ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

ARCHIVIO ISTITUZIONALE
DELLA RICERCA

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Density and gradient estimates for non degenerate Brownian SDEs with unbounded measurable drift

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:
Menozzi, S., Pesce, A., Zhang, X. (2021). Density and gradient estimates for non degenerate Brownian SDEs with unbounded measurable drift. JOURNAL OF DIFFERENTIAL EQUATIONS, 272, 330-369 [10.1016/j.jde.2020.09.004].

Availability:
This version is available at: https://hdl.handle.net/11585/818098 since: 2021-04-06

Published:
DOI: http://doi.org/10.1016/j.jde.2020.09.004

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/).
When citing, please refer to the published version.

(Article begins on next page)

12 October 2024

This is the final peer-reviewed accepted manuscript of:

S. Menozzi, A. Pesce, X. Zhang, Density and gradient estimates for non degenerate Brownian SDEs with unbounded measurable drift, Journal of Differential Equations, Volume 272, 2021, Pages 330-369,

The final published version is available online at: ***https://dx.doi.org/10.1016/j.jde.2020.09.004***

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/)

When citing, please refer to the published version.

Density and gradient estimates for non degenerate Brownian SDEs with unbounded measurable drift

S. Menozzi[*], **A. Pesce**[†], **X. Zhang**[‡]

September 18, 2020

Abstract

We consider non degenerate Brownian SDEs with Hölder continuous in space diffusion coefficient and unbounded drift with linear growth. We derive two sided bounds for the associated density and pointwise controls of its derivatives up to order two under some additional spatial Hölder continuity assumptions on the drift. Importantly, the estimates reflect the transport of the initial condition by the unbounded drift through an auxiliary, possibly regularized, flow.

Keywords: unbounded drift, heat kernel estimates, gradient estimates, parametrix method.
MSC 2010: Primary: 60H10, 35K10; Secondary: 60H30.

1 Introduction

1.1 Statement of the problem

We are interested in providing Aronson-like bounds and corresponding pointwise estimates for the derivatives up to order two for the transition probability density of the following d-dimensional, non-degenerate diffusion

$$\mathrm{d}X_t = b(t, X_t)\mathrm{d}t + \sigma(t, X_t)\mathrm{d}W_t, \quad t \geqslant 0, \ X_0 = x, \tag{1.1}$$

where $(W_t)_{t\geqslant 0}$ is a standard d-dimensional Brownian motion on the classical Wiener space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathbb{P})$. The diffusion coefficient σ is assumed to be rough in time, and Hölder continuous in space. The drift b is assumed to be measurable and to have linear growth in space. Importantly, we will always assume throughout the article that the diffusion coefficient σ is bounded and separated from 0 (usual uniform ellipticity condition).

When both coefficients b, σ are *bounded* and Hölder continuous, it is well known that there exists a unique weak solution to (1.1) which admits a density (see e.g. [26], [16], [10]), i.e. for all $A \in \mathcal{B}(\mathbb{R}^d)$ (Borel σ-field of $\mathbb{R}^d$),

$$\mathbb{P}[X_t \in A | X_0 = x] = \int_A p(0, x, t, y)\mathrm{d}y.$$

Furthermore, it can be proved by the parametrix method that the transition density $p(0, x, t, y)$ enjoys the following two sided Gaussian estimates on a compact set in time:

$$C^{-1} g_{\lambda^{-1}}(t, x-y) \leqslant p(0, x, t, y) \leqslant C g_\lambda(t, x-y) \tag{1.2}$$

as well as the following gradient estimate

$$|\nabla_x^j p(0, x, t, y)| \leqslant C t^{-\frac{j}{2}} g_\lambda(t, x-y), \quad j = 1, 2,$$

where

$$g_\lambda(t, x) := t^{-\frac{d}{2}} \exp\big(-\lambda |x|^2/t\big), \quad \lambda \in (0, 1], t > 0, \tag{1.3}$$

[*]Laboratoire de Modélisation Mathématique d'Evry (LaMME), Université d'Evry Val d'Essonne, 23 Boulevard de France 91037 Evry, France and Laboratory of Stochastic Analysis, HSE, Pokrovsky Blvd, 11, Moscow, Russian Federation. stephane.menozzi@univ-evry.fr

[†]Dipartimento di Matematica, Piazza di Porta San Donato, 5 Bologna (Italy), antonello.pesce2@unibo.it

[‡]School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, P.R. China, Email: XichengZhang@gmail.com

and the constants $\lambda, C \geqslant 1$ only depend on the regularity of the coefficients, the non-degeneracy constants of the diffusion coefficients, the dimension d, and for the constant C, on the maximal time considered (see [15] and [1], [2]). Such methods have been successfully applied to derive upper bounds up to the second order derivative for more general cases, such as operators satisfying a strong Hörmander condition (see [5]), and also Kolmogorov operators with linear drift (see [24] and [13]). A different approach consists in viewing a logarithmic transformation of p as the value function of a certain stochastic control problem (see [14]): such idea allows then to get the desired density estimates by choosing appropriate controls and eventually an upper bound for the *logarithmic gradient* (see [25]).

When the drift is unbounded and non-linear, fewer results are available. In fact, in this case it is no longer expected that the two sided estimates as given in (1.2) hold. Consider for instance the following Ornstein-Uhlenbeck (OU)-process

$$\mathrm{d}X_t = X_t \mathrm{d}t + \mathrm{d}W_t, \quad X_0 = x,$$

which has, with the notations of (1.3), the non-spatial homeogenous density

$$p_{\mathrm{OU}}(0, x, t, y) = (\pi(\mathrm{e}^{2t} - 1))^{-d/2} g_{t/(\mathrm{e}^{2t}-1)}(t, \mathrm{e}^t x - y).$$

In [12] the authors derive two sided density bounds for a class of degenerate operators with *unbounded and Lipschitz drift*, satisfying a weak Hörmander condition, by combining the two previous approaches: parametrix and logarithmic transform. Indeed, when the drift is unbounded it becomes difficult to get good controls for the iterated kernels in the *blunt* parametrix expansion. In our non-degenerate parabolic setting those bounds still hold provided the drift is *globally* Lipschitz continuous in space. Then, they read as:

$$C^{-1} g_{\lambda^{-1}}(t, \theta_t(x) - y) \leqslant p(0, x, t, y) \leqslant C g_\lambda(t, \theta_t(x) - y), \tag{1.4}$$

where θ stands for the deterministic flow associated with the drift, i.e.

$$\dot{\theta}_t(x) = b(t, \theta_t(x)), \ t \geqslant 0, \ \theta_0(x) = x$$

and $C, \lambda > 0$ enjoy the same type of dependence as in (1.2) but importantly λ now also depends on the maximal time considered. This means that the diffusion starting from x, oscillates around $\theta_t(x)$ at time t with fluctuations of order $t^{-\frac{1}{2}}$. Notice that if b is bounded, then (1.4) reduces to (1.2) since

$$t^{-\frac{1}{2}}|x - y| - \|b\|_\infty t^{\frac{1}{2}} \leqslant t^{-\frac{1}{2}}|\theta_t(x) - y| \leqslant t^{-\frac{1}{2}}|x - y| + \|b\|_\infty t^{\frac{1}{2}}.$$

Hence, taking or not into consideration the flow does not give much additional information. The above control also clearly emphasizes why C might depend on some maximal time interval considered. In the case where b is bounded but not necessarily *smooth*, the above bounds remain valid for *any* regularizing flow.

Diffusion with dynamics (1.1) and unbounded drifts appear in many applicative fields. Let us for instance mention the work [17] which was concerned with issues related to statistics of diffusions. We can also refer to [21] for the numerical approximation of ergodic diffusions.

In such frameworks, estimates on the density and its derivatives are naturally required. Some gradient estimates of the density were established in [17]. The approach developed therein relies on the Malliavin calculus. It thus required some extra regularity on the drift. Also, since the deterministic flow was not taken into consideration an additional penalizing exponential term in the right hand side[1] of the bounds appeared. Similar features appeared in the work [11] which established the existence of fundamental solutions for a strictly sublinear Hölder continuous drift.

We can also mention the work of Metafune *et al.* [19] in which non-Gaussian upper bounds are derived for the heat kernel associated with divergence form operators with unbounded one-sided drift (i.e. the divergence of the drift is bounded by a constant). The authors also obtain therein a diagonal lower bound when the diffusion coefficient is symmetric. Let us also refer to Boutiah *et al.* [7] who derive sharp two sided-bounds for the heat-kernel associated with very special operators which also involve unbounded terms in the diffusion part, which are also specifically linked to the growth of the drift term.

The point of the current work is to establish estimates for the derivatives that reflect both the singularities associated with the differentiation in the parabolic setting, as in equation (1.2) above, and also reflect the key importance of the flow for unbounded drifts as it appears in the two-sided heat kernel estimate (1.4). To the best of our knowledge, our is one of the first results for the derivatives of heat kernels with unbounded drifts.

We can actually address various frameworks. We manage to obtain two-sided heat kernel bounds for a Hölder continuous in space diffusion coefficient σ in (1.1) and a drift b which is uniformly bounded in time at

[1] r.h.s. in short

the origin and has linear growth in space (see assumptions $(\mathbf{H}^\sigma_\alpha)$ and $(\mathbf{H}^b_\beta)$ below). Importantly, when the drift b is itself not smooth, the heat kernel bounds can be stated in the form (1.4) for any flow associated with a mollification of b. In particular, if the drift is continuous in space they actually hold for *any* Peano flow. Those conditions are also sufficient to obtain gradient bounds w.r.t. the *backward* variable x. To derive controls for the second order derivatives w.r.t. x an additional spatial Hölder continuity assumption naturally appears for the drift. Eventually, imposing some additional spatial smoothness on the diffusion coefficient, we also succeed to establish a gradient bound w.r.t. the *forward* variable y.

The paper is organized as follows. Our main results are stated in details in Section 1.2; Section 2 is dedicated to the proof of our main results when the coefficients satisfy our previous assumptions and are also *smooth*. Importantly, we prove that the two-sided heat-kernel bounds do not depend on the smoothness of the coefficients but only on constants appearing in $(\mathbf{H}^\sigma_\alpha)$ and $(\mathbf{H}^b_\beta)$, the fixed final time horizon $T > 0$ and the dimension d. We also establish there bounds for the derivatives through Malliavin calculus techniques which is precisely possible because the coefficients are smooth. Those bounds serve as *a priori* controls to derive in Section 3, through a circular type argument based on the *Duhamel-parametrix type* representation of the density, that those bounds actually do not depend on the smoothness of the coefficients. We then deduce the main results passing to the limit in a mollification procedure through convergence in law and compactness arguments. We eventually discuss in Section 4 some possible extensions for the estimation of higher order derivatives of the heat kernel when the coefficients have some additional smoothness properties.

Let us mention that our approach developed here had previously been successfully used in [8] and [9] to derive respectively the strong well-posedness of kinetic degenerate SDEs or Schauder estimates for degenerate Kolmogorov equations. It actually seems sufficiently robust to be generalized, as soon as some suitable two sided bounds hold, in order to obtain estimates of the derivatives of the density.

1.2 Assumptions and main results

We make the following assumptions about σ and b in (1.1).

$(\mathbf{H}^\sigma_\alpha)$ (Non degeneracy). There exists a positive constant $\kappa_0 \geqslant 1$, such that

$$\kappa_0^{-1}|\xi|^2 \leqslant \langle \sigma\sigma^*(t,x)\xi,\xi\rangle \leqslant \kappa_0|\xi|^2, \quad x,\xi \in \mathbb{R}^d,\ t \geqslant 0, \tag{1.5}$$

and for some $\alpha \in (0,1)$,

$$|\sigma(t,x) - \sigma(t,y)| \leqslant \kappa_0|x-y|^\alpha, \quad t \geqslant 0,\ x,y \in \mathbb{R}^d. \tag{1.6}$$

$(\mathbf{H}^b_\beta)$ There exist positive constant $\kappa_1 > 0$ and $\beta \in [0,1]$ such that for all $x,y \in \mathbb{R}^d$ and $t \geqslant 0$,

$$|b(t,0)| \leqslant \kappa_1,\ |b(t,x) - b(t,y)| \leqslant \kappa_1(|x-y|^\beta \vee |x-y|). \tag{1.7}$$

It should be noticed that under $(\mathbf{H}^b_0)$, b can possibly be an *unbounded* measurable function with linear growth. For instance, $b(t,x) = x + b_0(t,x)$ with b_0 being bounded measurable satisfies (1.7). The drift $b(t,x) = c_1(t) + c_2(t)|x|^\beta,\ \beta \in [0,1]$ where c_1, c_2 are bounded measurable functions of time, also enters this class. We mention that the second condition in (1.7) could be stated only *locally*, i.e. for e.g. $|x-y| \leqslant 1$. This could be checked throughout the proofs below, we prefer to keep it global for simplicity.

Moreover, under $(\mathbf{H}^\sigma_\alpha)$ and $(\mathbf{H}^b_0)$, for any $(s,x) \in \mathbb{R}_+ \times \mathbb{R}^d$, it is well known that there exists a unique weak solution to (1.1) starting from x at time s (see e.g. [26], [4], [12], [18]).

For any $T \in (0,\infty]$ and $\varepsilon \in [0,T)$, we write

$$\mathbb{D}^T_\varepsilon := \{(s,t) \in [0,\infty)^2 : \varepsilon < t-s < T\}.$$

To state our main result, we need to prepare some deterministic regularized flow associated with the drift b. Let ρ be a nonnegative smooth function with support in the unit ball of $\mathbb{R}^d$ and such that $\int_{\mathbb{R}^d} \rho(x)\mathrm{d}x = 1$. For $\varepsilon \in (0,1]$, define

$$\rho_\varepsilon(x) := \varepsilon^{-d}\rho(\varepsilon^{-1}x), \quad b_\varepsilon(t,x) := b(t,\cdot) * \rho_\varepsilon(x) = \int_{\mathbb{R}^d} b(t,y)\rho_\varepsilon(x-y)dy, \tag{1.8}$$

i.e. $*$ stands for the usual spatial convolution. Then for each $n = 1, 2, \cdots$, it is easy to see that

$$\begin{aligned}|\nabla_x^n b_\varepsilon(t,x)| &= \left|\int_{\mathbb{R}^d} (b(t,y) - b(t,x))\nabla_x^n \rho_\varepsilon(x-y)\mathrm{d}y\right| \\ &\leqslant \int_{\mathbb{R}^d} |b(t,y) - b(t,x)||\nabla_x^n \rho_\varepsilon|(x-y)\mathrm{d}y \\ &\leqslant \kappa_1 \varepsilon^\beta \int_{\mathbb{R}^d} |\nabla_x^n \rho_\varepsilon|(x-y)\mathrm{d}y \leqslant c\varepsilon^{\beta-n}. \end{aligned} \tag{1.9}$$

On the other hand, from (1.7) we also have

$$|b_\varepsilon(t,x) - b(t,x)| \leqslant \int_{\mathbb{R}^d} |b(t,y) - b(t,x)|\rho_\varepsilon(x-y)\mathrm{d}y \leqslant \kappa_1 \varepsilon^\beta. \tag{1.10}$$

For fixed $(s,x) \in \mathbb{R}_+ \times \mathbb{R}^d$, we denote by $\theta_{t,s}^{(\varepsilon)}(x)$ the deterministic flow solving

$$\dot{\theta}_{t,s}^{(\varepsilon)}(x) = b_\varepsilon(t, \theta_{t,s}^{(\varepsilon)}(x)),\ t \geqslant 0, \quad \theta_{s,s}^{(\varepsilon)}(x) = x. \tag{1.11}$$

Note that $(\theta_{t,s}^{(\varepsilon)}(x))_{t\geqslant s}$ stands for a forward flow and $(\theta_{t,s}^{(\varepsilon)}(x))_{t\leqslant s}$ stands for a backward flow. Also, since we have regularized equation (1.11) is well posed.

The following lemma, which provides a kind of equivalence between mollified flows, is our starting point for treating the unbounded rough drifts.

Lemma 1.1 (Equivalence of flows)**.** *Under* $(\mathbf{H}_0^b)$*, for any* $\varepsilon \in (0,1]$*, the mapping* $x \mapsto \theta_{t,s}^{(\varepsilon)}(x)$ *is a* C^∞*-diffeomorphism and its inverse is given by* $x \mapsto \theta_{s,t}^{(\varepsilon)}(x)$*. Moreover, for any* $T > 0$*, there exists a constant* $C = C(T, \kappa_1, d) \geqslant 1$ *such that for any* $\varepsilon \in (0,1]$*, all* $|t-s| \leqslant T$ *and* $x, y \in \mathbb{R}^d$*,*

$$|\theta_{t,s}^{(1)}(x) - y| + |t-s| \asymp_C |\theta_{t,s}^{(\varepsilon)}(x) - y| + |t-s| \asymp_C |x - \theta_{s,t}^{(\varepsilon)}(y)| + |t-s|, \tag{1.12}$$

where $Q_1 \asymp_C Q_2$ *means that* $C^{-1}Q_2 \leqslant Q_1 \leqslant CQ_2$*.*

Proof. By (1.9), it is a classical fact that $x \mapsto \theta_{t,s}^{(\varepsilon)}(x)$ is a C^∞-diffeomorphism and its inverse is given by $x \mapsto \theta_{s,t}^{(\varepsilon)}(x)$. Below, without loss of generality, we assume $s < t$. By (1.11) and (1.8), (1.9), (1.10), we have

$$\begin{aligned}|\theta_{t,s}^{(\varepsilon)}(x) - \theta_{t,s}^{(1)}(x)| &\leqslant \int_s^t \left|b_\varepsilon(r, \theta_{r,s}^{(\varepsilon)}(x)) - b_1(r, \theta_{r,s}^{(\varepsilon)}(x))\right| \mathrm{d}r + \int_s^t \left|b_1(r, \theta_{r,s}^{(\varepsilon)}(x)) - b_1(r, \theta_{r,s}^{(1)}(x))\right| \mathrm{d}r \\ &\leqslant 2\kappa_1(t-s) + \|\nabla b_1\|_\infty \int_s^t |\theta_{r,s}^{(\varepsilon)}(x) - \theta_{r,s}^{(1)}(x)|\mathrm{d}r,\end{aligned}$$

which implies by the Gronwall inequality that

$$|\theta_{t,s}^{(\varepsilon)}(x) - \theta_{t,s}^{(1)}(x)| \leqslant 2\kappa_1(t-s)\mathrm{e}^{\|\nabla b_1\|_\infty (t-s)}.$$

Thus we obtain

$$|\theta_{t,s}^{(1)}(x) - y| \leqslant |\theta_{t,s}^{(\varepsilon)}(x) - y| + 2\kappa_1 \mathrm{e}^{\|\nabla b_1\|_\infty (t-s)}|t-s|.$$

By symmetry, we obtain the first $\asymp_C$. For the second one, note that by the Gronwall inequality,

$$|\theta_{t,s}^{(1)}(x) - \theta_{t,s}^{(1)}(y)| \asymp_{\mathrm{e}^{\|\nabla b_1\|_\infty (t-s)}} |x-y| \Rightarrow |\theta_{t,s}^{(1)}(x) - y| \asymp_{\mathrm{e}^{\|\nabla b_1\|_\infty (t-s)}} |x - \theta_{s,t}^{(1)}(y)|. \tag{1.13}$$

From this, by the first $\asymp_C$, we obtain the second $\asymp_C$. □

We introduce for notational convenience the following parameter set which gathers important quantities appearing in the assumptions:

$$\Theta := (T, \alpha, \beta, \kappa_0, \kappa_1, d), \tag{1.14}$$

where again $T > 0$ stands for the fixed considered final time, α denotes the Hölder regularity index of the diffusion coefficient σ (see (1.6)), κ_0 is the uniform ellipticity constant in (1.5), κ_1 and β are related to the behavior of the drift b in (1.7), and d is the current underlying dimension.

Our main result is the following theorem.

Theorem 1.2. *Let $\alpha \in (0,1]$. Under $(\mathbf{H}^\sigma_\alpha)$ and $(\mathbf{H}^b_0)$, for any $T>0$, $(s,t)\in \mathbb{D}^T_0$ and $x\in\mathbb{R}^d$, the unique weak solution $X_{t,s}(x)$ of (1.1) starting from x at time s admits a density $p(s,x,t,y)$ which is continuous in $x,y\in\mathbb{R}^d$. Moreover, $p(s,x,t,y)$ enjoys the following estimates:*

(i) (Two-sided density bounds) There exist constants $\lambda_0\in(0,1]$, $C_0\geqslant 1$ depending on Θ such that for any $(s,t)\in\mathbb{D}^T_0$ and $x,y\in\mathbb{R}^d$,

$$C_0^{-1} g_{\lambda_0^{-1}}(t-s,\theta^{(1)}_{t,s}(x)-y)\leqslant p(s,x,t,y)\leqslant C_0 g_{\lambda_0}(t-s,\theta^{(1)}_{t,s}(x)-y). \tag{1.15}$$

(ii) (Gradient estimate in x) There exist constants $\lambda_1\in(0,1]$, $C_1\geqslant 1$ depending on Θ such that for any $(s,t)\in\mathbb{D}^T_0$ and $x,y\in\mathbb{R}^d$,

$$|\nabla_x p(s,x,t,y)|\leqslant C_1(t-s)^{-\frac{1}{2}} g_{\lambda_1}(t-s,\theta^{(1)}_{t,s}(x)-y). \tag{1.16}$$

(iii) (Second order derivative estimate in x) If $(\mathbf{H}^b_\beta)$ holds for some $\beta\in(0,1]$, then there exist constants $\lambda_2, C_2>0$ depending on Θ such that for any $(s,t)\in\mathbb{D}^T_0$ and $x,y\in\mathbb{R}^d$,

$$\left|\nabla^2_x p(s,x,t,y)\right|\leqslant C_2(t-s)^{-1} g_{\lambda_2}(t-s,\theta^{(1)}_{t,s}(x)-y). \tag{1.17}$$

(iv) (Gradient estimate in y) If $(\mathbf{H}^b_\beta)$ holds for some $\beta\in(0,1]$ and for some $\alpha\in(0,1)$ and $\kappa_2>0$,

$$\|\nabla\sigma\|_\infty\leqslant\kappa_2,\quad |\nabla\sigma(t,x)-\nabla\sigma(t,y)|\leqslant\kappa_2|x-y|^\alpha, \tag{1.18}$$

then there exists constants $\lambda_3\in(0,1]$, $C_3\geqslant 1$ depending on Θ and κ_2 such that for any $(s,t)\in\mathbb{D}^T_0$ and $x,y\in\mathbb{R}^d$,

$$|\nabla_y p(s,x,t,y)|\leqslant C_3(t-s)^{-\frac{1}{2}} g_{\lambda_3}(t-s,\theta^{(1)}_{t,s}(x)-y). \tag{1.19}$$

Remark 1.3. By Lemma 1.1, the above $\theta^{(1)}_{t,s}(x)$ can be replaced by any regularized flow $\theta^{(\varepsilon)}_{t,s}(x)$. Importantly, if b satisfies $(\mathbf{H}^b_\beta)$ for some $\beta\in(0,1]$, then $\theta^{(1)}_{t,s}(x)$ can be replaced as well by **any** Peano flow solving $\dot\theta_{t,s}(x)=b(t,\theta_{t,s}(x))$, $\theta_{s,s}(x)=x$. Indeed, it is plain to check that, in this case, the result of Lemma 1.1 still holds with $\theta_{t,s}(x)$ instead of $\theta^{(\varepsilon)}_{t,s}(x)$.

Remark 1.4. Under the assumptions of the theorem, in fact, we can show the Hölder continuiy of $\nabla_x p$, $\nabla^2_x p$ and $\nabla_y p$ in the variables x and y (see Appendix A).

1.3 Notations

In the following we will denote by $\langle\cdot,\cdot\rangle$ and $|\cdot|$ the Euclidean scalar product and norm on $\mathbb{R}^d$. We will use the notation ∇,∇^2 to denote respectively the gradient and Hessian matrix of real valued smooth functions on $\mathbb{R}^d$. By extension, we denote by ∇^j the jth order derivative of a smooth function. We use the letter C for constants that may depend on the parameters in $\Theta=(T,\alpha,\beta,\kappa_0,\kappa_1,d)$ introduced in (1.14). We reserve the notation c for constant which only depend on the quantities $\alpha,\beta,\kappa_0,\kappa_1,d$ but not on T. Both type of constants c,C might change from line to line. Other possible dependencies are explicitly indicated when needed. Eventually, we will frequently use the notation $\lesssim$. For two quantities Q_1 and Q_2, we mean by $Q_1\lesssim Q_2$ that there exists C with the same previously described dependence such that $Q_1\leqslant CQ_2$.

2 A priori heat kernel estimates for SDEs with smooth coefficients

In this section we suppose $(\mathbf{H}^\sigma_\alpha)$ and $(\mathbf{H}^b_\beta)$, and consider the mollified b_ε and σ_ε. In particular, we have

$$\kappa^{(\varepsilon)}_j:=\sum_{k=1,\cdots,j}\left(\|\nabla^k_x b_\varepsilon\|_\infty+\|\nabla^k_x\sigma_\varepsilon\|_\infty\right)<\infty,\quad\forall j\in\mathbb{N}. \tag{2.1}$$

In the following, for simplicity of notations, we shall drop the subscripts ε. In other words, we assume

(**S**) b and σ satisfy $(\mathbf{H}^\sigma_\alpha)$, $(\mathbf{H}^b_\beta)$ and (2.1).

Under (**S**), it is well known that for each $(s,x)\in\mathbb{R}_+\times\mathbb{R}^d$, the following SDE has a unique strong solution:

$$\mathrm{d}X_{t,s}=b(t,X_{t,s})\mathrm{d}t+\sigma(t,X_{t,s})\mathrm{d}W_t,\quad X_{s,s}=x,\quad t\geqslant s. \tag{2.3}$$

The following theorem is well known and more or less standard in the theory of the Malliavin calculus. We refer to [20], [27, Remarks 2.1 and 2.2] or [29, Theorem 5.4] for more details.

Theorem 2.1. *Under* (**S**)*, for any $j,j'\in\mathbb{N}$, $p>1$ and $T>0$, there is a constant $C_p=C_p(\Theta,j,j',\kappa_{j+j'})$ such that for all $(s,t)\in\mathbb{D}_0^T$, $x\in\mathbb{R}^d$ and $f\in C_b^\infty(\mathbb{R}^d)$,*

$$|\nabla^j\mathbb{E}(\nabla^{j'}f(X_{t,s}(x)))|\leqslant C_p(t-s)^{-(j+j')/2}(\mathbb{E}|f(X_{t,s}(x)|^p)^{1/p}. \tag{2.4}$$

In particular, $X_{t,s}(x)$ has a density $p(s,x,t,y)$, which is smooth in x,y.

Remark 2.2. By Itô's formula, one sees that $p(s,x,t,y)$ satisfies the backward Kolmogorov equation

$$\partial_s p(s,x,t,y)+\mathcal{L}_{s,x}p(s,x,t,y)=0,\ p(s,\cdot,t,y)\longrightarrow\delta_y(\cdot)\text{ weakly as }s\uparrow t, \tag{2.5}$$

and the forward Kolmogorov equation (Fokker-Planck equation):

$$\partial_t p(s,x,t,y)-\mathcal{L}^*_{t,y}p(s,x,t,y)=0,\ p(s,x,t,\cdot)\longrightarrow\delta_x(\cdot)\text{ weakly as }t\downarrow s, \tag{2.6}$$

where, setting $a=\sigma\sigma^*/2$,

$$\mathcal{L}_{s,x}f(x)=\mathrm{Tr}\big(a(s,x)\nabla^2_x f(x)\big)+\langle b(s,x),\nabla_x f(x)\rangle$$

and

$$\mathcal{L}^*_{t,y}f(y)=\partial_{y_i}\partial_{y_j}(a_{ij}(t,y)f(y))-\mathrm{div}(b(t,\cdot)f)(y).$$

Here, we use the usual Einstein convention for the adjoint operator.

2.1 The Duhamel representation for $p(s,x,t,y)$

Fix now $(\tau,\xi)\in\mathbb{R}_+\times\mathbb{R}^d$ as *freezing parameters* to be chosen later on. Let $\widetilde{X}^{(\tau,\xi)}_{t,s}(x)$ denote the process starting at x at time s, with dynamics

$$d\widetilde{X}^{(\tau,\xi)}_{t,s}=b(t,\theta_{t,\tau}(\xi))\mathrm{d}t+\sigma(t,\theta_{t,\tau}(\xi))\mathrm{d}W_t,\ t\geqslant s,\ \widetilde{X}^{(\tau,\xi)}_{s,s}=x, \tag{2.7}$$

i.e. $\widetilde{X}^{(\tau,\xi)}_{t,s}$ denotes the process derived from (2.3), when freezing the spatial coefficients along the flow $\theta_{\cdot,\tau}(\xi)$, where $\theta_{\cdot,\tau}(\xi)$ is the unique solution of ODE (1.11) corresponding to b. It is clear that, for any choice of freezing couple (τ,ξ), $\widetilde{X}^{(\tau,\xi)}_{t,s}$ has a Gaussian density

$$\widetilde{p}^{(\tau,\xi)}(s,x,t,y)=\frac{\exp\{-\langle(\mathcal{C}^{(\tau,\xi)}_{t,s})^{-1}(\vartheta^{(\tau,\xi)}_{t,s}+x-y),\vartheta^{(\tau,\xi)}_{t,s}+x-y\rangle/2\}}{\sqrt{(2\pi)^d\det(\mathcal{C}^{(\tau,\xi)}_{t,s})}}, \tag{2.8}$$

where

$$\vartheta^{(\tau,\xi)}_{t,s}:=\int_s^t b(r,\theta_{r,\tau}(\xi))\mathrm{d}r,\quad \mathcal{C}^{(\tau,\xi)}_{t,s}:=\int_s^t\sigma\sigma^*(r,\theta_{r,\tau}(\xi))\mathrm{d}r.$$

In particular, $\widetilde{p}^{(\tau,\xi)}(s,x,t,y)$ satisfies for fixed $(t,y)\in\mathbb{R}_+\times\mathbb{R}^d$:

$$\partial_s\widetilde{p}^{(\tau,\xi)}(s,x,t,y)+\mathcal{L}^{(\tau,\xi)}_{s,x}\widetilde{p}^{(\tau,\xi)}(s,x,t,y)=0,\ (s,x)\in[0,t)\times\mathbb{R}^d, \tag{2.9}$$

subjected to the final condition

$$\widetilde{p}^{(\tau,\xi)}(s,\cdot,t,y)\longrightarrow\delta_y(\cdot)\text{ weakly as }s\uparrow t, \tag{2.10}$$

where

$$\mathcal{L}^{(\tau,\xi)}_{s,x}=\mathrm{Tr}\big(a(s,\theta_{s,\tau}(\xi))\cdot\nabla^2_x\big)+\langle b(s,\theta_{s,\tau}(\xi)),\nabla_x\rangle$$

denotes the generator of the diffusion with frozen coefficients in (2.7).

The following lemma is a straightforward consequence of the explicit representation (2.8), the uniform ellipticity condition (1.5) and the chain rule.

Lemma 2.3 (A priori controls for the frozen Gaussian density)**.** *For any $j = 0, 1, 2, \cdots$, there exist constants $\lambda_j, C_j > 0$ depending only on j, κ_0, d such that for all $(\tau, \xi) \in \mathbb{R}_+ \times \mathbb{R}^d$, $(s, t) \in \mathbb{D}_0^\infty$ and $x, y \in \mathbb{R}^d$,*

$$\widetilde{p}^{(\tau,\xi)}(s, x, t, y) \geqslant C_0 g_{\lambda_0^{-1}}\big(t - s, \vartheta_{t,s}^{(\tau,\xi)} + x - y\big),$$

and

$$\Big|\nabla_x^j \widetilde{p}^{(\tau,\xi)}(s, x, t, y)\Big| = \Big|\nabla_y^j \widetilde{p}^{(\tau,\xi)}(s, x, t, y)\Big| \leqslant C_j (t - s)^{-\frac{j}{2}} g_{\lambda_j}\big(t - s, \vartheta_{t,s}^{(\tau,\xi)} + x - y\big).$$

Moreover, for each $j, j' \in \mathbb{N}$, there are constants C', λ' depending on Θ and $\kappa_{j'}$ such that

$$\Big|\nabla_x^j \nabla_\xi^{j'} \widetilde{p}^{(\tau,\xi)}(s, x, t, y)\Big| \leqslant C' (t - s)^{-\frac{j}{2}} g_{\lambda'}\big(t - s, \vartheta_{t,s}^{(\tau,\xi)} + x - y\big). \tag{2.11}$$

Proof. We focus on (2.11) for which it suffices to note that for any $k \in \mathbb{N}$, $T > 0$,

$$|\nabla_\xi^k \vartheta_{t,s}^{(\tau,\xi)}| + |\nabla_\xi^k \mathcal{C}_{t,s}^{(\tau,\xi)}| \leqslant C_k |t - s|, \quad (s, t) \in \mathbb{D}_0^T, \ (\tau, \xi) \in [s, t] \times \mathbb{R}^d,$$

where the constant C_k depends on the bound of $\nabla^j b$ and $\nabla^j \sigma$, $j = 1, \cdots, k$. □

The starting point of our analysis is the following Duhamel type representation formula which readily follows in the current *smooth coefficients* setting from (2.5)-(2.6) and (2.9)-(2.10):

$$p(s, x, t, y) = \widetilde{p}^{(\tau,\xi)}(s, x, t, y) + \int_s^t \int_{\mathbb{R}^d} \widetilde{p}^{(\tau,\xi)}(s, x, r, z) \mathcal{A}_{r,z}^{(\tau,\xi)} p(r, z, t, y) \mathrm{d}z \mathrm{d}r \tag{2.12}$$

$$= \widetilde{p}^{(\tau,\xi)}(s, x, t, y) + \int_s^t \int_{\mathbb{R}^d} p(s, x, r, z) \mathcal{A}_{r,z}^{(\tau,\xi)} \widetilde{p}^{(\tau,\xi)}(r, z, t, y) \mathrm{d}z \mathrm{d}r, \tag{2.13}$$

where

$$\mathcal{A}_{r,z}^{(\tau,\xi)} := \mathcal{L}_{r,z} - \mathcal{L}_{r,z}^{(\tau,\xi)} = \mathrm{tr}(A_{r,z}^{(\tau,\xi)} \cdot \nabla_z^2) + B_{r,z}^{(\tau,\xi)} \cdot \nabla_z \tag{2.14}$$

and

$$A_{r,z}^{(\tau,\xi)} := a(r, z) - a(r, \theta_{r,\tau}(\xi)), \quad B_{r,z}^{(\tau,\xi)} := b(r, z) - b(r, \theta_{r,\tau}(\xi)). \tag{2.15}$$

If we take $(\tau, \xi) = (s, x)$ in (2.12) and set $\widetilde{p}_0(s, x, t, y) := \widetilde{p}^{(s,x)}(s, x, t, y)$, then we obtain the *backward* representation

$$p(s, x, t, y) = \widetilde{p}_0(s, x, t, y) + \int_s^t \int_{\mathbb{R}^d} \widetilde{p}_0(s, x, r, z) \mathcal{A}_{r,z}^{(s,x)} p(r, z, t, y) \mathrm{d}z \mathrm{d}r,$$

and in this case

$$\vartheta_{t,s}^{(s,x)} + x - y = \int_s^t b(r, \theta_{r,s}(x)) \mathrm{d}r + x - y = \theta_{t,s}(x) - y; \tag{2.16}$$

it involves the *forward* deterministic flow $\theta_{t,s}(x)$ in the frozen Gaussian density. If we now take $(\tau, \xi) = (t, y)$ in (2.13) and set $\widetilde{p}_1(s, x, t, y) := \widetilde{p}^{(t,y)}(s, x, t, y)$, we then obtain the *forward* representation

$$p(s, x, t, y) = \widetilde{p}_1(s, x, t, y) + \int_s^t \int_{\mathbb{R}^d} p(s, x, r, z) \mathcal{A}_{r,z}^{(t,y)} \widetilde{p}_1(r, z, t, y) \mathrm{d}z \mathrm{d}r, \tag{2.17}$$

and in this case

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

It involves the *backward* deterministic flow $\theta_{s,t}(y)$ in the frozen Gaussian density.

2.2 Two-sided estimates for the heat kernel

We first deal here with the two-sided estimates for the density in the current *smooth coefficients* setting. Importantly, we emphasize as much as possible that all the controls obtained are actually *independent* of the derivatives of the coefficients, or even of the continuity of the drift b, but only depend on the parameters gathered in Θ introduced in (1.14). We first iterate in Section 2.2.1 the Duhamel representation (2.17) (*forward case*) to obtain the so-called *parametrix series* expansion of the density. We then give some controls related to the smoothing effects in time of the parametrix kernel. A specific feature of the heat kernels associated with unbounded drifts is that the corresponding parametrix series needs to be handled with care. Indeed, it is not straightforward to prove that it converges and some truncation step is needed. This fact was already observed in [12] and we use here a similar kind of argument based on slightly different techniques deriving from the stochastic control representation of some Brownian functionals, see [6], [28] and Section 2.2.2 below.

We can assume here without loss of generality that $T \leqslant 1$. Indeed, once the two-sided estimates are established in this case, they can be easily extended to any compact time interval $[0,T]$ through Gaussian convolutions using the scaling properties (see Lemma 2.9).

2.2.1 Two-sided heat kernel estimates parametrix series

For notational convenience, we write from now on for $(s,t) \in \mathbb{D}_0^T$ and $x,y \in \mathbb{R}^d$,

$$\widetilde{p}_1(s,x,t,y) = \widetilde{p}^{(t,y)}(s,x,t,y), \quad H(s,x,t,y) := \mathcal{A}_{s,x}^{(t,y)}\widetilde{p}_1(s,x,t,y), \tag{2.18}$$

and

$$p \otimes H(s,x,t,y) = \int_s^t \int_{\mathbb{R}^d} p(s,x,r,z)H(r,z,t,y)\mathrm{d}z\mathrm{d}r. \tag{2.19}$$

Thus, from the Duhamel representation (2.17), we have

$$p(s,x,t,y) = \widetilde{p}_1(s,x,t,y) + (p \otimes H)(s,x,t,y). \tag{2.20}$$

For $N \geqslant 2$, by iterating $N-1$-times the identity (2.20), we obtain

$$p(s,x,t,y) = \widetilde{p}_1(s,x,t,y) + \sum_{j=1}^{N-1}(\widetilde{p}_1 \otimes H^{\otimes j})(s,x,t,y) + (p \otimes H^{\otimes N})(s,x,t,y). \tag{2.21}$$

We shall now use the following notational convention without mentioning the flow $\theta_{t,s}^{(1)}(x)$. For $(s,t) \in \mathbb{D}_0^T$, $x,y \in \mathbb{R}^d$, we define for $\lambda > 0$:

$$\mathbf{g}_\lambda(s,x,t,y) := g_\lambda\big(t-s, \theta_{t,s}^{(1)}(x) - y\big) = (t-s)^{-\frac{d}{2}} \exp\big(-\lambda|\theta_{t,s}^{(1)}(x) - y|^2/(t-s)\big), \tag{2.22}$$

recalling (1.3) for the last equality. We therefore derive from Lemmas 2.3 and 1.1 the following lemma.

Lemma 2.4. *For any $T > 0$ and $j = 0,1,2,\cdots$, there exist constants $\widetilde{\lambda}_j, \widetilde{C}_j > 0$ depending only on Θ such that for all $(s,t) \in \mathbb{D}_0^T$ and $x,y \in \mathbb{R}^d$,*

$$\widetilde{p}_1(s,x,t,y) \geqslant \widetilde{C}_0 \mathbf{g}_{\widetilde{\lambda}_0^{-1}}(s,x,t,y),$$

and for all $\alpha \in [0,1]$,

$$|x - \theta_{s,t}(y)|^\alpha \left|\nabla_x^j \widetilde{p}_1(s,x,t,y)\right| \leqslant \widetilde{C}_j (t-s)^{\frac{\alpha}{2}-\frac{j}{2}} \mathbf{g}_{\widetilde{\lambda}_j}(s,x,t,y). \tag{2.23}$$

The following convolution type inequality is also an easy consequence of Lemma 1.1.

Lemma 2.5. *For any $T > 0$, there is an $\varepsilon = \varepsilon(\Theta) \in (0,1)$ such that for any $\lambda > 0$, there is a $C_\varepsilon = C_\varepsilon(\Theta,\lambda) > 0$ such that for all $(s,t) \in \mathbb{D}_0^T$, $r \in [s,t]$ and $x,y \in \mathbb{R}^d$,*

$$\int_{\mathbb{R}^d} \mathbf{g}_\lambda(s,x,r,z)\mathbf{g}_\lambda(r,z,t,y)\mathrm{d}z \leqslant C_\varepsilon \mathbf{g}_{\varepsilon\lambda}(s,x,t,y).$$

Proof. By definition and Lemma 1.1, we have for some $0 < \varepsilon < \varepsilon' < 1$,

$$\begin{aligned}\int_{\mathbb{R}^d} \mathbf{g}_\lambda(s,x,r,z)\mathbf{g}_\lambda(r,z,t,y)\mathrm{d}z &= \int_{\mathbb{R}^d} g_\lambda\big(r-s,\theta^{(1)}_{r,s}(x)-z\big)g_\lambda\big(t-r,\theta^{(1)}_{t,r}(z)-y\big)\mathrm{d}z\\ &\lesssim \int_{\mathbb{R}^d} g_{\varepsilon'\lambda}\big(r-s,\theta^{(1)}_{r,s}(x)-z\big)g_{\varepsilon'\lambda}\big(t-r,z-\theta^{(1)}_{r,t}(y)\big)\mathrm{d}z\\ &= Cg_{\varepsilon'\lambda}\big(t-s,\theta^{(1)}_{r,s}(x)-\theta^{(1)}_{r,t}(y)\big) \lesssim \mathbf{g}_{\varepsilon\lambda}(s,x,t,y),\end{aligned}$$

where the second equality is due to the Chapman-Kolmogorov (C-K in short) property for the Gaussian semigroup, and the last inequality again follows from Lemma 1.1 and the following control

$$|\theta^{(1)}_{t,s}(x)-y| = |\theta^{(1)}_{t,r}\circ\theta^{(1)}_{r,s}(x)-\theta^{(1)}_{t,r}\circ\theta^{(1)}_{r,t}(y)| \lesssim |\theta^{(1)}_{r,s}(x)-\theta^{(1)}_{r,t}(y)|. \tag{2.24}$$

The proof is complete. □

The next lemma is crucial since it actually allows to control the iterated convolutions of the parametrix kernel H which appears in the expansion (2.21).

Lemma 2.6 (Control of the iterated parametrix kernel)**.** *Under* $(\mathbf{H}^\sigma_\alpha)$ *and* $(\mathbf{H}^b_0)$*, for any* $T>0$ *and* $N\in\mathbb{N}$*, there are constants* $C_N,\lambda_N>0$ *depending only on* Θ *such that for all* $(s,t)\in\mathbb{D}^T_0$ *and* $x,y\in\mathbb{R}^d$*,*

$$|H^{\otimes N}(s,x,t,y)| \leqslant C_N(t-s)^{-1+\frac{N\alpha}{2}}\mathbf{g}_{\lambda_N}(s,x,t,y),$$

where $\lambda_N\to 0$ *as* $N\to\infty$*.*

Proof. By the definition of H in (2.18), (2.15), Lemma 2.4 and (1.12), there exists $\lambda:=\lambda(\Theta)>0$ and $C:=C(\Theta)$ s.t.

$$\begin{aligned}|H(s,x,t,y)| &\leqslant |a(s,x)-a(s,\theta_{s,t}(y))|\cdot|\nabla^2_x\widetilde{p}_1(s,x,t,y)| + |b(s,x)-b(s,\theta_{s,t}(y))|\cdot|\nabla_x\widetilde{p}_1(s,x,t,y)|\\ &\lesssim |x-\theta_{s,t}(y)|^\alpha|\nabla^2_x\widetilde{p}_1(s,x,t,y)| + (1+|x-\theta_{s,t}(y)|)|\nabla_x\widetilde{p}_1(s,x,t,y)|\\ &\lesssim (t-s)^{-1+\frac{\alpha}{2}}\mathbf{g}_\lambda(s,x,t,y).\end{aligned} \tag{2.25}$$

This gives the stated estimate for $N=1$. From (2.25) and by Lemma 2.5, it is readily seen that:

$$|H^{\otimes 2}(s,x,t,y)| \lesssim \left(\int_s^t (r-s)^{-1+\frac{\alpha}{2}}(t-r)^{-1+\frac{\alpha}{2}}\mathrm{d}r\right)\mathbf{g}_{\varepsilon\lambda}(s,x,t,y) \lesssim (t-s)^{-1+\alpha}\mathbf{g}_{\varepsilon\lambda}(s,x,t,y).$$

For general $N\geqslant 2$, by direct induction we have

$$|H^{\otimes N}(s,x,t,y)| \leqslant C_N(t-s)^{-1+\frac{N\alpha}{2}}\mathbf{g}_{\varepsilon^{N-1}\lambda}(s,x,t,y).$$

The proof is complete. □

From the above lemma, (2.21) and (2.23), we thus derive that for all $N\in\mathbb{N}$, $(s,t)\in\mathbb{D}^T_0$, $x,y\in\mathbb{R}^d$:

$$p(s,x,t,y) \leqslant \bar{C}\mathbf{g}_{\lambda_{N-1}}(s,x,t,y) + |p\otimes H^{\otimes N}(s,x,t,y)|, \tag{2.26}$$

which is *almost* the expected upper-bound except that we explicitly have to control the remainder to stop the iteration at some fixed N to avoid the collapse to 0 of λ_N as N goes to infinity. This is precisely the purpose of the next subsection.

2.2.2 Stochastic control arguments and truncation of the parametrix series

In this section, we aim at controlling the remainder term $(p\otimes H^{\otimes N})(s,x,t,y)$ in the almost Gaussian upper-bound (2.26).

To this end, we use the variational representation formula to show the a priori derivative estimates of heat kernel when the coefficients are smooth. The idea is essentially the same as in [12]. The following variational representation formula was first proved by Boué and Dupuis [6]. The reader is referred to [28] for an extension to the abstract Wiener space.

Theorem 2.7. *Fix $T>0$ and let F be a bounded Wiener functional on the classical Wiener space $(\Omega,\mathscr{F},\mathbb{P})$*[2] *which is $\mathscr{F}_T$ measurable. Then it holds that*

$$-\ln \mathbb{E}\mathrm{e}^F=\inf_{h\in\mathcal{S}}\mathbb{E}\left(\frac{1}{2}\int_0^T|\dot{h}(s)|^2\mathrm{d}s-F(\omega+h)\right),$$

where $\mathcal{S}$ denotes the set of all $\mathbb{R}^d$-valued $\mathscr{F}_t$-adapted and absolutely continuous processes with

$$\mathbb{E}\int_0^T|\dot{h}(s)|^2\mathrm{d}s<\infty.$$

Using the above variational representation formula, we obtain the following important lemma.

Lemma 2.8. *Let $\ell:\mathbb{R}^d\to(0,\infty)$ be a measurable function s.t. for all $x\in\mathbb{R}^d,\zeta^{-1}\leqslant\ell(x)\leqslant\zeta$ for some $\zeta\geqslant 1$. Under $(\mathbf{H}^\sigma_\alpha)$ and $(\mathbf{H}^b_0)$, for any $T>0$, there is a constant $C=C(\Theta)>0$ such that for all $x\in\mathbb{R}^d$ and $(s,t)\in\mathbb{D}^T_0$,*

$$\mathbb{E}\ell(X_{t,s}(x))\leqslant C\sup_{z\in\mathbb{R}^d}\exp\Big\{\ln\ell(z)-C^{-1}|z-\theta^{(1)}_{t,s}(x)|^2\Big\}.$$

Proof. Without loss of generality, we assume $s=0$ and write $X_t:=X_{t,s}(x)$. Since in the current smooth coefficients setting X_t can be viewed as a functional of the Brownian path, taking $F=\ln(\ell(X_t))$, we derive from Theorem 2.7 that

$$-\ln\mathbb{E}\ell(X_t)=\inf_{h\in\mathcal{S}}\mathbb{E}\left(\frac{1}{2}\int_0^t|\dot{h}(s)|^2\mathrm{d}s-\ln\ell(X^h_t)\right),$$

where X^h solves the following SDE:

$$\mathrm{d}X^h_t=\Big(b(t,X^h_t)+\sigma(t,X^h_t)\dot{h}(t)\Big)\mathrm{d}t+\sigma(t,X^h_t)\mathrm{d}W_t,\ X^h_0=x,$$

i.e. the control process h enters the dynamics in the drift part. Note that $\theta_t:=\theta_{t,0}(x)$ solves the following ODE:

$$\dot{\theta}_t=b(t,\theta_t),\quad\theta_0=x.$$

By Itô's formula, we have

$$\mathbb{E}|X^h_t-\theta_t|^2=\mathbb{E}\int_0^t\Big(2\langle X^h_s-\theta_s,b(s,X^h_s)-b(s,\theta_s)+\sigma(s,X^h_s)\dot{h}(s)\rangle+(\sigma\sigma^*)(s,X^h_s)\Big)\mathrm{d}s.$$

Recalling

$$|b(t,x)-b(t,y)|\leqslant\kappa_1(1+|x-y|),$$

the Young inequality yields

$$\mathbb{E}|X^h_t-\theta_t|^2\lesssim\mathbb{E}\int_0^t|X^h_s-\theta_s|^2\mathrm{d}s+\mathbb{E}\int_0^t|\dot{h}(s)|^2\mathrm{d}s+t.$$

From the Gronwall inequality, we thus obtain

$$\mathbb{E}|X^h_t-\theta_t|^2\lesssim\mathbb{E}\int_0^t|\dot{h}(s)|^2\mathrm{d}s+t.$$

Hence, for some $C>0$,

$$\frac{1}{2}\mathbb{E}\int_0^t|\dot{h}(s)|^2\mathrm{d}s\geqslant C^{-1}\mathbb{E}|X^h_t-\theta_t|^2-Ct.$$

Therefore, we eventually derive

$$-\ln\mathbb{E}\ell(X_t)\geqslant\inf_{h\in\mathcal{S}}\mathbb{E}\left(C^{-1}|X^h_t-\theta_t|^2-\ln\ell(X^h_t)\right)-C\geqslant\inf_{z\in\mathbb{R}^d}\left(C^{-1}|z-\theta_t|^2-\ln\ell(z)\right)-C.$$

The desired estimate eventually follows from Lemma 1.1. □

[2] We recall that for the Wiener space, the fundamental set $\Omega=C(\mathbb{R}_+,\mathbb{R}^d)$ and in this framework $\omega\in\Omega$ simply stands here for an $\mathbb{R}^d$-valued continuous function. Also, the canonical process $\big(\omega(t)\big)_{t\geqslant 0}$ is a standard d-dimensional Brownian motion under the measure $\mathbb{P}$.

We now state a direct yet important scaling lemma. We refer to Section 2.3 of [12] for additional details.

Lemma 2.9 (Scaling property of the density)**.** *Fix $(s,t) \in \mathbb{D}_0^T$ and let $\lambda := t-s$. Introduce for $u \in [0,1]$, $\widehat{X}_u^\lambda := \lambda^{-\frac{1}{2}} X_{s+u\lambda}$. Then, $(\widehat{X}_u^\lambda)_{u\in[0,1]}$ satisfies the SDE*

$$\mathrm{d}\widehat{X}_u^\lambda = \lambda^{\frac{1}{2}} b\big(s+u\lambda, \widehat{X}_u^\lambda \lambda^{\frac{1}{2}}\big)\mathrm{d}u + \sigma\big(s+u\lambda, \widehat{X}_u^\lambda \lambda^{\frac{1}{2}}\big)\mathrm{d}\widehat{W}_u^\lambda = \widehat{b}^\lambda(u, \widehat{X}_u^\lambda)\mathrm{d}u + \widehat{\sigma}^\lambda(u, \widehat{X}_u^\lambda)\mathrm{d}\widehat{W}_u^\lambda,$$

where $\widehat{W}_u^\lambda = \lambda^{-\frac{1}{2}} W_{u\lambda}$ is a Brownian motion. It also holds that:

$$p(s,x,t,y) = \lambda^{-\frac{d}{2}} \widehat{p}^\lambda\Big(0, \lambda^{-\frac{1}{2}}x, 1, \lambda^{-\frac{1}{2}}y\Big),$$

and introducing for $z \in \mathbb{R}^d$, $u \in [0,1]$, $\partial_u \widehat{\theta}_{u,0}^\lambda(z) = \widehat{b}^\lambda(u, \widehat{\theta}_{u,0}(z))$, $\widehat{\theta}_{0,0}^\lambda(z) = z$,

$$\big|\widehat{\theta}_{1,0}^\lambda(\lambda^{-\frac{1}{2}}x) - \lambda^{-\frac{1}{2}}y\big|^2 = \lambda^{-1}|\theta_{t,s}(x) - y|^2.$$

Proof. We only focus on the last statement. The other ones readily follow from the change of variable. Write:

$$\lambda^{-\frac{1}{2}}\theta_{t,s}(x) = \lambda^{-\frac{1}{2}}x + \lambda^{-\frac{1}{2}}\int_s^t b(r, \theta_{r,s}(x))\mathrm{d}r = \lambda^{-\frac{1}{2}}x + \lambda^{\frac{1}{2}}\int_0^1 b(s+u\lambda, \theta_{s+u\lambda,s}(x))\mathrm{d}u.$$

Setting now for $u \in [0,1]$, $\bar{\theta}_{u,0}(x) = \theta_{s+u\lambda,s}(x)$, the above equation rewrites:

$$\lambda^{-\frac{1}{2}}\bar{\theta}_{1,0}(x) = \lambda^{-\frac{1}{2}}x + \lambda^{\frac{1}{2}}\int_0^1 b(s+u\lambda, \bar{\theta}_{u,0}(x))\mathrm{d}u = \lambda^{-\frac{1}{2}}x + \int_0^1 \widehat{b}^\lambda(u, \lambda^{-\frac{1}{2}}\bar{\theta}_{u,0}(x))\mathrm{d}u$$

from which we readily derive by uniqueness of the solution to the ODE that for $u \in [0,1]$,

$$\lambda^{-\frac{1}{2}}\bar{\theta}_{u,0}(x) = \widehat{\theta}_{u,0}^\lambda(\lambda^{-\frac{1}{2}}x) = \lambda^{-\frac{1}{2}}\theta_{s+u\lambda,s}(x),$$

which gives the statement. □

We will now use the previous Lemmas 2.8 and 2.9 to establish the following result from which the Gaussian upper-bound will readily follow.

Lemma 2.10 (Control of the remainder)**.** *Choose N large enough in order to have:*

$$-1 + \frac{N\alpha}{2} > \frac{d}{2}. \tag{2.27}$$

There exists constants $C_0, \lambda_0 > 0$ depending only on Θ such that for all $(s,t) \in \mathbb{D}_0^T$ and $x, y \in \mathbb{R}^d$,

$$|(p \otimes H^{\otimes N})(s,x,t,y)| \leqslant C_0 \mathbf{g}_{\lambda_0}(s,x,t,y).$$

Proof. From the scaling property obtained in Lemma 2.9 above, we can assume without loss of generality that $s = 0$ and $t = 1$. From the definition in (2.19) and Lemmas 2.6, 2.8, we have

$$\begin{aligned}
|(p \otimes H^{\otimes N})(0,x,1,y)| &\leqslant \int_0^1 \left|\int_{\mathbb{R}^d} p(0,x,r,z)H^{\otimes N}(r,z,1,y)\mathrm{d}z\right|\mathrm{d}r = \int_0^1 |\mathbb{E}H^{\otimes N}(r, X_{r,0}(x), 1, y)|\mathrm{d}r \\
&\leqslant C_N \int_0^1 (1-r)^{-1+\frac{N\alpha}{2}} \mathbb{E}\mathbf{g}_{\lambda_N}(r, X_{r,0}(x), 1, y)\mathrm{d}r \\
&\leqslant C_N \int_0^1 (1-r)^{-1+\frac{N\alpha}{2}} \sup_{z\in\mathbb{R}^d} \exp\Big\{\ln \mathbf{g}_{\lambda_N}(r,z,1,y) - C^{-1}|z - \theta_{r,0}^{(1)}(x)|^2\Big\}\mathrm{d}r.
\end{aligned}$$

Since by (1.3),

$$\ln \mathbf{g}_{\lambda_N}(r,z,1,y) = \ln g_{\lambda_N}(1-r, \theta_{1,r}^{(1)}(z) - y) = -\frac{d}{2}\ln(1-r) - \lambda_N|\theta_{1,r}^{(1)}(z) - y|^2/(1-r),$$

we have

$$\begin{aligned}
&\sup_z(\ln g_{\lambda_N}(1-r,\theta^{(1)}_{1,r}(z)-y)-C^{-1}|z-\theta^{(1)}_{r,0}(x)|^2)\\
&\leqslant -\frac{d}{2}\ln(1-r)-\inf_z(\lambda_N|\theta^{(1)}_{1,r}(z)-y|^2/(1-r)+C^{-1}|z-\theta^{(1)}_{r,0}(x)|^2)\\
&\leqslant -\frac{d}{2}\ln(1-r)-\lambda'_N\inf_z(|z-\theta^{(1)}_{r,1}(y)|^2/(1-r)-C(1-r)+|z-\theta^{(1)}_{r,0}(x)|^2)\\
&\leqslant -\frac{d}{2}\ln(1-r)-\lambda'_N|\theta^{(1)}_{r,1}(y)-\theta^{(1)}_{r,0}(x)|^2/2+C\\
&\leqslant -\frac{d}{2}\ln(1-r)-\lambda''_N|\theta^{(1)}_{1,0}(x)-y|^2+C,
\end{aligned}$$

where the last step is due to (2.24). Therefore, from the condition (2.27) and the above computations, there exist constants $C_0,\lambda_0>0$ depending only on Θ such that

$$|(p\otimes H^{\otimes N})(0,x,1,y)|\leqslant C_0 g_{\lambda_0}(1,\theta^{(1)}_{1,0}(x)-y)=C_0\mathbf{g}_{\lambda_0}(0,x,1,y).$$

The general statement for arbitrary $(s,t)\in\mathbb{D}^T_0$ again follows from the scaling arguments of Lemma 2.9. □

2.2.3 Final derivation of the two-sided heat kernel estimates

We are now in position to prove the following two-sided estimates.

Theorem 2.11. *Under* $(\mathbf{H}^\sigma_\alpha)$ *and* $(\mathbf{H}^b_0)$*, for any* $T>0$*, there exists constants* $C_0,\lambda_0>0$ *depending only on* Θ *such that for all* $(s,t)\in\mathbb{D}^T_0$ *and* $x,y\in\mathbb{R}^d$*,*

$$C_0^{-1}\mathbf{g}_{\lambda_0^{-1}}(s,x,t,y)\leqslant p(s,x,t,y)\leqslant C_0\mathbf{g}_{\lambda_0}(s,x,t,y).$$

Proof. (i) **(Upper bound)** The upper bound is a direct consequence of the expansion (2.26) and the previous Lemma 2.10 up to a possible modification of the constants C_0,λ_0 that anyhow still only depend on Θ.

(ii) **(Lower bound)** By the upper bound and Lemmas 2.6 and 2.5, we get for some $\lambda_1<\lambda_0$ and $\varepsilon\in(0,1)$:

$$|p\otimes H(s,x,t,y)|\lesssim\int_s^t(t-r)^{-1+\frac{\alpha}{2}}\int_{\mathbb{R}^d}\mathbf{g}_{\lambda_1}(s,x,r,z)\mathbf{g}_{\lambda_1}(r,z,t,y)\mathrm{d}z\mathrm{d}r\lesssim(t-s)^{\frac{\alpha}{2}}\mathbf{g}_{\varepsilon\lambda_1}(s,x,t,y).$$

Hence, for $|\theta^{(1)}_{t,s}(x)-y|\leqslant\sqrt{t-s}$, recalling (2.20) and (2.22), we have

$$p(s,x,t,y)\geqslant\left(C_1-C_2(t-s)^{\frac{\alpha}{2}}\right)\mathbf{g}_{\varepsilon\lambda_1}(s,x,t,y)\geqslant\left(C_1-C_2(t-s)^{\frac{\alpha}{2}}\right)(t-s)^{-d/2}\mathrm{e}^{-\varepsilon\lambda_1}.$$

In particular, letting $t-s\leqslant\delta$ with δ small enough, we obtain that

$$p(s,x,t,y)\geqslant C_3(t-s)^{-d/2}\text{ on }\mathbb{D}^\delta_0\text{ and }|\theta^{(1)}_{t,s}(x)-y|\leqslant\sqrt{t-s}. \tag{2.28}$$

We now precisely use a chaining argument to obtain the lower bound when $|\theta^{(1)}_{t,s}(x)-y|\geqslant\sqrt{t-s}$. The idea is to consider a suitable path between the points x and y. The point is then to associate to this path a sequence of balls, with centers on that path, for which the diagonal lower estimate (2.28) holds for points belonging to consecutive balls, and which also have a large enough volume to consent to derive the global *off-diagonal* lower bound from the iterative application of the local *diagonal* one. The usual strategy to build such balls consists in considering the "geodesic" line between x and y. In the non-degenerate case, when the coefficients are bounded, this is nothing but the straight-line joining x and y, see e.g. [3]. When dealing with unbounded coefficients, recall that b has linear growth and is *smooth*, a possibility is to consider the optimal path associated with the deterministic controllability problem $\dot\phi_u=b(u,\phi_u)+\varphi_u,\ u\in[s,t],\ \phi_s=x,\phi_t=y$ with $\varphi\in L^2([s,t],\mathbb{R}^d)$. This was the choice in [12] for a Lipschitz continuous b. The constants in the lower bound estimates obtained therein actually depend on the Lipschitz modulus b. We adopt here a slightly different strategy which only involves the *mollified* flow $\theta^{(1)}$ but which will have the main advantage to provide constants that will again only depend on Θ and not on the smoothness of b, using thoroughly the controls established in Lemma 1.1. We now detail such a construction which is in some sense *original* though pretty natural.

From the scaling arguments of Lemma 2.9, we can assume without loss of generality that $\delta=1$, $s=0$ and $t=1$. Suppose $|\theta^{(1)}_{1,0}(x)-y|>1$ and let M be the smallest integer greater than $4\mathrm{e}^{2\|\nabla b_1\|_\infty}|\theta^{(1)}_{1,0}(x)-y|^2$, i.e.,

$$M-1\leqslant 4\mathrm{e}^{2\|\nabla b_1\|_\infty}|\theta^{(1)}_{1,0}(x)-y|^2<M. \tag{2.29}$$

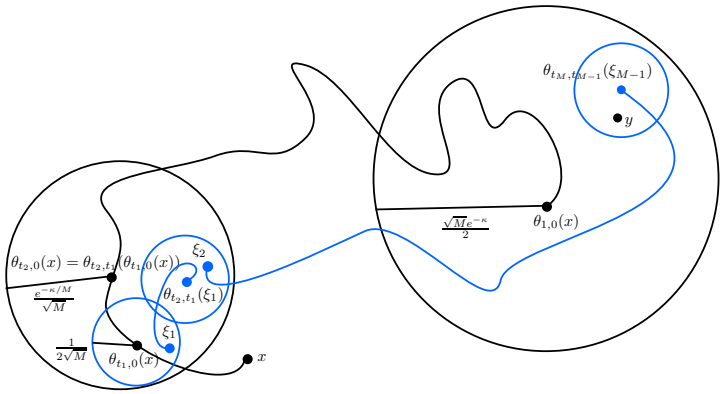

Figure 1: Construction of the chaining balls for the lower bound.

Importantly, we recall from (1.9) that under $(\mathbf{H}_\alpha^\sigma)$ and $(\mathbf{H}_0^b)$, $\|\nabla b_1\|_\infty \leqslant C(\kappa_1)$. Let

$$t_j := j/M, \quad j = 0, 1, \cdots, M.$$

The important point for the proof is the following claim.

Claim: Set $\xi_0 := x$ and $\xi_M := y$. There exist $(M+1)$-points $\xi_0, \xi_1, \cdots, \xi_M$ such that

$$|\xi_{j+1} - \theta^{(1)}_{t_{j+1},t_j}(\xi_j)| \leqslant \tfrac{1}{2\sqrt{M}}, \ j = 0, 1, \cdots, M-1.$$

Indeed, let $Q_1 := B_{1/(2\sqrt{M})}(\theta^{(1)}_{t_1,0}(x))$ and recursively define for $j = 2, \cdots, M$,

$$Q_j := \bigcup_{z \in Q_{j-1}} B_{1/(2\sqrt{M})}(\theta^{(1)}_{t_j,t_{j-1}}(z)) = \Big\{ z : \mathrm{dist}\Big(z, \theta^{(1)}_{t_j,t_{j-1}}(Q_{j-1})\Big) \leqslant 1/(2\sqrt{M}) \Big\}.$$

Letting $\kappa := \|\nabla b_1\|_\infty$ and noting that (see (1.13))

$$\mathrm{e}^{-\kappa/M}|z - z'| \leqslant |\theta^{(1)}_{t_{j+1},t_j}(z) - \theta^{(1)}_{t_{j+1},t_j}(z')| \leqslant \mathrm{e}^{\kappa/M}|z - z'|,$$

by the previous induction method and noting that $\theta^{(1)}_{t_{j+1},t_j} \circ \theta^{(1)}_{t_j,0}(x) = \theta^{(1)}_{t_{j+1},0}(x)$, we have

$$B_{j\mathrm{e}^{-(j-1)\kappa/M}/(2\sqrt{M})}(\theta^{(1)}_{t_j,0}(x)) \subset Q_j, \quad j = 1, 2, \cdots, M.$$

Intuitively, the image of a ball with radius r under the flow $\theta^{(1)}_{t_j,t_{j-1}}$ contains a ball with radius $\mathrm{e}^{-\kappa/M}r$. In particular, by (2.29),

$$\xi_M = y \in B_{\sqrt{M}\mathrm{e}^{-\kappa}/2}(\theta^{(1)}_{1,0}(x)) \subset B_{M\mathrm{e}^{-(M-1)\kappa/M}/(2\sqrt{M})}(\theta^{(1)}_{t_M,0}(x)) \subset Q_M.$$

The claim then follows. The idea of the construction is illustrated in Figure 1.

Now let $\gamma := 1/(2(\mathrm{e}^{\|\nabla b_1\|_\infty} + 1))$ and $z_0 := x$, $z_{M+1} := y$ and $\Sigma_j := B_{\gamma/\sqrt{M}}(\xi_j)$. From the previous claim, we have that for $z_j \in \Sigma_j$ and $z_{j+1} \in \Sigma_{j+1}$,

$$\begin{aligned}
|\theta^{(1)}_{t_{j+1},t_j}(z_j) - z_{j+1}| &\leqslant |\theta^{(1)}_{t_{j+1},t_j}(z_j) - \theta^{(1)}_{t_{j+1},t_j}(\xi_j)| + |\theta^{(1)}_{t_{j+1},t_j}(\xi_j) - \xi_{j+1}| + |\xi_{j+1} - z_{j+1}| \\
&\leqslant \mathrm{e}^{\|\nabla b_1\|_\infty}|z_j - \xi_j| + |\theta^{(1)}_{t_{j+1},t_j}(\xi_j) - \xi_{j+1}| + |\xi_{j+1} - z_{j+1}| \\
&\leqslant \frac{\gamma(\mathrm{e}^{\|\nabla b_1\|_\infty} + 1)}{\sqrt{M}} + \frac{1}{2\sqrt{M}} = \frac{1}{\sqrt{M}} = \sqrt{t_{j+1} - t_j}.
\end{aligned}$$

This precisely means that the previous diagonal lower bound holds for $p(t_j, z_j, t_{j+1}, z_{j+1})$. Thus, by the

Chapman-Kolmogorov equation and (2.28), we have

$$\begin{aligned}
p(0,x,1,y) &= \int_{\mathbb{R}^d}\cdots\int_{\mathbb{R}^d} p(t_0,z_0,t_1,z_1)\cdots p(t_{M-1},z_{M-1},t_M,z_M)\mathrm{d}z_1\cdots\mathrm{d}z_{M-1}\\
&\geqslant \int_{\Sigma_1}\cdots\int_{\Sigma_{M-1}} p(t_0,z_0,t_1,z_1)\cdots p(t_{M-1},z_{M-1},t_M,z_M)\mathrm{d}z_1\cdots\mathrm{d}z_{M-1}\\
&\geqslant (C_3M^{d/2})^M\int_{\Sigma_1}\cdots\int_{\Sigma_{M-1}}\mathrm{d}z_1\cdots\mathrm{d}z_{M-1} = (C_3M^{d/2})^M(M^{-d/2}\gamma^d|B_1|)^{M-1}\\
&= C_3^M M^{d/2}(\gamma^d|B_1|)^{M-1} = M^{d/2}\exp\{M\log(C_3\gamma^d|B_1|)\}/(\gamma^d|B_1|)\\
&\geqslant C_4\exp\{M\log(C_3\gamma^d|B_1|)\}\geqslant C_5\exp\{-C_6|\theta^{(1)}_{1,0}(x)-y|^2\},
\end{aligned}$$

recalling the definition of M in (2.29) and that $C_3\gamma^d|B_1|\leqslant 1$ for the last inequality. □

2.3 Estimates for the derivatives of the heat kernel with smooth coefficients

We insist in the previous section on the fact that, even though we consider smooth coefficients, all our estimates for the two-sided Gaussian bounds are actually uniform w.r.t. Θ which only depends on parameters appearing in $(\mathbf{H}^b_\beta)$ and $(\mathbf{H}^\sigma_\alpha)$.

Our point of view is here different since we mainly want to derive some *a priori bounds* on the derivatives of the heat-kernel when the coefficients are smooth which will then serve in a second time, namely in the circular argument developed in Section 3, to prove that those estimates are actually again independent of the smoothness of the coefficients. Anyhow, in the current section, we fully exploit such a smoothness and obtain controls on the derivatives which *do* depend on the derivatives of b,σ. To this end, we restart from the representation (2.20) of the density and exploit the gradient estimate (2.4).

2.3.1 Proof of the main estimates

Theorem 2.12 (Controls on the derivatives of the heat kernel with smooth coefficients)**.** *Under* **(S)**, *for* $j\in\{1,2\}$, *there exist constants* $C_j := C_j(\Theta,\kappa_j), \lambda_j := \lambda_j(\Theta)>0$, *such that*

$$|\nabla^j_x p(s,x,t,y)|\leqslant C_j(t-s)^{-\frac{j}{2}}\mathbf{g}_{\lambda_j}(s,x,t,y),\ \ |\nabla_y p(s,x,t,y)|\leqslant C_1(t-s)^{-\frac{1}{2}}\mathbf{g}_{\lambda_1}(s,x,t,y).$$

Proof. (i) Let us first establish the estimates on the derivatives w.r.t. the backward variable x. Write from (2.20):

$$\nabla^j_x p(s,x,t,y) = \nabla^j_x\widetilde{p}_1(s,x,t,y)+\nabla^j_x\big(p\otimes H\big)(s,x,t,y).$$

From Lemma 2.4 it readily follows that

$$|\nabla^j_x\widetilde{p}_1(s,x,t,y)|\leqslant C_j(t-s)^{-\frac{j}{2}}\mathbf{g}_{\lambda_j}(s,x,t,y).$$

For the other contribution recall that for $u:=\frac{s+t}{2}$,

$$p\otimes H(s,x,t,y) = \int_u^t \mathbb{E}[H(r,X_{r,s}(x),t,y)]\mathrm{d}r+\int_s^u\mathbb{E}[H(r,X_{r,s}(x),t,y)]\mathrm{d}r =: I_1(x)+I_2(x).$$

For I_1, choosing $p>1$ such that $\frac{d+\alpha-2}{2p}>\frac{d}{2}-1$, by (2.4) and Lemmas 2.6 and 2.5, we get

$$\begin{aligned}
|\nabla^j_x I_1(x)| &\lesssim \int_u^t (r-s)^{-j/2}(\mathbb{E}|H(r,X_{r,s}(x),t,y)|^p)^{1/p}\mathrm{d}r \qquad (2.30)\\
&= \int_u^t (r-s)^{-j/2}\left(\int_{\mathbb{R}^d} p(s,x,r,z)|H(r,z,t,y)|^p\mathrm{d}z\right)^{1/p}\mathrm{d}r\\
&\lesssim (t-s)^{-j/2}\int_u^t (t-r)^{-\frac{1}{p}+\frac{\alpha}{2p}+\frac{d}{2p}-\frac{d}{2}}\left(\int_{\mathbb{R}^d}\mathbf{g}_{\lambda_0}(s,x,r,z)\mathbf{g}_{p\lambda_1}(r,z,t,y)\mathrm{d}z\right)^{1/p}\mathrm{d}r\\
&\lesssim (t-s)^{-j/2}\left(\int_u^t (t-r)^{-\frac{1}{p}+\frac{\alpha}{2p}+\frac{d}{2p}-\frac{d}{2}}\mathrm{d}r\right)\mathbf{g}^{1/p}_{\lambda_2}(s,x,t,y)\\
&\lesssim (t-s)^{-j/2}(t-s)^{1-\frac{1}{p}+\frac{\alpha}{2p}+\frac{d}{2p}-\frac{d}{2}}\mathbf{g}^{1/p}_{\lambda_2}(s,x,t,y)\lesssim (t-s)^{-j/2}\mathbf{g}_{\lambda_2/p}(s,x,t,y).
\end{aligned}$$

To treat $I_2(x)$, we only consider $j=1$ since the case $j=2$ is similar. By the chain rule, we have

$$\nabla_x \mathbb{E}[\big(H(r, X_{r,s}(x), t, y)\big)] = \mathbb{E}[\big(\nabla_x H\big)(r, X_{r,s}(x), t, y) \cdot \nabla_x X_{r,s}(x)],$$

and for all $k \in \{1, \cdots, d\}$,

$$\begin{aligned}\partial_{x_k} H(s,x,t,y) := {} & \mathrm{tr}\big(\partial_{x_k} a(s,x) \cdot \nabla_x^2 \widetilde{p}_1(s,x,t,y)\big) + \partial_{x_k} b(s,x) \cdot \nabla_x \widetilde{p}_1(s,x,t,y) \\ & + \mathrm{tr}(a(s,x) - a(s,\theta_{s,t}(y))) \cdot \partial_{x_k} \nabla_x^2 \widetilde{p}_1(s,x,t,y) \\ & + (b(s,x) - b(s,\theta_{s,t}(y))) \cdot \partial_{x_k} \nabla_x \widetilde{p}_1(s,x,t,y).\end{aligned}$$

Thus by Lemma 2.4, (2.1) and (2.23), it is easy to see that for some $\lambda_3 > 0$,

$$|\nabla_x H(s,x,t,y)| \lesssim (t-s)^{-1} \mathbf{g}_{\lambda_3}(s,x,t,y).$$

We carefully emphasize that the constant implicitly associated with the symbol $\lesssim$ here **does** depend on the smoothness of the coefficients. Using the same argument as above, from the Hölder inequality, one sees that for $p = \frac{d}{d-1}$,

$$\begin{aligned}|\nabla_x I_2(x)| &\lesssim \int_s^u (\mathbb{E}|(\nabla_x H)(r, X_{r,s}(x), t, y)|^p)^{1/p} \mathrm{d}r \\ &\lesssim \int_s^u (t-r)^{-1} (\mathbb{E}\mathbf{g}_{\lambda_3}^p(r, X_{r,s}(x), t, y))^{1/p} \mathrm{d}r \lesssim (t-s)^{-\frac{1}{2}} \mathbf{g}_{\lambda_4}(s,x,t,y).\end{aligned}$$

We thus obtain the gradient estimate in the variable x.

(ii) Let us now turn to the gradient estimate w.r.t. y. We restart from (2.17) differentiating first w.r.t. y. This can be done for arbitrary freezing parameters (τ, ξ). Write:

$$\begin{aligned}\nabla_y p(s,x,t,y) &= \nabla_y \widetilde{p}^{(\tau,\xi)}(s,x,t,y) + \int_s^t \int_{\mathbb{R}^d} p(s,x,r,z) \mathcal{A}_{r,z}^{(\tau,\xi)} \nabla_y \widetilde{p}^{(\tau,\xi)}(r,z,t,y) \mathrm{d}z \mathrm{d}r \\ &= -\nabla_x \widetilde{p}^{(\tau,\xi)}(s,x,t,y) - \int_s^t \int_{\mathbb{R}^d} p(s,x,r,z) \mathcal{A}_{r,z}^{(\tau,\xi)} \nabla_z \widetilde{p}^{(\tau,\xi)}(r,z,t,y) \mathrm{d}z \mathrm{d}r, \qquad (2.31)\end{aligned}$$

where we have used the explicit expression (2.8) for the second equality. Still letting $u = \frac{s+t}{2}$ and taking $(\tau, \xi) = (t, y)$, we can split

$$\nabla_y p(s,x,t,y) = -\nabla_x \widetilde{p}_1(s,x,t,y) - J_1(y) - J_2(y),$$

where

$$\begin{aligned}J_1(y) &:= \int_s^u \int_{\mathbb{R}^d} p(s,x,r,z) \mathcal{A}_{r,z}^{(t,y)} \nabla_z \widetilde{p}_1(r,z,t,y) \mathrm{d}z \mathrm{d}r, \\ J_2(y) &:= \int_u^t \int_{\mathbb{R}^d} p(s,x,r,z) \mathcal{A}_{r,z}^{(t,y)} \nabla_z \widetilde{p}_1(r,z,t,y) \mathrm{d}z \mathrm{d}r.\end{aligned}$$

For $J_1(y)$, from the Gaussian upper-bound of Theorem 2.11, (2.23) and Lemma 2.5 (see also Lemma 2.6), we have

$$|J_1(y)| \lesssim \int_s^u (t-r)^{-3/2} \int_{\mathbb{R}^d} \mathbf{g}_{\lambda_0}(s,x,r,z) \mathbf{g}_{\lambda_5}(r,z,t,y) \mathrm{d}z \mathrm{d}r \lesssim (t-s)^{-1/2} \mathbf{g}_{\lambda_6}(s,x,t,y).$$

For J_2, integrating by parts and recalling (2.14) and (2.15), we have

$$\begin{aligned}|J_2(y)| \leqslant {} & \int_u^t \left| \int_{\mathbb{R}^d} \nabla_z p(s,x,r,z) \mathcal{A}_{r,z}^{(t,y)} \widetilde{p}_1(r,z,t,y) \mathrm{d}z \right| \mathrm{d}r \\ & + \int_u^t \left| \int_{\mathbb{R}^d} p(s,x,r,z) \nabla_z b \cdot \nabla_z \widetilde{p}_1(r,z,t,y) \mathrm{d}z \right| \mathrm{d}r \\ & + \int_u^t \left| \int_{\mathbb{R}^d} p(s,x,r,z) \nabla_z a \cdot \nabla_z^2 \widetilde{p}_1(r,z,t,y) \mathrm{d}z \right| \mathrm{d}r \\ =: {} & J_{21}(y) + J_{22}(y) + J_{23}(y).\end{aligned}$$

For $J_{21}(y)$, recalling from (2.18) that $\mathcal{A}^{(t,y)}_{r,z}\widetilde{p}_1(r,z,t,y) = H(r,z,t,y)$, we derive from (2.4) and as in (2.30) that

$$J_{21}(y) = \int_u^t \left| \int_{\mathbb{R}^d} p(s,x,r,z)\nabla_z H(r,z,t,y)\mathrm{d}z \right| \mathrm{d}r = \int_u^t |\mathbb{E}(\nabla_z H)(r, X_{r,s}(x), t, y)|\,\mathrm{d}r$$
$$\lesssim \int_u^t (r-s)^{-1/2}(\mathbb{E}|H(r, X_{r,s}(x), t, y)|^p)^{1/p}\mathrm{d}r \lesssim (t-s)^{-1/2}\mathbf{g}_{\lambda_7}(s,x,t,y).$$

For $J_{22}(y)$, from the upper bound in Theorem 2.11 and (2.23), we have

$$J_{22}(y) \lesssim \int_u^t (t-r)^{-1/2} \int_{\mathbb{R}^d} \mathbf{g}_{\lambda_0}(s,x,r,z)\mathbf{g}_{\lambda_1}(r,z,t,y)\mathrm{d}z\mathrm{d}r \lesssim (t-s)^{-1/2}\mathbf{g}_{\lambda_8}(s,x,t,y).$$

For J_{23}, since $|\nabla_z^2 \widetilde{p}_1(r,z,t,y)|$ has the singularity $(t-r)^{-1}$, noting that

$$\nabla_z a \cdot \nabla_z^2 \widetilde{p}_1 = \nabla_z^2(\nabla_z a \cdot \widetilde{p}_1) - \nabla_z^3 a \cdot \widetilde{p}_1 - \nabla_z^2 a \cdot \nabla_z \widetilde{p}_1,$$

as above, by (2.4) we still have

$$J_{23}(y) \lesssim (t-s)^{-1/2}\mathbf{g}_{\lambda_9}(s,x,t,y).$$

Combining the above estimates, we obtain the derivative estimate in y. The proof is complete. □

Remark 2.13. *We point out that Theorem 2.12 anyhow has some interest by itself. A careful reading of the proof shows that actually the statements about the derivatives w.r.t. x hold true if additionally to* $(\mathbf{H}^\sigma_\alpha)$, $(\mathbf{H}^b_\beta)$, *the coefficients b, σ are twice continuously differentiable with bounded derivatives and that the second order derivatives are themselves Hölder continuous. In this framework, the Duhamel representation* (2.20) *coupled to the heat-kernel estimates of Theorem 2.11 provides an alternative approach to the full Malliavin calculus viewpoint developed in [17].*

3 Proof of the main theorem

In the following proof, the final time horizon $T > 0$ is fixed. We first work under the assumptions **(S)** aiming at obtaining constants in the estimates of Section 2.3 that only depend on $\Theta := (T, \alpha, \beta, \kappa_0, \kappa_1, d)$ introduced in (1.14). To this end, we introduce for $\delta > 0$ the SDE (2.3) with diffusion coefficient $\sigma(t,x) = \delta \mathbb{I}_{d\times d}$ and denote by $\bar{p}_\delta$ the corresponding heat kernel. By the lower bound estimate proven in Theorem 2.11 and scaling techniques similar to those presented in Lemma 2.9, it holds that for any $\lambda > 0$, there exists $\delta := \delta(\lambda)$ large enough and $\bar{C}_\delta > 0, \lambda'$ depending on $\bar{\Theta} = (T, \beta, \delta, \kappa_1, d)$ such that for all $(s,t) \in \mathbb{D}_0^T$ and $x, y \in \mathbb{R}^d$,

$$\bar{C}_\delta^{-1}\mathbf{g}_\lambda(s,x,t,y) \leqslant \bar{p}_\delta(s,x,t,y) \leqslant \bar{C}_\delta \mathbf{g}_{\lambda'}(s,x,t,y). \tag{3.1}$$

We carefully mention that the Chapman-Kolmogorov equation satisfied by $\bar{p}_\delta$ plays a key role in the following proof when we use a Gronwall type argument. This important property had already been successfully used in [22] to derive Aronson type estimates for some degenerate SPDEs. We can as well refer to [23] for other applications of the parametrix method to non degenerate parabolic SPDEs.

Importantly, with the notations of Section 2.1, we will here choose λ, and then $\delta := \delta(\lambda)$ s.t. for all $\gamma \in [0,1]$, $(s,t) \in \mathbb{D}_0^T,\ x, y \in \mathbb{R}^d$ and $j \in \{0,1,2\}$,

$$|\theta_{t,s}(x) - y|^\gamma |\nabla_y^j \widetilde{p}_0(s,x,t,y)| + |x - \theta_{s,t}(y)|^\gamma |\nabla_x^j \widetilde{p}_1(s,x,t,y)| \leqslant C_\delta (t-s)^{\frac{\gamma}{2} - \frac{j}{2}} \bar{p}_\delta(s,x,t,y), \tag{3.2}$$

where C_δ here only depends on $\Theta = (T, \alpha, \beta, \kappa_0, \kappa_1, d)$ and δ, γ.

Without further declaration, we shall fix from now on a δ such that (3.2) holds. From the definition of H in (2.18) and the proof of Lemma 2.6, we also derive from this choice of δ that, under the sole assumptions $(\mathbf{H}^\sigma_\alpha)$ and $(\mathbf{H}^b_\beta)$, there exists $C := C(\Theta)$ s.t. for all $(s,t) \in \mathbb{D}_0^T,\ x, y \in \mathbb{R}^d$:

$$|H(s,x,t,y)| \leqslant C(t-s)^{-1+\frac{\alpha}{2}} \bar{p}_\delta(s,x,t,y). \tag{3.3}$$

For simplicity we will write from now on $\bar{p} = \bar{p}_\delta$. In particular, for all $(s,t) \in \mathbb{D}_0^T,\ x, y \in \mathbb{R}^d,\ r \in [s,t]$:

$$\int_{\mathbb{R}^d} \bar{p}(s,x,r,z)\bar{p}(r,z,t,y)\mathrm{d}z = \bar{p}(s,x,t,y). \tag{CK}$$

For the rest of section, we use the convention that all the constants appearing below only depend on $\Theta = (T, \alpha, \beta, \kappa_0, \kappa_1, d)$. Again, we have shown in the previous section that for smooth coefficients the expected bounds for the derivatives hold. The constants in Theorem 2.12 however do depend on the derivatives of the coefficients, since we use the gradient estimate (2.4). We aim here at proving that we can obtain the same type of estimates as in Theorem 2.12 under $(\mathbf{H}^{\sigma}_{\alpha})$, $(\mathbf{H}^{b}_{\beta})$ and (2.1) but for constants that only depend on Θ. This is the purpose of Sections 3.1 to 3.3. We will then eventually derive in Section 3.4 the main results of Theorem 1.2 thanks to some compactness arguments (Ascoli-Arzelà theorem) from the uniformity of the controls obtained for mollified parameters.

3.1 First order derivative estimates with respect to the backward variable x

Without loss of generality we shall assume $s = 0$ and for $t \in (0, T]$, we define

$$f_1(t) := \sup_{x,y} |\nabla_x p(0, x, t, y)|/\bar{p}(0, x, t, y).$$

From Theorem 2.12 and (3.1), we know that

$$\int_0^T f_1(t)\mathrm{d}t < \infty.$$

By the forward representation formula (2.20), we have

$$|\nabla_x p(0, x, t, y)| \leqslant |\nabla_x \widetilde{p}_1(0, x, t, y)| + |\nabla_x p| \otimes |H|(0, x, t, y).$$

Observe first that, from Lemma 2.4 and (3.2)

$$|\nabla_x \widetilde{p}_1(0, x, t, y)| \lesssim t^{-1/2}\mathbf{g}_\lambda(0, x, t, y) \lesssim t^{-1/2}\bar{p}(0, x, t, y).$$

Secondly, (3.3) yields

$$\begin{aligned}
|\nabla_x p| \otimes |H|(0, x, t, y) &\leqslant \int_0^t \int_{\mathbb{R}^d} f_1(r)\bar{p}(0, x, r, z)|H(r, z, t, y)|\mathrm{d}z\mathrm{d}r \\
&\lesssim \int_0^t f_1(r)(t-r)^{-1+\frac{\alpha}{2}} \int_{\mathbb{R}^d} \bar{p}(0, x, r, z)\bar{p}(r, z, t, y)\mathrm{d}z\mathrm{d}r \\
&= \left(\int_0^t f_1(r)(t-r)^{-1+\frac{\alpha}{2}}\mathrm{d}r\right)\bar{p}(0, x, t, y),
\end{aligned}$$

using also (**CK**) for the last identity. Thus,

$$f_1(t) \lesssim t^{-\frac{1}{2}} + \int_0^t (t-r)^{-1+\frac{\alpha}{2}} f_1(r)\mathrm{d}r.$$

By the Volterra type Gronwall inequality, we obtain

$$f_1(t) \lesssim t^{-\frac{1}{2}} \Rightarrow |\nabla_x p(0, x, t, y)| \lesssim t^{-\frac{1}{2}}\bar{p}(0, x, t, y). \tag{3.4}$$

3.2 Second order derivative estimates with respect to the backward variable x

We assume for this section that $(\mathbf{H}^{b}_{\beta})$ holds for some $\beta \in (0, 1]$. It is crucial to take here $\beta > 0$. Below we fix $t \in (0, T]$ and define for $s < t$

$$f_2(s) := (t-s) \cdot \sup_{x,y} |\nabla^2_x p(s, x, t, y)|/\bar{p}(s, x, t, y). \tag{3.5}$$

By Theorem 2.12 and (3.1), we have

$$\sup_{s \leqslant t} f_2(s) < \infty.$$

To derive the estimate of the second order derivative of the heat kernel, we use the backward Duhamel representation (2.12). And for fixed freezing parameters (τ, ξ) we differentiate twice w.r.t. x to derive:

$$\begin{aligned}
\nabla^2_x p(s, x, t, y) &= \nabla^2_x \widetilde{p}^{(\tau,\xi)}(s, x, t, y) + \int_s^t \int_{\mathbb{R}^d} \nabla^2_x \widetilde{p}^{(\tau,\xi)}(s, x, r, z)\mathcal{A}^{(\tau,\xi)}_{r,z} p(r, z, t, y)\mathrm{d}z\mathrm{d}r \\
&= \nabla^2_y \widetilde{p}^{(\tau,\xi)}(s, x, t, y) + \int_s^t \int_{\mathbb{R}^d} \nabla^2_z \widetilde{p}^{(\tau,\xi)}(s, x, r, z)\mathcal{A}^{(\tau,\xi)}_{r,z} p(r, z, t, y)\mathrm{d}z\mathrm{d}r,
\end{aligned} \tag{3.6}$$

using again the explicit expression (2.8) for the second equality. Let us now denote for a parameter $\varepsilon > 0$ that might depend on r to be specified later on,

$$A^{\varepsilon,(\tau,\xi)}_{r,z} := a_\varepsilon(r,z) - a_\varepsilon(r,\theta_{r,\tau}(\xi)), \quad \bar{A}^{(\tau,\xi)}_{r,z} := A^{(\tau,\xi)}_{r,z} - A^{\varepsilon,(\tau,\xi)}_{r,z}, \tag{3.7}$$

where similarly to (1.8), $a_\varepsilon(r,z) = a(r,\cdot) * \rho_\varepsilon(z)$. Choosing the freezing point $(\tau,\xi) = (s,x)$ and setting as well

$$\widetilde{p}_0(s,x,t,y) = \widetilde{p}^{(s,x)}(s,x,t,y), \quad u := (t+s)/2,$$

we decompose the expression in (3.6) as follows:

$$\nabla^2_x p(s,x,t,y) =: \sum_{i=1}^5 I_i(s,x,t,y), \tag{3.8}$$

where $I_1(s,x,t,y) := \nabla^2_y \widetilde{p}_0(s,x,t,y)$ and

$$\begin{aligned}
I_2(s,x,t,y) &:= \int_s^u \int_{\mathbb{R}^d} \nabla^2_z \widetilde{p}_0(s,x,r,z) \mathrm{tr}(A^{(s,x)}_{r,z} \cdot \nabla^2_z p(r,z,t,y)) \mathrm{d}z\mathrm{d}r \\
I_3(s,x,t,y) &:= \int_u^t \int_{\mathbb{R}^d} \nabla^2_z \widetilde{p}_0(s,x,r,z) \mathrm{tr}(A^{\varepsilon,(s,x)}_{r,z} \cdot \nabla^2_z p(r,z,t,y)) \mathrm{d}z\mathrm{d}r \\
I_4(s,x,t,y) &:= \int_u^t \int_{\mathbb{R}^d} \nabla^2_z \widetilde{p}_0(s,x,r,z) \mathrm{tr}(\bar{A}^{\varepsilon,(s,x)}_{r,z} \cdot \nabla^2_z p(r,z,t,y)) \mathrm{d}z\mathrm{d}r \\
I_5(s,x,t,y) &:= \int_s^t \int_{\mathbb{R}^d} \nabla^2_z \widetilde{p}_0(s,x,r,z) B^{(s,x)}_{r,z} \cdot \nabla_z p(r,z,t,y) \mathrm{d}z\mathrm{d}r.
\end{aligned}$$

By Lemma 2.3, (2.16) and (3.2), it is easy to see that

$$|I_1(s,x,t,y)| \lesssim (t-s)^{-1} g_\lambda(t-s, \theta_{t,s}(x) - y) \lesssim (t-s)^{-1} \bar{p}(s,x,t,y).$$

For I_2, by (1.6) and again (3.2), we have

$$\begin{aligned}
|I_2(s,x,t,y)| &\lesssim \int_s^u \int_{\mathbb{R}^d} \frac{g_\lambda(r-s, \theta_{r,s}(x) - z)}{r-s} |z - \theta_{r,s}(x)|^\alpha |\nabla^2_x p|(r,z,t,y)) \mathrm{d}z\mathrm{d}r \\
&\lesssim \int_s^u \frac{(r-s)^{\alpha/2} f_2(r)}{(r-s)(t-r)} \int_{\mathbb{R}^d} \bar{p}(s,x,r,z) \bar{p}(r,z,t,y) \mathrm{d}z\mathrm{d}r \\
&\leqslant (t-s)^{-1} \bar{p}(s,x,t,y) \int_s^t (r-s)^{-1+\frac{\alpha}{2}} f_2(r) \mathrm{d}r.
\end{aligned}$$

For I_3, integrating by parts, we have

$$\begin{aligned}
|I_3(s,x,t,y)| \lesssim& \int_u^t \int_{\mathbb{R}^d} \left| \nabla^3_z \widetilde{p}_0(s,x,r,z) \right| \cdot |A^{\varepsilon,(s,x)}_{r,z}| \cdot |\nabla_z p(r,z,t,y)| \mathrm{d}z\mathrm{d}r \\
&+ \int_u^t \int_{\mathbb{R}^d} \left| \nabla^2_z \widetilde{p}_0(s,x,r,z) \right| \cdot |\nabla_z A^{\varepsilon,(s,x)}_{r,z}| \cdot |\nabla_z p(r,z,t,y)| \mathrm{d}z\mathrm{d}r.
\end{aligned}$$

Note that by the property of convolutions,

$$|\nabla_z A^{\varepsilon,(s,x)}_{r,z}| \lesssim \varepsilon^{-1+\alpha}, \quad |A^{\varepsilon,(s,x)}_{r,z}| \lesssim |z - \theta_{r,s}(x)|^\alpha, \quad |\bar{A}^{\varepsilon,(s,x)}_{r,z}| \lesssim \varepsilon^\alpha.$$

In particular, taking $\varepsilon = (t-r)^{1/2}$, by Lemma 2.3, (2.16), (3.2) and using as well the bound (3.4) on the gradient established in the previous section, we obtain

$$\begin{aligned}
|I_3(s,x,t,y)| \lesssim& \int_u^t \int_{\mathbb{R}^d} \frac{\bar{p}(s,x,r,z)}{(r-s)^{\frac{3}{2}}} \cdot (r-s)^{\frac{\alpha}{2}} \cdot \frac{\bar{p}(r,z,t,y)}{(t-r)^{\frac{1}{2}}} \mathrm{d}z\mathrm{d}r \\
&+ \int_u^t \int_{\mathbb{R}^d} \frac{\bar{p}(s,x,r,z)}{r-s} \cdot (t-r)^{\frac{\alpha}{2}} \cdot \frac{\bar{p}(r,z,t,y)}{t-r} \mathrm{d}z\mathrm{d}r \\
\lesssim& \bar{p}(s,x,t,y) \int_u^t (r-s)^{-\frac{3-\alpha}{2}} (t-r)^{-\frac{1}{2}} \mathrm{d}r \\
&+ \bar{p}(s,x,t,y) \int_u^t (r-s)^{-1} (t-r)^{-1+\frac{\alpha}{2}} \mathrm{d}r \\
\lesssim& \bar{p}(s,x,t,y)(t-s)^{-1+\frac{\alpha}{2}},
\end{aligned}$$

and

$$|I_4(s,x,t,y)| \lesssim \int_u^t \int_{\mathbb{R}^d} \frac{\bar{p}(s,x,r,z)}{r-s} \cdot (t-r)^{\frac{\alpha}{2}} \cdot \frac{f_2(r)\bar{p}(r,z,t,y)}{t-r} \mathrm{d}z\mathrm{d}r$$
$$\lesssim \bar{p}(s,x,t,y)(t-s)^{-1} \int_u^t (t-r)^{-1+\frac{\alpha}{2}} f_2(r)\mathrm{d}r.$$

For I_5, from (3.2), we derive similarly to I_2 that

$$|I_5(s,x,t,y)| \lesssim \int_s^t \int_{\mathbb{R}^d} \frac{g_\lambda(r-s,\theta_{r,s}(x)-z)}{r-s} (|z-\theta_{r,s}(x)|^\beta + |z-\theta_{r,s}(x)|) \frac{\bar{p}(r,z,t,y)}{(t-r)^{\frac{1}{2}}} \mathrm{d}z\mathrm{d}r$$
$$\lesssim \bar{p}(s,x,t,y) \int_s^t \frac{(r-s)^{\frac{\beta}{2}} + (r-s)^{\frac{1}{2}}}{(r-s)(t-r)^{\frac{1}{2}}} \mathrm{d}r \lesssim \bar{p}(s,x,t,y)(t-s)^{-1}.$$

Combining the above estimates for the $(I_j)_{j\in\{1,\cdots,5\}}$, we obtain from (3.8) and (3.5) that:

$$f_2(s) \lesssim 1 + \int_s^t (r-s)^{-1+\frac{\alpha}{2}} f_2(r)\mathrm{d}r + \int_s^t (t-r)^{-1+\frac{\alpha}{2}} f_2(r)\mathrm{d}r.$$

Finally, from the Volterra type Gronwall inequality, we obtain

$$\sup_{s\in[0,t]} f_2(s) \lesssim 1 \Rightarrow |\nabla_x^2 p(s,x,t,y)| \lesssim (t-s)^{-1}\bar{p}(s,x,t,y). \tag{3.9}$$

3.3 First order derivative estimate in y

We assume for this section that $(\mathbf{H}_\beta^b)$ holds for some $\beta > 0$ and the gradient of the diffusion coefficient σ satisfies (1.18). Fix $s > 0$. For $t \in (s,T]$, we define

$$f_3(t) := \sup_{x,y} |\nabla_y p(s,x,t,y)|/\bar{p}(s,x,t,y). \tag{3.10}$$

By Theorem 2.12 and (3.1) we know that

$$\int_s^T f_3(t)\mathrm{d}t < \infty.$$

In (2.31), taking $(\tau,\xi) = (t,y)$ and recalling the notations of (2.15) and $\widetilde{p}_1(s,x,t,y) = \widetilde{p}^{(t,y)}(s,x,t,y)$, by the integration by parts, we have

$$\nabla_y p(s,x,t,y) = -\nabla_x \widetilde{p}_1(s,x,t,y) + \int_s^t \int_{\mathbb{R}^d} \nabla_z p(s,x,r,z)\mathrm{tr}(A_{r,z}^{(t,y)} \cdot \nabla_z^2 \widetilde{p}_1)(r,z,t,y)\mathrm{d}z\mathrm{d}r$$
$$+ \int_s^t \int_{\mathbb{R}^d} p(s,x,r,z)\mathrm{tr}((\nabla_z a)(r,z) \cdot \nabla_z^2 \widetilde{p}_1)(r,z,t,y)\mathrm{d}z\mathrm{d}r$$
$$- \int_s^t \int_{\mathbb{R}^d} p(s,x,r,z) B_{r,z}^{(t,y)} \cdot \nabla_z^2 \widetilde{p}_1(r,z,t,y)\mathrm{d}z\mathrm{d}r =: \sum_{i=1}^4 J_i(s,x,t,y). \tag{3.11}$$

For J_1, we readily get from (3.2)

$$|J_1(s,x,t,y)| \lesssim (t-s)^{-\frac{1}{2}} \bar{p}(s,x,t,y).$$

For J_2, using again (3.2) and (3.10) gives:

$$|J_2(s,x,t,y)| \lesssim \int_s^t \int_{\mathbb{R}^d} |\nabla_z p(s,x,r,z)| \cdot (t-r)^{-1+\frac{\alpha}{2}} \bar{p}(r,z,t,y)\mathrm{d}z\mathrm{d}r$$
$$\lesssim \int_s^t f_3(r) \int_{\mathbb{R}^d} \bar{p}(s,x,r,z) \cdot (t-r)^{-1+\frac{\alpha}{2}} \bar{p}(r,z,t,y)\mathrm{d}z\mathrm{d}r$$
$$\lesssim \bar{p}(s,x,t,y) \int_s^t f_3(r)(t-r)^{-1+\frac{\alpha}{2}} \mathrm{d}r.$$

For J_3, we further write

$$\begin{aligned}J_3(s,x,t,y) &= \int_s^t \int_{\mathbb{R}^d} p(s,x,r,z)\mathrm{tr}\Big(((\nabla_z a)(r,z) - (\nabla_z a)(r,\theta_{r,t}(y)))\cdot \nabla_z^2 \widetilde{p}_1\Big)(r,z,t,y)\mathrm{d}z\mathrm{d}r \\ &\quad + \int_s^t \int_{\mathbb{R}^d} p(s,x,r,z)\mathrm{tr}\Big((\nabla_z a)(r,\theta_{r,t}(y))\cdot \nabla_z^2 \widetilde{p}_1\Big)(r,z,t,y)\mathrm{d}z\mathrm{d}r \\ &=: J_{31}(s,x,t,y) + J_{32}(s,x,t,y).\end{aligned}$$

For J_{31}, as above, by (3.2) we have

$$|J_{31}(s,x,t,y)| \lesssim \int_s^t \int_{\mathbb{R}^d} \bar{p}(s,x,r,z)\cdot (t-r)^{\frac{\alpha}{2}-1}\bar{p}(r,z,t,y)\mathrm{d}z\mathrm{d}r \lesssim \bar{p}(s,x,t,y).$$

For J_{32}, by the integration by parts again, we derive

$$\begin{aligned}|J_{32}(s,x,t,y)| &\lesssim \int_s^t \int_{\mathbb{R}^d} |\nabla_z p(s,x,r,z)|\cdot|\nabla_z \widetilde{p}_1(r,z,t,y)|\mathrm{d}z\mathrm{d}r \\ &\lesssim \int_s^t f_3(r)\int_{\mathbb{R}^d} \bar{p}(s,x,r,z)\cdot(t-r)^{-\frac{1}{2}}\bar{p}(r,z,t,y)\mathrm{d}z\mathrm{d}r \\ &\lesssim \bar{p}(s,x,t,y)\int_s^t f_3(r)(t-r)^{-\frac{1}{2}}\mathrm{d}r.\end{aligned}$$

For J_4, we derive similarly to the term J_{31} that

$$|J_4(s,x,t,y)| \lesssim \bar{p}(s,x,t,y).$$

Combining the above estimates for the $(J_i)_{i\in\{1,\cdots,4\}}$, we obtain from (3.11) and (3.10) that

$$f_3(t) \lesssim (t-s)^{-\frac{1}{2}} + \int_s^t f_3(r)(t-r)^{-1+\frac{\alpha}{2}}\mathrm{d}r,$$

which in turn yields

$$f_3(t) \lesssim (t-s)^{-\frac{1}{2}} \Rightarrow |\nabla_y p(s,x,t,y)| \lesssim (t-s)^{-\frac{1}{2}}\bar{p}(s,x,t,y). \tag{3.12}$$

3.4 Proof of Theorem 1.2

Now we turn to the notations of the beginning of Section 2 and keep the index ε, associated with the spatial mollification of the coefficients. Thus, let p_ε be the corresponding heat kernel and $X^\varepsilon_{t,s}(x)$ the solution of SDE (2.3) corresponding to b_ε and σ_ε. It is well known, see e.g. Theorem 11.1.4 in [26], that under $(\mathbf{H}^b_\beta)$ and $(\mathbf{H}^\sigma_\alpha)$, for any $f \in C^\infty_c(\mathbb{R}^d)$

$$\lim_{\varepsilon\to 0}\mathbb{E}f(X^\varepsilon_{t,s}(x)) = \mathbb{E}f(X_{t,s}(x)).$$

Moreover, from Theorem 2.11 we have the following uniform estimate: there exist constants $\lambda_0, C_0 > 0$ depending only on Θ such that for all $\varepsilon \in (0,1)$,

$$C_0^{-1}\mathbf{g}_{\lambda_0^{-1}}(s,x,t,y) \leqslant p_\varepsilon(s,x,t,y) \leqslant C_0\mathbf{g}_{\lambda_0}(s,x,t,y).$$

Similarly, we derive from (3.4), (3.9) and (3.12) that under $(\mathbf{H}^\sigma_\alpha)$ and $(\mathbf{H}^b_0)$,

$$\sup_\varepsilon |\nabla_x p_\varepsilon(s,x,t,y)| \leqslant C_1(t-s)^{-1/2}\mathbf{g}_{\lambda_1}(s,x,t,y), \tag{3.13}$$

and under $(\mathbf{H}^\sigma_\alpha)$ and $(\mathbf{H}^b_\beta)$ with $\beta \in (0,1)$, $j \in \{1,2\}$,

$$\sup_\varepsilon |\nabla^j_x p_\varepsilon(s,x,t,y)| \leqslant C_2(t-s)^{-j/2}\mathbf{g}_{\lambda_2}(s,x,t,y), \tag{3.14}$$

and under $(\mathbf{H}^\sigma_\alpha)$, $(\mathbf{H}^b_\beta)$ with $\beta \in (0,1)$ and (1.18),

$$\sup_\varepsilon |\nabla_y p_\varepsilon(s,x,t,y)| \leqslant C'_1(t-s)^{-1/2}\mathbf{g}_{\lambda'_1}(s,x,t,y), \tag{3.15}$$

where in the above equations (3.13)-(3.15) the constants C_1, C_2, C_1' only depend on Θ and *not* on the mollification parameter ε.

In particular, for nonnegative measurable function f, we eventually derive

$$C_0^{-1}\int_{\mathbb{R}^d}\mathbf{g}_{\lambda_0^{-1}}(s,x,t,y)f(y)\mathrm{d}y \leqslant \mathbb{E}f(X_{t,s}(x)) \leqslant C_0\int_{\mathbb{R}^d}\mathbf{g}_{\lambda_0}(s,x,t,y)f(y)\mathrm{d}y,$$

which implies that $X_{t,s}(x)$ has a density $p(s,x,t,y)$ having lower and upper bound as in (1.15). This proves point *(i)* of the theorem.

Moreover, for each $s<t$, we now aim at proving that

$$(x,y)\mapsto \nabla_x p_\varepsilon(s,x,t,y) \text{ is equi-continuous on any compact subset of } \mathbb{R}^d\times\mathbb{R}^d, \tag{$\mathbf{C}_1$}$$

and

$$(x,y)\mapsto \nabla_x^2 p_\varepsilon(s,x,t,y) \text{ is equi-continuous on any compact subset of } \mathbb{R}^d\times\mathbb{R}^d, \tag{$\mathbf{C}_2$}$$

$$(x,y)\mapsto \nabla_y p_\varepsilon(s,x,t,y) \text{ is equi-continuous on any compact subset of } \mathbb{R}^d\times\mathbb{R}^d. \tag{$\mathbf{C}_3$}$$

Assume for a while that such a continuity condition holds. Then, from the Ascoli-Arzelà theorem, one can find a subsequence ε_k such that for each $x,y\in\mathbb{R}^d$,

$$\nabla_x^j p_{\varepsilon_k}(s,x,t,y)\to\nabla_x^j p(s,x,t,y),\quad j=0,1,2,\ \nabla_y p_{\varepsilon_k}(s,x,t,y)\to\nabla_y p(s,x,t,y).$$

The gradient and second order derivative estimates follow, under the previously recalled additional assumptions when needed, from (3.13), (3.14) and (3.15). This completes the proof of points *(ii)* to *(iv)* of the theorem up to the proof of $(\mathbf{C}_1)$, $(\mathbf{C}_2)$ and $(\mathbf{C}_3)$. This equicontinuity property is proved in Appendix A. □

4 Extension to higher order derivatives

We explain here how the estimates (1.16), (1.17), (1.19) can be extended for higher order derivatives in our analysis. We claim that under **(S)** the *a-priori* bounds of Theorem 2.12 can be obtained for any $j\in\mathbb{N}$, using the same techniques based on the Duhamel representation of the density and (2.4). On the other hand the circular arguments used in Section 3 can be repeated as well, provided that the coefficients are smooth enough.

For instance, let us assume **(S)** to be in force; assume as well that $\|\nabla\sigma\|_\infty+\|\nabla b\|_\infty<\infty$ and for some $\alpha,\beta\in(0,1]$, $\kappa_2\geqslant 1$,

$$|\nabla\sigma(t,x)-\nabla\sigma(t,y)|\leqslant\kappa_2|x-y|^\alpha,\quad |\nabla b(t,x)-\nabla b(t,y)|\leqslant\kappa_2|x-y|^\beta,\quad x,y\in\mathbb{R}^d. \tag{4.1}$$

We aim here at proving that we can obtain bounds on the third order derivatives which only depend on $(\mathbf{H}_1^\sigma)$, $(\mathbf{H}_1^b)$ and the constants in (4.1). Namely, we want to illustrate a kind of *parabolic bootstrap property*, i.e. in (4.1) we give some Hölder conditions on the first derivatives of the coefficients which together with the assumptions $(\mathbf{H}_1^\sigma)$, $(\mathbf{H}_1^b)$ lead to a uniform control of the third order derivatives.

As in (3.6), for the choice of the freezing parameters $(\tau,\xi)=(s,x)$ and recalling $\widetilde{p}_0(s,x,t,y)=\widetilde{p}^{(s,x)}(s,x,t,y)$, we have the following representation for the derivatives of order three:

$$\nabla_x^3 p(s,x,t,y)=-\nabla_y^3\widetilde{p}_0(s,x,t,y)-\int_s^t\int_{\mathbb{R}^d}\nabla_z^3\widetilde{p}_0(s,x,r,z)\mathcal{A}_{r,z}^{(s,x)}p(r,z,t,y)\mathrm{d}z\mathrm{d}r. \tag{4.2}$$

Let us now concentrate on the most singular term in (4.2). Setting $u=(t+s)/2$ and $A_{r,z}^{(\tau,\xi)}$, $A_{r,z}^{\varepsilon,(\tau,\xi)}$, $\bar{A}_{r,z}^{\varepsilon,(\tau,\xi)}$ as in (2.15), (3.7) we write

$$\begin{aligned}
&\int_s^t\int_{\mathbb{R}^d}\nabla_z^3\widetilde{p}_0(s,x,r,z)\mathrm{tr}\left(A_{r,z}^{(s,x)}\cdot\nabla_z^2 p(r,z,t,y)\right)\mathrm{d}z\mathrm{d}r\\
=&\int_u^t\int_{\mathbb{R}^d}\nabla_z^3\widetilde{p}_0(s,x,r,z)\mathrm{tr}\left((A_{r,z}^{\varepsilon,(s,x)}+\bar{A}_{r,z}^{\varepsilon,(s,x)})\cdot\nabla_z^2 p(r,z,t,y)\right)\mathrm{d}z\mathrm{d}r\\
&+\int_s^u\int_{\mathbb{R}^d}\nabla_z^3\widetilde{p}_0(s,x,r,z)\mathrm{tr}\left(A_{r,z}^{(s,x)}\cdot\nabla_z^2 p(r,z,t,y)\right)\mathrm{d}z\mathrm{d}r=:G_1(s,x,t,y)+G_2(s,x,t,y).
\end{aligned}$$

When $r\in[u,t]$, $(r-s)^{-\frac{3}{2}}\asymp(t-s)^{-\frac{3}{2}}$ is not singular. Therefore we may control G_1 similarly to the terms I_3 and I_4 appearing in Section 3.2, owing to the fact that the upper bound on $\nabla_z^2 p$ is already available at this point.

When $r \in [s,u]$, then $(r-s)^{-\frac{3}{2}}$ is indeed singular. Thus, to control G_2 the point is precisely to exploit the regularity of the coefficients and perform an integration by parts to balance the singularity. We write

$$\begin{aligned} G_2(s,x,t,y) = &-\int_s^u \int_{\mathbb{R}^d} \nabla_z^2 \widetilde{p}_0(s,x,r,z) \mathrm{tr}\left(\nabla_z A_{r,z}^{(s,x)} \cdot \nabla_z^2 p(r,z,t,y)\right) \mathrm{d}z\mathrm{d}r \\ &-\int_s^u \int_{\mathbb{R}^d} \nabla_z^2 \widetilde{p}_0(s,x,r,z) \mathrm{tr}\left(A_{r,z}^{(s,x)} \cdot \nabla_z^3 p(r,z,t,y)\right) \mathrm{d}z\mathrm{d}r, \end{aligned}$$

and define

$$f_3(s) := (t-s)^{\frac{3}{2}} \sup_{x,y} |\nabla_x^3 p(s,x,t,y)|/\bar{p}(s,x,t,y);$$

Then, exploiting the uniform bounds for the derivatives of order lower or equal than 2 obtained in Section 3, we eventually derive

$$f_3(t) \lesssim 1 + \int_s^u (r-s)^{-1+\frac{\alpha}{2}} f_3(s)\mathrm{d}s \Rightarrow \sup_{s\in[0,t]} f_3(s) \lesssim 1,$$

which yields the desired estimate for $\nabla_x^3 p$. In the same manner, starting from the Duhamel expansion (2.17), and assuming in addition that $\|\nabla^2\sigma\|_\infty < \infty$ and $|\nabla^2\sigma(t,x) - \nabla^2\sigma(t,y)| \leqslant \kappa_3|x-y|^\alpha$ for some $\alpha \in (0,1)$ we could derive

$$|\nabla_y^2 p(s,x,t,y)| \lesssim (t-s)^{-1}\bar{p}(s,x,t,y).$$

A careful reading of the proof suggests that the above arguments may be repeated for any derivative of order $j > 3$ in the backward variable x as soon as we have appropriate regularity assumptions on $\nabla^{j-2}\sigma$ and $\nabla^{j-2}b$. More precisely, assuming that

$$\|\nabla^{j'}\sigma\|_\infty + \|\nabla^{j'}b\|_\infty < \infty,\ j' = 1,\cdots,j-2,$$

and for some $\alpha,\beta \in (0,1]$, $\kappa_{j-2} \geqslant 1$,

$$|\nabla^{j-2}\sigma(t,x) - \nabla^{j-2}\sigma(t,y)| \leqslant \kappa_{j-2}|x-y|^\alpha, \quad |\nabla^{j-2}b(t,x) - \nabla^{j-2}b(t,y)| \leqslant \kappa_{j-2}|x-y|^\beta, \quad x,y \in \mathbb{R}^d,$$

then we may derive

$$|\nabla_x^j p(s,x,t,y)| \lesssim (t-s)^{-\frac{j}{2}}\bar{p}(s,x,t,y).$$

On the other hand, the derivative with respect to the forward variable ∇_y^{j-1} requires an additional assumption on $\nabla^{j-1}\sigma$. Again, assuming that for some $\alpha \in (0,1)$, $|\nabla^{j-1}\sigma(t,x) - \nabla^{j-1}\sigma(t,y)| \leqslant \kappa_{j-1}|x-y|^\alpha$ for any $x,y \in \mathbb{R}^d$, then we may derive

$$|\nabla_y^{j-1} p(s,x,t,y)| \lesssim (t-s)^{-\frac{j-1}{2}}\bar{p}(s,x,t,y).$$

A Proof of the equicontinuity $(\mathbf{C}_1)$, $(\mathbf{C}_2)$ and $(\mathbf{C}_3)$

Importantly, we mention that we drop in this section the subscripts and superscripts in ε for notational convenience. However, it must be recalled that we aim at proving some equicontinuity properties for the densities associated with the SDE (2.3) with mollified coefficients and their derivatives.

In this section we devote to proving the following Hölder continuity of the derivatives.

Lemma A.1. *Suppose that* $(\mathbf{H}_\alpha^\sigma)$ *and* $(\mathbf{H}_\beta^b)$ *hold. Let* $T > 0$ *and* $\gamma_1 \in (0,1)$ *and* $\gamma_2 \in (0,\alpha)$ *and* $\gamma_3 \in (0, \alpha\wedge\beta)$.

$(\mathbf{C}_1)$ *There exist constants* $C, \lambda > 0$ *depending only on* Θ, γ_1, γ_2 *such that for all* $(s,t) \in \mathbb{D}_0^T$ *and* $x, x', y, y' \in \mathbb{R}^d$,

$$\begin{aligned} |\nabla_x p(s,x,t,y) - \nabla_x p(s,x',t,y)| &\lesssim_C \frac{|x-x'|^{\gamma_1}}{(t-s)^{(1+\gamma_1)/2}}\Big(\mathbf{g}_\lambda(s,x,t,y) + \mathbf{g}_\lambda(s,x',t,y)\Big) \\ |\nabla_x p(s,x,t,y) - \nabla_x p(s,x,t,y')| &\lesssim_C \frac{|y-y'|^{\gamma_2}}{(t-s)^{(1+\gamma_2)/2}}\Big(\mathbf{g}_\lambda(s,x,t,y) + \mathbf{g}_\lambda(s,x,t,y')\Big). \end{aligned}$$

($\mathbf{C}_2$) *If $\beta \in (0,1]$, there exist constants $C, \lambda > 0$ depending only on Θ such that for all $(s,t) \in \mathbb{D}_0^T$ and $x, x', y \in \mathbb{R}^d$,*

$$\begin{aligned}|\nabla_x^2 p(s,x,t,y) - \nabla_x^2 p(s,x',t,y)| &\lesssim_C \left(\frac{|x-x'|}{(t-s)^{\frac{3}{2}}} + \frac{|x-x'|^\alpha + |x-x'|^\beta}{t-s}\right) \\ &\quad \times \Big(\mathbf{g}_\lambda(s,x,t,y) + \mathbf{g}_\lambda(s,x',t,y)\Big), \\ |\nabla_x^2 p(s,x,t,y) - \nabla_x^2 p(s,x,t,y')| &\lesssim_C \left(\frac{|y-y'|^{\gamma_2}}{(t-s)^{1+\frac{\gamma_2}{2}}} + \frac{|y-y'|^\alpha + |y-y'|^\beta}{t-s}\right) \\ &\quad \times \Big(\mathbf{g}_\lambda(s,x,t,y) + \mathbf{g}_\lambda(s,x',t,y)\Big).\end{aligned} \tag{A.1}$$

($\mathbf{C}_3$) *If σ also satisfies (1.18) and $\beta \in (0,1)$, then there exist constants $C, \lambda > 0$ depending only on Θ, γ_1, γ_3 such that for all $(s,t) \in \mathbb{D}_0^T$ and $x, x', y, y' \in \mathbb{R}^d$,*

$$\begin{aligned}|\nabla_y p(s,x,t,y) - \nabla_y p(s,x,t,y')| &\lesssim_C \frac{|y-y'|^{\gamma_3}}{(t-s)^{(1+\gamma_3)/2}}\Big(\mathbf{g}_\lambda(s,x,t,y) + \mathbf{g}_\lambda(s,x,t,y')\Big), \\ |\nabla_y p(s,x,t,y) - \nabla_y p(s,x',t,y)| &\lesssim_C \frac{|x-x'|^{\gamma_1}}{(t-s)^{(1+\gamma_1)/2}}\Big(\mathbf{g}_\lambda(s,x,t,y) + \mathbf{g}_\lambda(s,x,t,y')\Big).\end{aligned}$$

Proof. We only prove **($\mathbf{C}_2$)** and focus on the sensitivity w.r.t the variable x. The sensitivity w.r.t. the variable y could be established similarly. Also, the inequalities in conditions **($\mathbf{C}_1$)** and **($\mathbf{C}_3$)** could be shown more directly.

First of all, if $|x-x'|^2 > (t-s)/4$, then by (3.14), we clearly have

$$|\nabla_x^2 p(s,x,t,y) - \nabla_x^2 p(s,x',t,y)| \lesssim (t-s)^{-1}\Big(\mathbf{g}_\lambda(s,x,t,y) + \mathbf{g}_\lambda(s,x',t,y)\Big) \lesssim \text{r.h.s. of (A.1)}.$$

Next we restrict to the so-called *diagonal case*

$$|x-x'|^2 \leqslant (t-s)/4.$$

For any fixed freezing point (τ,ξ) and $r \in (s,t)$, by (2.12), one sees that

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

where, with the notations of (2.7),

$$\widetilde{P}_{s,r}^{(\tau,\xi)} f(x) = \int_{\mathbb{R}^d} \widetilde{p}^{(\tau,\xi)}(s,x,r,z) f(z)\mathrm{d}z.$$

Let us now differentiate w.r.t. r. We obtain for all $(\tau,\xi) \in [0,T] \times \mathbb{R}^d$:

$$0 = \partial_r[\widetilde{P}_{s,r}^{(\tau,\xi)} p(r,\cdot,t,y)(x)] + \int_{\mathbb{R}^d} \widetilde{p}^{(\tau,\xi)}(s,x,r,z)\mathcal{A}_{r,z}^{(\tau,\xi)} p(r,z,t,y)\mathrm{d}z. \tag{A.2}$$

Fix $s_0 \in (s,t)$. Now, integrating (A.2) between s_0 and t and taking $\xi = \xi'$, we get

$$0 = \widetilde{p}^{(\tau,\xi')}(s,x,t,y) - \widetilde{P}_{s,s_0}^{(\tau,\xi')} p(s_0,\cdot,t,y)(x) + \int_{s_0}^t \mathrm{d}r \int_{\mathbb{R}^d} \widetilde{p}^{(\tau,\xi')}(s,x,r,z)\mathcal{A}_{r,z}^{(\tau,\xi')} p(r,z,t,y)\mathrm{d}z.$$

Moreover, integrating (A.2) between s and s_0, we obtain

$$0 = \widetilde{P}_{s,s_0}^{(\tau,\xi)} p(s_0,\cdot,t,y)(x) - p(s,x,t,y) + \int_s^{s_0} \mathrm{d}r \int_{\mathbb{R}^d} \widetilde{p}^{(\tau,\xi)}(s,x,r,z)\mathcal{A}_{r,z}^{(\tau,\xi)} p(r,z,t,y)\mathrm{d}z;$$

Summing up the two equalities we get the following new representation for $p(s,x,t,y)$:

$$\begin{aligned}p(s,x,t,y) =& \widetilde{p}^{(\tau,\xi')}(s,x,t,y) + \Big(\widetilde{P}_{s,s_0}^{(\tau,\xi)} - \widetilde{P}_{s,s_0}^{(\tau,\xi')}\Big) p(s_0,\cdot,t,y)(x) \\ &+ \int_{s_0}^t \mathrm{d}r \int_{\mathbb{R}^d} \widetilde{p}^{(\tau,\xi')}(s,x,r,z)\mathcal{A}_{u,z}^{(\tau,\xi')} p(r,z,t,y)\mathrm{d}z \\ &+ \int_s^{s_0} \mathrm{d}r \int_{\mathbb{R}^d} \widetilde{p}^{(\tau,\xi)}(s,x,r,z)\mathcal{A}_{r,z}^{(\tau,\xi)} p_\varepsilon(r,z,t,y)\mathrm{d}z,\end{aligned}$$

which, together with (2.12) by taking $x = x'$ and $\xi = \xi'$ there, yields

$$\begin{aligned} p(s,x,t,y) - p(s,x',t,y) =& \widetilde{p}^{(\tau,\xi')}(s,x,t,y) - \widetilde{p}^{(\tau,\xi')}(s,x',t,y) + \Big(\widetilde{P}^{(\tau,\xi)}_{s,s_0} - \widetilde{P}^{(\tau,\xi')}_{s,s_0}\Big)p(s_0,\cdot,t,y)(x) \\ &+ \Delta^{\tau,\xi',\xi'}_{\mathrm{diag}}(s,t,x,x',y) + \Delta^{\tau,\xi,\xi'}_{\mathrm{off-diag}}(s,t,x,x',y), \end{aligned} \tag{A.3}$$

where

$$\Delta^{\tau,\xi',\xi'}_{\mathrm{diag}}(s,t,x,x',y) = \int_{s_0}^{t} \mathrm{d}r \int_{\mathbb{R}^d} \Big[\widetilde{p}^{(\tau,\xi')}(s,x,r,z) - \widetilde{p}^{(\tau,\xi')}(s,x',r,z)\Big]\mathcal{A}^{(\tau,\xi')}_{r,z} p(r,z,t,y)\mathrm{d}z$$

and

$$\Delta^{\tau,\xi,\xi'}_{\mathrm{off-diag}}(s,t,x,x',y) = \int_{s}^{s_0} \mathrm{d}r \int_{\mathbb{R}^d} \Big[\widetilde{p}^{(\tau,\xi)}(s,x,r,z)\mathcal{A}^{(\tau,\xi)}_{r,z} p(r,z,t,y) - \widetilde{p}^{(\tau,\xi')}(s,x',r,z)\mathcal{A}^{(\tau,\xi')}_{r,z} p(r,z,t,y)\Big]\mathrm{d}z.$$

Observe that for any freezing couple (τ,ξ) and $h \in \mathbb{R}^d$,

$$\nabla^2_x \widetilde{p}^{(\tau,\xi)}(s,x+h,t,y) = \nabla^2_y \widetilde{p}^{(\tau,\xi)}(s,x,t,y-h).$$

After differentiating twice in x for both sides of (A.3) and taking $\tau = s$ and $\xi = x,\ \xi' = x'$, we obtain

$$\nabla^2_x p(s,x,t,y) - \nabla^2_x p(s,x',t,y) = \sum_{i=1}^{4} I_i(s,t,x,x',y),$$

where, with the notation $\widetilde{p}_0(s,x,t,y) = \widetilde{p}^{(s,x)}(s,x,t,y)$,

$$\begin{aligned} I_1(s,t,x,x',y) &:= \nabla^2_y \widetilde{p}_0(s,x',t,y+x'-x) - \nabla^2_y \widetilde{p}_0(s,x',t,y), \\ I_2(s,t,x,x',y) &:= \int_{\mathbb{R}^d} \nabla^2_z (\widetilde{p}^{(s,x)}(s,x,s_0,z) - \widetilde{p}^{(s,x')}(s,x,s_0,z))p(s_0,z,t,y)\mathrm{d}z, \\ I_3(s,t,x,x',y) &:= \int_{s_0}^{t} \mathrm{d}r \int_{\mathbb{R}^d} \Big[\nabla^2_z \widetilde{p}_0(s,x',r,z+x'-x) - \nabla^2_z \widetilde{p}_0(s,x',r,z)\Big]\mathcal{A}^{(s,x')}_{r,z} p(r,z,t,y)\mathrm{d}z, \\ I_4(s,t,x,x',y) &:= \int_{s}^{s_0} \mathrm{d}r \int_{\mathbb{R}^d} \Big[\nabla^2_z \widetilde{p}_0(s,x,r,z)\mathcal{A}^{(s,x)}_{r,z} p(r,z,t,y) - \nabla^2_z \widetilde{p}_0(s,x',r,z)\mathcal{A}^{(s,x')}_{r,z} p(r,z,t,y)\Big]\mathrm{d}z. \end{aligned}$$

Note that by Lemma 2.3, for $j \in \mathbb{N}$ and $h \in \mathbb{R}^d$ with $|h|^2 \leqslant (t-s)/4$,

$$\begin{aligned} \Big|\nabla^j_y \widetilde{p}_0(s,x,t,y+h) - \nabla^j_y \widetilde{p}_0(s,x,t,y)\Big| &\leqslant |h| \sup_{\varrho\in[0,1]} \big|\nabla^{j+1}_y \widetilde{p}_0(s,x,t,y+\varrho h)\big| \\ &\lesssim |h|(t-s)^{-(j+1)/2} \sup_{\varrho\in[0,1]} g_\lambda(t-s,\theta_{t,s}(x)-(y+\varrho h)) \\ &\lesssim |h|(t-s)^{-(j+1)/2} g_\lambda(t-s,\theta^{(1)}_{t,s}(x)-y), \end{aligned} \tag{A.4}$$

using Lemma 1.1 for the last step. On the other hand, we also have

$$\Big|\nabla^j_y \widetilde{p}_0(s,x,t,y+h)\Big| \lesssim (t-s)^{-j/2} g_\lambda(t-s,\theta^{(1)}_{t,s}(x)-y).$$

Thus, by interpolation, we get for any $\gamma \in (0,1)$,

$$\Big|\nabla^j_y \widetilde{p}_0(s,x,t,y+h) - \nabla^j_y \widetilde{p}_0(s,x,t,y)\Big| \lesssim |h|^\gamma (t-s)^{-(j+\gamma)/2} g_\lambda(t-s,\theta^{(1)}_{t,s}(x)-y).$$

Hence,

$$|I_1(s,t,x,x',y)| \lesssim |x-x'|^\gamma (t-s)^{-1-\frac{\gamma}{2}} \mathbf{g}_\lambda(s,x',t,y).$$

To treat the remaining terms, we take $s_0 = s + |x-x'|^2$. We have the following claim:

$$\Big|\widetilde{p}^{(s,x)}(s,x,s_0,y) - \widetilde{p}^{(s,x')}(s,x,s_0,y)\Big| \lesssim \big(|x-x'|^\alpha + |x-x'|^\beta\big)\mathbf{g}_\lambda(s,x,s_0,y). \tag{A.5}$$

Indeed, by Lemma 1.1, there is a constant $C = C(\Theta)$ such that

$$|\theta_{r,s}(x) - \theta_{r,s}(x')| \lesssim_C |x - x'| + |r - s|, \quad x, x' \in \mathbb{R}^d,\ r \in [s, s_0].$$

Recalling $\vartheta^{(\tau,\xi)}_{s_0,s} = \int_s^{s_0} b(r, \theta_{r,\tau}(\xi))\mathrm{d}r$ from the notations of Section 2.1, we have:

$$\begin{aligned}|\vartheta^{(s,x)}_{s_0,s} - \vartheta^{(s,x')}_{s_0,s}| &\leqslant \int_s^{s_0} |b(r, \theta_{r,t}(x)) - b(r, \theta_{r,t}(x'))|\mathrm{d}r \\ &\lesssim \int_s^{s_0} |\theta_{r,s}(x) - \theta_{r,s}(x')|^\beta \mathrm{d}r \lesssim |x - x'|^{2+\beta},\end{aligned}$$

where the last step is due to $|r - s| \leqslant |x - x'|^2 \leqslant |t - s|/4$. The desired claim (A.5) follows by (2.8) and elementary but cubersome calculations.

Now, integrating by parts, we get from (3.9), (A.5) and Lemma 2.5

$$\begin{aligned}|I_2(s,t,x,x',y)| &\leqslant \int_{\mathbb{R}^d} |\widetilde{p}^{(s,x)}(s,x,s_0,z) - \widetilde{p}^{(s,x')}(s,x,s_0,z)| \cdot |\nabla^2_z p(s_0,z,t,y)|\mathrm{d}z, \\ &\lesssim \big(|x - x'|^\alpha + |x - x'|^\beta\big)(t - s_0)^{-1} \int_{\mathbb{R}^d} \mathbf{g}_\lambda(s,x,s_0,z)\mathbf{g}_{\lambda'}(s_0,z,t,y)\mathrm{d}z \\ &\lesssim \big(|x - x'|^\alpha + |x - x'|^\beta\big)(t - s)^{-1}\mathbf{g}_{\lambda''}(s,x,t,y).\end{aligned}$$

For I_3, by (A.4) and using arguments completely similar to those of Section 3.2, we have

$$|I_3(s,t,x,x',y)| \lesssim |x - x'|^\alpha (t - s)^{-1}\mathbf{g}_{\lambda''}(s,x,t,y).$$

Finally, for I_4, from (3.2), we have

$$\begin{aligned}|I_4(s,t,x,x',y)| &\lesssim \int_s^{s_0}\int_{\mathbb{R}^d} \frac{(\mathbf{g}_{\lambda_1}(s,x,r,z) + \mathbf{g}_{\lambda_1}(s,x',r,z))\mathbf{g}_{\lambda_2}(r,z,t,y)}{(r-s)^{1-\frac{\alpha}{2}}(t-r)}\mathrm{d}z\mathrm{d}r \\ &\lesssim \big(\mathbf{g}_{\lambda_3}(s,x,t,y) + \mathbf{g}_{\lambda_3}(s,x',t,y)\big)\int_s^{s_0} \frac{\mathrm{d}r}{(r-s)^{1-\frac{\alpha}{2}}(t-r)} \\ &\lesssim \frac{|x - x'|^\alpha}{t - s}\big(\mathbf{g}_{\lambda_3}(s,x,t,y) + \mathbf{g}_{\lambda_3}(s,x',t,y)\big),\end{aligned}$$

where we have used that $|x - x'|^2 \leqslant (t - s)/4$ and $s_0 = s + |x - x'|$. Combining the above calculations, we obtain (A.1). □

Acknowledgment

Research of S. Menozzi is funded by the Russian Academic Excellence Project '5-100'. Research of X. Zhang is funded by NNSFC grant of China (No. 11731009) and the DFG through the CRC 1283 "Taming uncertainty and profiting from randomness and low regularity in analysis, stochastics and their applications".

References

[1] Aronson, D. G. The fundamental solution of a linear parabolic equation containing a small parameter. *Ill. Journ. Math. 3* (1959), 580–619.

[2] Aronson, D. G. Bounds for the fundamental solution of a parabolic equation. *Bull. Amer. Math. Soc. 73* (1967), 890–896.

[3] Bass, R. F. *Diffusions and Elliptic Operators.* Springer, 1997.

[4] Bass, R. F., and Perkins, E. A new technique for proving uniqueness for martingale problems. *Astérisque*, 327 (2009), 47–53 (2010).

[5] Bonfiglioli, A., Lanconelli, E., and Uguzzoni, F. Fundamental solutions for non-divergence form operators on stratified groups. *Transactions of the American Mathematical Society 356*, 7 (2004), 2709–2737.

[6] Boué, M., and Dupuis, P. A variational representation for certain functionals of Brownian motion. *Ann. Probab. 26*, 4 (1998), 1641–1659.

[7] Boutiah, S. E., Rhandi, A., and Tacelli, C. Kernel estimates for elliptic operators with unbounded diffusion, drift and potential terms. *Discrete Contin. Dyn. Syst. 39*, 2 (2019), 803–817.

[8] Chaudru de Raynal, P. E. Strong existence and uniqueness for degenerate SDE with Hölder drift. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques 53*, 1 (Feb. 2017), 259–286.

[9] Chaudru de Raynal, P.-E., Honoré, I., and Menozzi, S. Sharp Schauder estimates for some Degenerate Kolmogorov Equations. *To appear in Annali Scuola Normale Superiore di Pisa* (2018).

[10] Chen, Z., Hu, E., Xie, L., and Zhang, X. Heat kernels for non-symmetric diffusion operators with jumps. *J. Differential Equations 263* (2017), 6576–6634.

[11] Deck, T., and Kruse, S. Parabolic differential equations with unbounded coefficients - a generalization of the parametrix method. *Acta Appl. Math. 74*, 1 (2002), 71–91.

[12] Delarue, F., and Menozzi, S. Density estimates for a random noise propagating through a chain of differential equations. *J. Funct. Anal. 259*, 6 (2010), 1577–1630.

[13] Di Francesco, M., and Pascucci, A. On a class of degenerate parabolic equations of Kolmogorov type. *AMRX Appl. Math. Res. Express 3* (2005), 77–116.

[14] Fleming, W. H., and Sheu, S. J. Stochastic variational formula for fundamental solutions of parabolic PDE. *Appl. Math. Optim. 13*, 3 (1985), 193–204.

[15] Friedman, A. *Partial differential equations of parabolic type.* Prentice-Hall Inc., Englewood Cliffs, N.J., 1964.

[16] Friedman, A. *Stochastic differential equations.* Chapmann-Hall, 1975.

[17] Gobet, E. LAN property for ergodic diffusions with discrete observations. *Ann. Inst. H. Poincaré Probab. Statist. 38*, 5 (2002), 711–737.

[18] Menozzi, S. Parametrix techniques and martingale problems for some degenerate Kolmogorov equations. *Electron. Commun. Probab. 16* (2011), 234–250.

[19] Metafune, G., Ouhabaz, E. M., and Pallara, D. Long time behavior of heat kernels of operators with unbounded drift terms. *J. Math. Anal. Appl. 377*, 1 (2011), 170–179.

[20] Nualart, D. *The Malliavin Calculus and Related Topics.* Springer, 2006.

[21] Pagès, G., and Panloup, F. Total Variation and Wasserstein bounds for the ergodic Euler-Naruyama scheme for diffusions. *Preprint* (2020).

[22] Pascucci, A., and Pesce, A. On stochastic Langevin and Fokker-Planck equations: the two-dimensional case. *arXiv:1910.05301* (2019).

[23] Pascucci, A., and Pesce, A. The parametrix method for parabolic SPDEs. *Stochastic Processes and Applications, https://doi.org/10.1016/j.spa.2020.05.008* (2020).

[24] Polidoro, S. On a class of ultraparabolic operators of Kolmogorov-Fokker-Planck type. *Matematiche (Catania) 49*, 1 (1994), 53–105.

[25] Sheu, S. J. Some estimates of the transition density of a nondegenerate diffusion Markov process. *Ann. Probab. 19*, 2 (1991), 538–561.

[26] Stroock, D. W., and Varadhan, S. R. S. *Multidimensional diffusion processes*, vol. 233 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, 1979.

[27] Wang, F. Y., and Zhang, X. Derivative formula and applications for degenerate diffusion semigroups. *J. Math. Pures Appl. 99* (2013), 726–740.

[28] Zhang, X. A variational representation for random functionals on abstract Wiener spaces. *J. Math. Kyoto Univ. 49*, 3 (2009), 475–490.

[29] Zhang, X. Fundamental solutions of nonlocal Hörmander's operators. *Commun. Math. Stat. 4* (2016), 359–402.