, we obtain that for all $s \in (u,t)$,

$$\Gamma_u^t(s,x) \lesssim P_{u,t}|f|(x).$$

Taking limit $u \uparrow s$, we obtain

$$(t-s)^{1/\alpha} |\nabla P_{s,t}f|(x) \lesssim P_{s,t}|f|(x),$$

which eventually yields the desired gradient estimate.

*Step 6.* — Finally, by (4.7) and the dominated convergence theorem, one can show that

$$\lim_{x\to x_0} \sup_{(b,a)\in\mathscr{C}} \left|\nabla P_{s,t}^{b,a}f(x) - \nabla P_{s,t}^{b,a}f(x_0)\right| = 0.$$

Indeed, from the above proof, it suffices to show that

$$\lim_{x\to x_0} \sup_{(b,a)\in\mathscr{C}} \left|H_{s,t}^{b,a}f(x) - H_{s,t}^{b,a}f(x_0)\right| = 0.$$

This follows by Lemma 2.8.

## 5. Proof of Theorem 1.1

The point is here to prove Theorem 1.1. Namely, we want to extend the bounds of Theorem 3.5, Lemma 3.7 and Theorem 4.1 under the sole assumptions $(\mathbf{H}_a^\gamma)$, $(\mathbf{H}_b^\beta)$. In Section 5.1 we consider bounded drift $b$ (without additional smoothness than in $(\mathbf{H}_a^\gamma)$, $(\mathbf{H}_b^\beta)$). In Section 5.2 we drop the bounded assumption on $b$ by a truncation argument.

5.1. Bounded drift $b$. — In this subsection we assume $(\mathbf{H}_a^\gamma)$, $(\mathbf{H}_b^\beta)$ and $b$ is bounded. Let $a_\varepsilon$ and $b_\varepsilon$ be the smooth approximations of $a$ and $b$, respectively. Hence, assumptions $(\mathbf{H}_a^\gamma)$, $(\mathbf{H}_b^\beta)$ and (3.1) are met by $a_\varepsilon, b_\varepsilon$ for the SDE

$$(5.1) \qquad \mathrm{d}X_t^\varepsilon = b_\varepsilon(t, X_t^\varepsilon)\,\mathrm{d}t + a_\varepsilon(t, X_{t-}^\varepsilon)\,\mathrm{d}L_t^{(\alpha)}.$$

The following convergence in law result was established in [7], see Theorem 1.1 therein.

Theorem 5.1. — *Let $X_{s,t}^\varepsilon(x)$ be the unique solution of SDE* (5.1) *starting from $x$ at time $s$. Then $X_{s,t}^\varepsilon(x)$ weakly converges to $X_{s,t}(x)$.*

*Proof.* — For fixed $(s, x) \in \mathbb{R}_+ \times \mathbb{R}^d$, since the coefficients $b, a$ have linear growth, it is by now standard to show that the law of $X_{s,\cdot}^\varepsilon(x)$ is tight in the space of all càdlàg functions (see e.g. the Aldous criterion in the monographs [18, Ch. VI, Th. 4.5], [1]). Through the martingale problem formulation, one can show as well that any weak accumulation point of the law of $X_{s,\cdot}^\varepsilon(x)$ is a weak solution of SDE (1.1). Finally, by the weak uniqueness, one sees that $X_{s,t}^\varepsilon(x)$ weakly converges to $X_{s,t}(x)$. $\quad\square$

Denoting by $p_\varepsilon$ the associated density, it therefore holds from Theorem 3.5, Lemma 3.7 and Theorem 4.1 that

(i) (Two-sides estimate) For any $T > 0$, there is a constant $C_1 = C_1(T, \Theta) > 0$ such that for all $0 \leqslant s < t \leqslant T$ and $x, y \in \mathbb{R}^d$,

$$(5.2) \qquad p_\varepsilon(s, x, t, y) \asymp_{C_1} \phi_{0,\alpha}^{(\alpha)}(s, x, t, y).$$

(ii) (Fractional derivative estimate) For any $T > 0$, there is a constant $C_2 = C_2(T, \Theta) > 0$ such that for all $0 \leqslant s < t \leqslant T$ and $x, y \in \mathbb{R}^d$,

$$(5.3) \qquad |\mathcal{D}^{(\alpha)}p_\varepsilon(s, \cdot, t, y)|(x) \lesssim_{C_2} \phi_{0,0}^{(\alpha)}(s, x, t, y).$$

(iii) (Gradient estimate in $x$) For any $T > 0$, there is a constant $C_3 = C_3(T, \Theta) > 0$ such that for all $0 \leqslant s < t \leqslant T$ and $x, y \in \mathbb{R}^d$,

$$(5.4) \qquad |\nabla P_{s,t}^\varepsilon f(x)| \lesssim_{C_3} (t - s)^{-1/\alpha} P_{s,t}^\varepsilon |f|(x),$$

where the constants in the above controls only depend on $(\mathbf{H}_a^\gamma)$, $(\mathbf{H}_b^\beta)$ through $\Theta$ (see precisely (1.22)).

By Theorem 5.1, we have for any $f \in C_b(\mathbb{R}^d)$,

$$(5.5) \qquad \lim_{\varepsilon \to 0} P_{s,t}^\varepsilon f(x) := \lim_{\varepsilon \to 0} \mathbb{E}f(X_{s,t}^\varepsilon(x)) = \mathbb{E}f(X_{s,t}(x)) =: P_{s,t}f(x).$$

(i) (Two-sided estimates) For nonnegative measurable function $f$, we get from (5.2)

$$C_1^{-1} \int_{\mathbb{R}^d} \phi_{0,\alpha}^{(\alpha)}(s, x, t, y)f(y)\,\mathrm{d}y \leqslant \mathbb{E}f(X_{s,t}(x)) \leqslant C_1 \int_{\mathbb{R}^d} \phi_{0,\alpha}^{(\alpha)}(s, x, t, y)f(y)\,\mathrm{d}y,$$

which implies that $X_{s,t}(x)$ has a density $p(s, x, t, y)$ having lower and upper bound as in (1.24). On the other hand, for fixed $s < t$, by Theorem 3.5 we have

$$(x, y) \longmapsto p_\varepsilon(s, x, t, y) \text{ is equi-continuous in } \varepsilon \in (0, 1).$$

From the Ascoli-Arzelà theorem, there are a subsequence $\varepsilon_k$ and a continuous function $\overline{p}(s,x,t,y)$ as a function of $x, y \in \mathbb{R}^d$ such that

$$(5.6) \qquad p_{\varepsilon_k}(s,x,t,y) \longrightarrow \overline{p}(s,x,t,y) \text{ locally uniformly in } x, y \in \mathbb{R}^d,$$

which together with (5.5) yields that

$$(5.7) \qquad p(s,x,t,y) = \overline{p}(s,x,t,y) \text{ is continuous as a function of } x, y \in \mathbb{R}^d.$$

(ii) (Fractional derivative estimates) It follows by (5.3), (5.6), (5.7) and Fatou's lemma that

$$
\begin{aligned}
|\mathcal{D}^{(\alpha)}p(s,\cdot,t,y)|(x) &= \int_{\mathbb{R}^d} \lim_{k\to\infty} \left|\delta^{(2)}_{p_{\varepsilon_k}(s,\cdot,t,y)}(x;z)\right| \frac{\mathrm{d}z}{|z|^{d+\alpha}} \\
&\leqslant \varliminf_{k\to\infty} \int_{\mathbb{R}^d} \left|\delta^{(2)}_{p_{\varepsilon_k}(s,\cdot,t,y)}(x;z)\right| \frac{\mathrm{d}z}{|z|^{d+\alpha}} \\
&= \varliminf_{k\to\infty} |\mathcal{D}^{(\alpha)}p_{\varepsilon_k}(s,\cdot,t,y)|(x) \lesssim_{C_2} \phi^{(\alpha)}_{0,0}(s,x,t,y).
\end{aligned}
$$

(iii) (Gradient estimates) For fixed $f \in C_b(\mathbb{R}^d)$, by (5.4),

$$x \longmapsto \nabla P^{\varepsilon}_{s,t}f(x) \text{ is equi-continuous in } \varepsilon,$$

which together with (5.5) implies that $x \mapsto P_{s,t}f(x)$ is continuous differentiable. By taking limits along a subsequence $\varepsilon_k$ for (5.4), we obtain

$$|\nabla P_{s,t}f(x)| \lesssim_{C_3} (t-s)^{-1/\alpha} P_{s,t}|f|(x).$$

Finally, for fixed $t' > t$ and $y \in \mathbb{R}^d$, we let $f(x) := p(t,x,t',y)$, then by the Chapman-Kolmogorov equation, we obtain

$$|\nabla p(s,\cdot,t',y)(x)| \lesssim_{C_3} (t-s)^{-1/\alpha} p(s,x,t',y).$$

This then readily gives estimate (4.2) (logarithmic derivative) of Theorem 1.1.

5.2. Unbounded drift $b$. — In this subsection we assume $(\mathbf{H}^\gamma_a)$ and $(\mathbf{H}^\beta_b)$. For $n \in \mathbb{N}$, define

$$b^n(t,x) := (-n) \vee b(t,x) \wedge n.$$

With the $\kappa_0$ as in $(\mathbf{H}^\beta_b)$, we have

$$(5.8) \qquad |b^n(t,0)| \leqslant \kappa_0, \ |b^n(t,x) - b^n(t,y)| \leqslant \kappa_0(|x-y|^\beta \vee |x-y|).$$

Consider the following SDE

$$(5.9) \qquad \mathrm{d}X^n_t = b^n(t,X^n_t)\,\mathrm{d}t + a(t,X^n_{t-})\,\mathrm{d}L^{(\alpha)}_t.$$

The following result is the very similar to Theorem 5.1.

THEOREM 5.2. — *Let $X^n_{s,t}(x)$ be the unique solution of SDE* (5.9) *starting from $x$ at time $s$. Then $X^n_{s,t}(x)$ weakly converges to $X_{s,t}(x)$.*

Moreover, for fixed $s > 0$, consider the following ODE:

$$(5.10) \qquad \dot{\theta}^n_{s,t} = b^n_{|t-s|^{1/\alpha}}(t,\theta^n_{s,t}), \quad \theta^n_{s,s} = x, \quad t \geqslant 0.$$

We have the following convergence.

Lemma 5.3. — *For each $s, t > 0$ and $x \in \mathbb{R}^d$, it holds that*

$$\lim_{n \to \infty} |\theta^n_{s,t}(x) - \theta_{s,t}(x)| = 0.$$

*Proof.* — We drop the starting point $x$ and assume $t > s$. From dynamics (1.21) and (5.10),

$$|\theta^n_{s,t} - \theta_{s,t}| \leqslant \int_s^t \|\nabla b^n_{|r-s|^{1/\alpha}}(r, \cdot)\|_\infty |\theta^n_{s,r} - \theta_{s,r}| \, dr$$
$$+ \int_s^t \left| b^n_{|r-s|^{1/\alpha}}(r, \theta_{s,r}) - b_{|r-s|^{1/\alpha}}(r, \theta_{s,r}) \right| dr.$$

By (5.8), (1.20) and Gronwall's inequality, we have

$$|\theta^n_{s,t} - \theta_{s,t}| \leqslant C \int_s^t \left| b^n_{|r-s|^{1/\alpha}}(r, \theta_{s,r}) - b_{|r-s|^{1/\alpha}}(r, \theta_{s,r}) \right| dr,$$

which gives the desired limit by the dominated convergence theorem. □

Now by Theorem 5.2 and Lemma 5.3, using exactly the same argument as in Section 5.1, we can show Theorem 1.1.

## References

[1] R. F. Bass – *Stochastic processes*, Cambridge Series in Statistical and Probabilistic Math., vol. 33, Cambridge University Press, Cambridge, 2011.

[2] R. F. Bass & Z.-Q. Chen – "Systems of equations driven by stable processes", *Probab. Theory Relat. Fields* **134** (2006), no. 2, p. 175–214.

[3] A. Bendikov – "Asymptotic formulas for symmetric stable semigroups", *Exposition. Math.* **12** (1994), no. 4, p. 381–384.

[4] K. Bogdan & T. Jakubowski – "Estimates of heat kernel of fractional Laplacian perturbed by gradient operators", *Comm. Math. Phys.* **271** (2007), no. 1, p. 179–198.

[5] B. Böttcher, R. Schilling & J. Wang – *Lévy matters. III. Lévy-type processes: construction, approximation and sample path properties*, Lect. Notes in Math., vol. 2099, Springer, Cham, 2013.

[6] P.-É. Chaudru de Raynal, S. Menozzi & E. Priola – "Schauder estimates for drifted fractional operators in the supercritical case", *J. Funct. Anal.* **278** (2020), no. 8, article no. 108425 (57 pages).

[7] Z.-Q. Chen, Z. Hao & X. Zhang – "Hölder regularity and gradient estimates for SDEs driven by cylindrical $\alpha$-stable processes", *Electron. J. Probab.* **25** (2020), p. Paper No. 137, 23.

[8] Z.-Q. Chen, E. Hu, L. Xie & X. Zhang – "Heat kernels for non-symmetric diffusion operators with jumps", *J. Differential Equations* **263** (2017), no. 10, p. 6576–6634.

[9] Z.-Q. Chen & X. Zhang – "Heat kernels and analyticity of non-symmetric jump diffusion semigroups", *Probab. Theory Relat. Fields* **165** (2016), no. 1-2, p. 267–312.

[10] _____, "Heat kernels for time-dependent non-symmetric mixed Lévy-type operators", 2016, arXiv:1604.02681.

[11] _____, "Heat kernels for time-dependent non-symmetric stable-like operators", *J. Math. Anal. Appl.* **465** (2018), no. 1, p. 1–21.

[12] Z.-Q. Chen, X. Zhang & G. Zhao – "Supercritical SDEs driven by multiplicative stable-like Lévy processes", *Trans. Amer. Math. Soc.* **374** (2021), no. 11, p. 7621–7655.

[13] A. Debussche & N. Fournier – "Existence of densities for stable-like driven SDE's with Hölder continuous coefficients", *J. Funct. Anal.* **264** (2013), no. 8, p. 1757–1778.

[14] F. Delarue & S. Menozzi – "Density estimates for a random noise propagating through a chain of differential equations", *J. Funct. Anal.* **259** (2010), no. 6, p. 1577–1630.

[15] K. Du & X. Zhang – "Optimal gradient estimates of heat kernels of stable-like operators", *Proc. Amer. Math. Soc.* **147** (2019), no. 8, p. 3559–3565.

[16] A. Friedman – *Partial differential equations of parabolic type*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1964.

[17] L. Huang – "Density estimates for sdes driven by tempered stable processes", 2015, `arXiv: 1504.04183`.

[18] J. Jacod & A. N. Shiryaev – *Limit theorems for stochastic processes*, second ed., Grundlehren Math. Wiss., vol. 288, Springer-Verlag, Berlin, 2003.

[19] V. Knopova & A. M. Kulik – "Parametrix construction of the transition probability density of the solution to an SDE driven by $\alpha$-stable noise", *Ann. Inst. H. Poincaré Probab. Statist.* **54** (2018), no. 1, p. 100–140.

[20] V. Knopova, A. M. Kulik & R. Schilling – "Construction and heat kernel estimates of general stable-like Markov processes", 2020, `arXiv:2005.08491`.

[21] A. N. Kochubej – "Parabolic pseudodifferential equations, hypersingular integrals, and Markov processes", *Math. USSR-Izv.* **33** (1989), no. 2, p. 233–259.

[22] V. Kolokoltsov – "Symmetric stable laws and stable-like jump-diffusions", *Proc. London Math. Soc. (3)* **80** (2000), no. 3, p. 725–768.

[23] V. Konakov, S. Menozzi & S. Molchanov – "Explicit parametrix and local limit theorems for some degenerate diffusion processes", *Ann. Inst. H. Poincaré Probab. Statist.* **46** (2010), no. 4, p. 908—923.

[24] F. Kühn – "Interior Schauder estimates for elliptic equations associated with Lévy operators", *Potential Anal.* **56** (2022), no. 3, p. 459–481.

[25] T. Kulczycki, A. M. Kulik & M. Ryznar – "On weak solution of SDE driven by inhomogeneous singular Lévy noise", 2021, `arXiv:2104.08129`.

[26] T. Kulczycki & M. Ryznar – "Semigroup properties of solutions of SDEs driven by Lévy processes with independent coordinates", *Stochastic Process. Appl.* **130** (2020), no. 12, p. 7185–7217.

[27] T. Kulczycki, M. Ryznar & P. Sztonyk – "Strong Feller property for SDEs driven by multiplicative cylindrical stable noise", *Potential Anal.* **55** (2021), no. 1, p. 75–126.

[28] A. M. Kulik – "On weak uniqueness and distributional properties of a solution to an SDE with $\alpha$-stable noise", *Stochastic Process. Appl.* **129** (2019), no. 2, p. 473–506.

[29] A. M. Kulik, S. Peszat & E. Priola – "Gradient formula for transition semigroup corresponding to stochastic equation driven by a system of independent Lévy processes", 2020, `arXiv:2006.09133`.

[30] E. E. Levi – "Sulle equazioni lineari totalmente ellittiche alle derivate parziali", *Rend. Circ. Mat. Palermo* **24** (1907), no. 1, p. 275–317.

[31] W. Liu, R. Song & L. Xie – "Gradient estimates for the fundamental solution of Lévy type operator", *Adv. in Nonlinear Anal.* **9** (2020), no. 1, p. 1453–1462.

[32] A. J. Majda & A. L. Bertozzi – *Vorticity and incompressible flow*, Cambridge Texts in Applied Math., vol. 27, Cambridge University Press, Cambridge, 2002.

[33] H. P. McKean, Jr. & I. M. Singer – "Curvature and the eigenvalues of the Laplacian", *J. Differential Geom.* **1** (1967), no. 1, p. 43–69.

[34] S. Menozzi, A. Pesce & X. Zhang – "Density and gradient estimates for non degenerate Brownian SDEs with unbounded measurable drift", *J. Differential Equations* **272** (2021), p. 330–369.

[35] R. Mikulevicius & H. Pragarauskas – "On the Cauchy problem for integro-differential operators in Hölder classes and the uniqueness of the martingale problem", *Potential Anal.* **40** (2014), no. 4, p. 539–563.

[36] H. Tanaka, M. Tsuchiya & S. Watanabe – "Perturbation of drift-type for Lévy processes", *J. Math. Kyoto Univ.* **14** (1974), p. 73–92.

[37] F.-Y. Wang, L. Xu & X. Zhang – "Gradient estimates for SDEs driven by multiplicative Lévy noise", *J. Funct. Anal.* **269** (2015), no. 10, p. 3195–3219.

[38] T. Watanabe – "Asymptotic estimates of multi-dimensional stable densities and their applications", *Trans. Amer. Math. Soc.* **359** (2007), no. 6, p. 2851–2879.

[39] X. Zhang & G. Zhao – "Dirichlet problem for supercritical non-local operators", 2018, `arXiv: 1809.05712`.

Stéphane Menozzi, Laboratoire de Modélisation Mathématique d'Evry (LaMME), UMR CNRS 8071,
Université d'Evry Val d'Essonne (Université Paris Saclay)
23 Boulevard de France 91037 Evry, France
& Laboratory of Stochastic Analysis, HSE
Pokrovsky Blvd, 11, Moscow, Russian Federation
*E-mail :* `stephane.menozzi@univ-evry.fr`
*Url :* `http://www.math-evry.cnrs.fr/members/smenozzi/`

Xicheng Zhang, School of Mathematics and Statistics, Wuhan University
Wuhan, Hubei 430072, P.R. China
*E-mail :* `XichengZhang@gmail.com`