Strong regularization by Brownian noise propagating through a weak Hörmander structure

Paul-Éric Chaudru de Raynal · Igor Honoré · Stéphane Menozzi

Abstract
We establish strong uniqueness for a class of degenerate SDEs of weak Hörmander type under suitable Hölder regularity conditions for the associated drift term. Our approach relies on the Zvonkin transform which requires to exhibit good smoothing properties of the underlying parabolic PDE with rough, here Hölder, drift coefficients and source term. Such regularizing effects are established through a perturbation technique (forward parametrix approach) which also heavily relies on appropriate duality properties on Besov spaces. For the method employed, we exhibit some sharp thresholds on the Hölder exponents for the strong uniqueness to hold.

Mathematics Subject Classification Primary 60H30 · 60H10; Secondary 34F05 · 35H10
1 Introduction

1.1 Statement of the problem

In this work, we aim at establishing a strong well-posedness result outside the classical Cauchy-Lipschitz framework for the following degenerate Stochastic Differential Equation (SDE) of Kolmogorov type:

\[
\begin{align*}
\frac{dX_1}{dt} &= F_1(t, X_1^{t}, \ldots, X_n^{t})dt + \sigma(t, X_1^{t}, \ldots, X_n^{t})dW_t, \\
\frac{dX_2}{dt} &= F_2(t, X_1^{t}, \ldots, X_n^{t})dt, \\
\frac{dX_3}{dt} &= F_3(t, X_2^{t}, \ldots, X_n^{t})dt, \\
& \vdots \\
\frac{dX_n}{dt} &= F_n(t, X_{n-1}^{t}, X_n^{t})dt,
\end{align*}
\tag{1.1}
\]

where \((W_t)_{t \geq 0}\) stands for a \(d\)-dimensional Brownian motion on some filtered probability space \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)\) and for all \(i \in \{1, n\}\), \(t \geq 0\) the component \(X_i^t\) is \(\mathbb{R}^d\)-valued as well (i.e. \(X_i \in \mathbb{R}^{nd}\)). We suppose that the \((F_i)_{i \in \{2, n\}}\) satisfy a kind of weak Hörmander condition, i.e. the matrices \(\left(D_{X^{t-1}_i}F_i(t, \cdot)\right)_{i \in \{2, n\}}\) have full rank. However, the coefficients \((F_i)_{i \in \{2, n\}}\) can be rather rough in their other entries, namely, Hölder continuous. We assume as well that the diffusion coefficient \(\sigma\) is bounded from above and below and spatially Lipschitz continuous.

For a system of Ordinary Differential Equation (ODE) it may be a real challenge to prove the well-posedness outside the Lipschitz framework (see e.g. [23]) and, as shown by Peano’s example, uniqueness may fail as soon as the drift of the system of interest is only Hölder continuous. For an SDE, the story is rather different since the presence of the noise may allow to restore well-posedness. Such a phenomenon, called regularization by noise (see the Saint Flour Lecture notes of Flandoli [28] and the references therein for an overview of the topic), has been widely studied since the pioneering works of Zvonkin [69] and Veretennikov [64] who establish, respectively in the scalar and multidimensional setting, strong well-posedness for non-degenerate Brownian SDEs with bounded and measurable drift. We recall that a strong solution is adapted to the Brownian filtration generated by the driving noise and that non-degenerate means that the noise has the same dimension as the underlying system on which it acts.

Let us mention, among others, and still within the non-degenerate setting, the works of Krylov and Röckner [40] (\(L_q - L_p\) drift), Zhang [67] (\(L_q - L_p\) drift and weakly Lipschitz diffusion matrix), Fedrizzi and Flandoli [24] (\(L_q - L_p\) and Hölder drift). We also mention Flandoli et al. [26] and Beck et al. [7] for connections with stochastic transport equations.

The crucial assumption, shared by all the aforementioned results, is the non-degeneracy condition on the noise added in the considered system. A possible approach to relax this hypothesis was proposed by Veretennikov in [65], where the author extended the result in [64] to some specific cases of the considered chain (1.1) for \(n = 2\). In comparison with our setting, the author does not impose any non-degeneracy

---

1 We will use throughout the paper the notation \([\cdot, \cdot]\) for integer intervals.
condition on $D_{x_i}F_2(t, \cdot)$. The price to pay is anyhow that all coefficients (i.e. with the notations of (1.1) $F_1, F_2, \sigma$) need to be twice continuously differentiable functions with bounded derivatives w.r.t. the degenerate component, meaning that no regularization by noise is investigated in the degenerate direction. More generally, it is useless to expect a generalization of the previous results without any additional assumption: we can benefit from the regularization by noise phenomenon only in the directions submitted to the noise.

In our current framework, the non-degeneracy assumption on the Jacobian $(D_{x_i}F_i(t, \cdot)), i \in \mathbb{[2, n]}$ precisely allows the noise to propagate through the chain passing from the $i^{th}$ to the $(i + 1)^{th}$ level thanks to the drift, hence leading to a propagation of the noise in the whole considered space. The main idea is then to take advantage of this particular propagation, known as weak Hörmander setting (in reference to the work of Hörmander on hypoelliptic differential operator [35]), to restore strong well-posedness under our current Hölder framework. This feature has already been considered in the literature for the system (1.1) in the kinetic case (i.e. when $n = 2$), see the works of Chaudru de Raynal [10], Wang and Zhang [66], Fedrizzi et al. [25], Zhang [68]. In any cases, in addition to the weak Hörmander structure, the regularity of the drift w.r.t. the second space (and hence degenerate) argument is required to be of regularity index superior or equal (depending on the work) to $2/3$ (usually called critical Hölder index or critical weak differentiation index). As a generalization of these results, we prove in this paper that strong well-posedness holds as soon as each drift component $F_i$ is $\beta_j$-Hölder continuous in the $j^{th}$ variable for some $\beta_j \in \left(\frac{2j - 2}{2j - 1}, 1\right]$ when $i \leq j$, so that we recover the critical index mentioned above when $j = 2$. We refer to Sect. 1.5 for a thorough discussion about those facts.

1.2 Notations, assumptions and main result

Some notations. We will denote by a bold letter, e.g. $x$ or $y$, any element of $\mathbb{R}^{nd}$, writing as well $x = (x_1, \ldots, x_n)$ where for $i \in \mathbb{[1, n]}$, $x_i \in \mathbb{R}^d$. For practical purpose we will be led in our analysis to consider subcomponents of a vector $x \in \mathbb{R}^{nd}$. Namely, for any $1 \leq i \leq j \leq n$ and $x \in \mathbb{R}^{nd}$, we introduce the notation $x_{i:j} := (x_i, \ldots, x_j)$. Accordingly, we write the drift as the mapping

$$(s, x) \in \mathbb{R}^+ \times \mathbb{R}^{nd} \mapsto F(s, x) = \left(\begin{array}{c}
F_1(s, x) \\
\vdots \\
F_n(s, x)
\end{array}\right) = \left(\begin{array}{c}
F_1(s, x) \\
F_2(s, x) \\
F_3(s, x_{2:n}) \\
\vdots \\
F_n(s, x_{n-1:n})
\end{array}\right),$$

from the specific structure of the drift appearing in (1.1).

For $f \in C^1(\mathbb{R}^{nd}, \mathbb{R}^k), k \in \{1, d\}$, we denote for each $i \in \mathbb{[1, n]}$, by $D_{x_i}f(x)$ the Jacobian matrix of the derivative of $f$ w.r.t. its $\mathbb{R}^d$-valued variable $x_i$. As shortened form, and when no ambiguity is possible, we also write for all $x, y \in \mathbb{R}^{nd}$,
\[ D_{x_j} f(x) = D_i f(x) \text{ and } D_{y_j} f(y) = D_i f(y). \] Also, if \( k = 1 \) we denote by \( D f(x) = (D_1 f(x), \ldots, D_n f(x))^* \), where \( \cdot^* \) stands for the transpose, the full gradient of the function \( f \) at point \( x \).

Let \( f : \mathbb{R}^{nd} \to \mathbb{R}^k \) and \( \beta : (= (\beta_1, \ldots, \beta_n) \in (0, 1]^n \) be a multi-index. We say that \( f \) is uniformly \( \beta \)-Hölder continuous if for each \( j \in \llbracket 1, n \rrbracket \)

\[
[(f_j)_{\beta_j} := \sup_{(z_1:j-1, z_{j+1:n}) \in \mathbb{R}^{(n-1)d}, z \neq z', (z, z') \in (\mathbb{R}^d)^2} \frac{|f(z_1:j-1, z, z_{j+1:n}) - f(z_1:j-1, z', z_{j+1:n})|}{|z - z'|^{\beta_j}} < +\infty. \] (1.2)

For a smooth function \( \Psi : [0, T] \times \mathbb{R}^{nd} \to \mathbb{R}^{nd} \), where \( T > 0 \) is a fixed given time, writing for \( (t, x) \in [0, T] \times \mathbb{R}^{nd}, \Psi(t, x) = (\Psi_1(t, x), \ldots, \Psi_n(t, x)) \) where for each \( i \in \llbracket 1, n \rrbracket, \Psi_i \) is \( \mathbb{R}^d \) valued, we denote by

\[
\|D \Psi\|_\infty := \sum_{i=1}^n \sup_{(t, x) \in [0, T] \times \mathbb{R}^{nd}} \|D \Psi_i(t, x)\|, \|D(D_1 \Psi)\|_\infty
:= \sum_{i=1}^n \sup_{(t, x) \in [0, T] \times \mathbb{R}^{nd}} \|D(D_1 \Psi_i)(t, x)\|,
\] (1.3)

where in the above equation \( \| \cdot \| \) stands for a tensor norm in the appropriate corresponding dimension. Precisely, \( D \Psi_i(t, x) \in \mathbb{R}^{nd} \otimes \mathbb{R}^d \) and \( D(D_1 \Psi_i)(t, x) \in \mathbb{R}^{nd} \otimes \mathbb{R}^d \otimes \mathbb{R}^d \).

**Assumptions.** We will assume throughout the paper that the following conditions hold.

**ML** The coefficients \( \Phi \) and \( \sigma \) are measurable in time and \( \Phi(t, 0) \) is bounded uniformly in time. Also, the diffusion coefficient \( \sigma \) is uniformly Lipschitz continuous in space, uniformly in time, i.e. there exists \( \kappa > 0 \) s.t. for all \( t \geq 0, (x, x') \in (\mathbb{R}^{nd})^2 \):

\[
|\sigma(t, x) - \sigma(t, x')| \leq \kappa |x - x'|.
\]

**UE** The diffusion matrix \( a := \sigma \sigma^* \) is uniformly elliptic and bounded, uniformly in time, i.e. there exists \( \Lambda \geq 1 \) s.t. for all \( t \geq 0, (x, \xi) \in \mathbb{R}^{nd} \times \mathbb{R}^d \):

\[
\Lambda^{-1} |\xi|^2 \leq \langle a(t, x) \xi, \xi \rangle \leq \Lambda |\xi|^2.
\]

**T(β)** For all \( j \in \llbracket 1, n \rrbracket \), the functions \( (F_j)_{i \in \llbracket 1, j \rrbracket} \) are uniformly \( \beta_j \)-Hölder continuous in the \( j^{th} \) spatial variable with \( \beta_j \in ((2j - 2)/(2j - 1), 1] \), uniformly w.r.t. the other spatial variables of \( F_i \) and in time. In particular, there exists a finite constant \( C_\beta > 0 \) s.t. with the notations of (1.2),

\[
\max_{i \leq j \in \llbracket 1, n \rrbracket} \sup_{s \in [0, T]} [(F_j s, \cdot)]_{\beta_j} \leq C_\beta.
\]
For all $i \in \llbracket 2, n \rrbracket$, there exists a closed convex subset $E_{i-1} \subset GL_d(\mathbb{R})$ (set of invertible $d \times d$ matrices) s.t., for all $t \geq 0$ and $(x_{i-1}, \ldots, x_n) \in \mathbb{R}^{(n-i+2)d}$, $D_{x_{i-1}}F_i(t, x_{i-1}, \ldots, x_n) \in E_{i-1}$. For example, $E_{i-1}$ may be a closed ball included in $GL_d(\mathbb{R})$ the latter being an open set. Moreover, $D_{x_{i-1}}F_i$ is $\eta$-Hölder continuous w.r.t. $x_{i-1}$ uniformly in $x_{i:n}$ and time. We also assume without loss of generality that $\eta \in \left(0, \inf_{j \in [2, n]} \left\{ \beta_j - (2j - 2)/(2j - 1) \right\} \right)$, i.e. $\eta$ is meant to be small. Importantly, we assume that $D_{x_{i-1}}F_i$ is also bounded (which is automatically the case if the $E_{i-1}$ are balls).

From now on, we will say that assumption (A) is in force provided that (ML), (UE), (T) and (H$_\eta$) hold.

**Main result** The main result of this work is the following theorem.

**Theorem 1** (Strong uniqueness for the degenerate system (1.1)) Under (A) there exists a unique strong solution to system (1.1).

**Remark 1** Still in comparison with the results obtained in the non-degenerate cases, and especially the one of Krylov and Röckner [40], we do not tackle the case of drifts in $L_q - L_p$ w.r.t. the first (and then non-degenerate) variable. This is only to keep our result as clear as possible and to concentrate on the novelty of the approach we use here. We are anyhow confident that these specific drifts could be handled. Indeed, all the intermediate results needed to perform the analysis in that setting seem to be already available. We refer to Subsect. 1.5 for further details.

**1.3 Related discussions and perspectives**

One may wonder if the thresholds in (T$_\beta$) are sharp. To investigate this question, we should recast our result within the framework of the regularization by noise phenomenon, see e.g. the lecture notes by Flandoli [28]. Such phenomenon has been considered for ODEs perturbed by many noise classes (in particular $\alpha$-stable, $\alpha \in (0, 2]$ and fractional Brownian motion with Hurst index $H$ in $(0, 1)$) and has given rise to a large literature. Before going further, we feel that it is important to specify what “regularization” means when speaking about regularization by noise. Hereafter, we say that a noise “regularizes” an ill posed system if the resulting perturbed equation has a unique solution. Within our probabilistic setting, we can thus distinguish three types of regularizations: the weak regularization, for which the law of the solution is concerned; the strong regularization and path by path regularization which relate to the path of the solution. While the difference between weak and path by path or strong well-posedness is clear, we refer to the recent work of Shaposhnikov and Wresch [61] for the more subtle differences between strong and path by path uniqueness (let us only mention that path by path uniqueness implies strong uniqueness while the converse is not true).

Also, in order to illustrate the following discussion on the thresholds in (T$_\beta$), we introduce, for a given scalar $\alpha$-stable$^2$, $\alpha \in (0, 2]$ or $H$-fractional Brownian, $H \in (0, 1)$ the usual fractional Laplacian.

---

$^2$ For simplicity reasons, we restrain our considerations to rotationally invariant stable processes with generator $1/2$ the usual fractional Laplacian.
(0, 1) noise \( \mathcal{W} \) of self similarity index \( \gamma > 0 \), the following simplified version of (1.1)

\[
dX_t = AX_t dt + B d\mathcal{W}_t + P(X_t) dt, \quad A = \begin{pmatrix} 0 & \cdots & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}, \quad X_0 = 0, \tag{1.4}
\]

where \( d = 1, P = (P_1, \ldots, P_n)^* \) and for each \( i \in [1, n], P_i(x) = \sum_{j=i}^{n} P_i^j(x) \) with \( P_i^j(x) := c_{i,j} \text{sgn}(x_j) (|x_j|^{\beta_j} \wedge 1) \) with \( c_{i,j} \in \mathbb{R} \) and \( B = (1, 0, \ldots, 0)^* = (1, 0_{1,n-1})^* \) where “*” stands for the transpose. The dynamics (1.4) can be viewed as a perturbation of \( \tilde{X}_t = A\tilde{X}_t dt + B d\mathcal{W}_t, \tilde{X}_0 = 0 \), corresponding to the noise and its iterated integrals, i.e. \( \tilde{X}_t = (\mathcal{W}_t, \int_0^t \mathcal{W}_s ds, \ldots, \int_0^t \int_0^{t_1} \int_0^{t_2} \cdots \int_0^{t_{n-1}} dt_n \mathcal{W}_{t_n}^*)^* \), by a Peano type drift \( P \). With the notations of (1.1), the full drift of (1.4) writes \( F(t, x) = F(x) = Ax + P(x) \).

Note that the above truncation in \( P \) is only introduced to avoid technical considerations in the following discussion (simple statement of parabolic bootstrap results). However, our approach allows to consider general non-linear unbounded drifts satisfying (T_\beta) and (H_n). We emphasize that, when \( n = 1 \), setting \( \beta_1^1 := \beta, c_{1,1} = 1 \), we obtain the following dynamics

\[
dX_t = \text{sgn}(X_t)(|X_t|^{\beta_1} \wedge 1) dt + d\mathcal{W}_t, \quad X_0 = 0, \tag{1.5}
\]

known as the (localized version of the) stochastic Peano example and when \( n = 2, P_1 \equiv 0, c_{2,2} = 1 \), we get that

\[
dX_1^2 = d\mathcal{W}_t, \quad dX_1^2 = \text{sgn}(X_2^2)(|X_2^2|^{\beta_2} \wedge 1) dt + d \left\{ \int_0^t ds \mathcal{W}_s \right\}, \quad X_0^1 = X_0^2 = 0, \tag{1.6}
\]

which is the kinetic version of the above stochastic Peano example.

**Weak regularization for the stochastic Peano example (1.5).** Let us start with the weak regularization by noise, as it seems to us that it is the most understood and understandable setting. Weak well-posedness relies on the well-posedness of the martingale formulation for the system which itself relies, in turn, on a good theory for the associated PDE (see e.g. [60]). Roughly speaking, it consists in showing that the transport term for the associated PDE is somehow a negligible perturbation of the equation. This can be seen, at least heuristically, through scaling arguments. Consider indeed the dynamics (1.5) and assume that \( \mathcal{W} \) is an \( \alpha \)-stable process. The formal associated PDE writes

\[
\partial_t u(t, x) + (\text{sgn}(x)(|x|^{\beta} \wedge 1), Du(t, x)) + \frac{1}{2} \Delta^\alpha u(t, x) = 0,
\]
where $\Delta^{\alpha/2}$ stands for the usual fractional Laplacian. Introduce for $\lambda > 0$ the corresponding rescaled function $u_\lambda(t, x) := u(\lambda t, \lambda^{1/\alpha} x)$ (scaling reflecting the parabolic scale for $t$ and $x$). We get that $u_\lambda$ satisfies the equation:

$$
\partial_t u_\lambda(t, x) + \langle \lambda \frac{1}{\alpha} \text{sgn}(x) \left( |\lambda^{\frac{1}{\alpha}} x|^\beta \wedge 1 \right), D u_\lambda(t, x) \rangle + \frac{1}{2} \Delta^{\alpha/2} u_\lambda(t, x) = 0,
$$

so that the terms associated with the principal part of the partial differential operator in the above PDE, namely $\partial_t u_\lambda$ and $\Delta^{\alpha/2} u_\lambda(t, x)$, are comparable. On the other hand, when $x$ belongs to compact subsets of $\mathbb{R}$:

- if $\beta > 1 - \alpha$, the scaled drift coefficient goes to zero with $\lambda$ and the smoothing effect of the principal part of the partial differential operator dominates;
- if $\beta = 1 - \alpha$, the scaled drift coefficient stays at a macro scale and the rescaled drift has the same order as the principal part of the partial differential operator (critical case);
- otherwise, the drift explodes when $\lambda$ goes to zero.

In other words, a sufficient condition on the exponent $\beta$ to ensure that the drift is a negligible perturbation of the equation is

$$
\beta > 1 - \alpha.
$$

This rule can also be obtained by analyzing the path associated with the solution of (1.5), see e.g. the work [21] and the counter-examples to weak uniqueness in [11, 14]. From these pathwise analysis, one can formally generalize the above heuristic and get that, for the system (1.5) driven by a general $\gamma$ self-similar noise $\mathcal{W}$, the condition

$$
\beta > 1 - \frac{1}{\gamma},
$$

(1.7)

is needed for the weak well-posedness to hold. When $\gamma = 1/\alpha$, $\alpha$ in (0, 1] we refer to e.g. [15, 16] for results in that direction and to [9, 13, 20, 27, 45, 70] for results related to the case where $\alpha$ in (1, 2] so that the above rule allows $\beta$ to take negative values, the drift being then a distribution.

**Thresholds for weak well-posedness and thresholds in $(T_\beta)$.** Let us come back to the more tricky system (1.4) and let us consider first the kinetic version (1.6) with $\mathcal{W} = W$. In this last case, the associated PDE writes

$$
\partial_t u(t, x_1, x_2) + \langle x_1, D x_2 u(t, x_1, x_2) \rangle + \langle \text{sgn}(x_2) (|x_2|^{\beta_2} \wedge 1), D x_2 u(t, x_1, x_2) \rangle
\]

$$
+ \frac{1}{2} \Delta_{x_1} u(t, x_1, x_2) = 0.
$$

We introduce the corresponding rescaled function $u_\lambda(t, x) := u(\lambda t, \lambda^{1/2} x_1, \lambda^{3/2} x_2)$ (scaling reflecting the usual parabolic scale for $t$ and $x_1$ and the fast regime associ-
associated with the integral of the Brownian motion) which solves the equation

$$\begin{align*}
\frac{\partial_t u\lambda(t, x_1, x_2)}{} + \langle x_1, D x_2 u\lambda(t, x_1, x_2) \rangle \\
+ \lambda^{-\frac{1}{2}} \langle \text{sgn}(x_2) \left( \left| \frac{x_2}{\lambda} \right|^2 \right)^{\beta_2^2} \wedge 1, D x_2 u\lambda(t, x_1, x_2) \rangle \\
+ \frac{1}{2} \Delta x_1 u\lambda(t, x_1, x_2) = 0,
\end{align*}$$

so that, again, the terms $\partial_t u\lambda$, $\Delta x_1 u\lambda(t, x_1, x_2)$ and $\langle x_1, D x_2 u\lambda(t, x_1, x_2) \rangle$ associated with the principal part of the partial differential operator in the above PDE are comparable. On the other hand, when $x_2$ belongs to compact subsets of $\mathbb{R}$:

- if $\beta_2^2 > 1/3$, the scaled drift term $\lambda^{-1/2} \langle \text{sgn}(x_2) \left( \left| \frac{x_2}{\lambda} \right|^2 \right)^{\beta_2^2} \wedge 1, D x_2 u\lambda(t, x_1, x_2) \rangle$ goes to zero with $\lambda$, and the smoothing effect of the principal part of the partial differential operator dominates;
- if $\beta_2^2 = 1/3$, the previous scaled drift coefficient stays at a macro scale and has the same order as the principal part of the partial differential operator (critical case);
- otherwise, the drift explodes when $\lambda$ goes to zero.

Observe that $1/3 = 1 - 1/(3/2)$ and that $3/2$ is precisely the self similarity index of $\int_0^t W_s ds$, so that we recover the rule (1.7). This threshold has also been shown to be (almost) sharp through counter-examples in [11]. In comparison, the threshold introduced in $(T_\beta)$ gives $\beta_2^2 \beta_2 = 2/3 = 1 - 1/\gamma > 1/3$ so that, in this case, we lose a factor 2 w.r.t. the thresholds predicted by (1.7).

Reproducing then the above analysis for the PDE formally associated with (1.4), we obtain that the degenerate part of the scaled drift, i.e. $\sum_{i=2}^n \left| \lambda^{-i+3/2} \langle \text{sgn}(x_j) \left( \left| \frac{x_j}{\lambda} \right|^2 \right)^{\beta_j^i} \wedge 1, D x_i \rangle \right|$, explodes when $\beta_j^i < (2i - 3)/(2j - 1)$, $i \leq j$ in $\mathbb{I}[1, n]^2$. These thresholds have also been shown to be (almost) sharp in [14] through counter-examples. On the other hand, along the diagonal of the system (i.e. for the indexes $\beta_j^j$), this gives that weak uniqueness fails as soon as $\beta_j^j < (2j - 3)/(2j - 1) = 1 - 1/(j - 1/2)$ where $(j - 1/2)$ again precisely corresponds to the self similarity index of the $(j - 1/2)$th iterated in time integral of the Brownian motion. Still in comparison, we assumed in $(T_\beta)$ that each $\beta_j^j = \beta_j$ is (strictly) greater than $(2j - 2)/(2j - 1) = 1 - 1/\gamma > 1/2 \times (j - 1/2)]$. Therefore, we lose this factor 2 for each variable of the chain (note further that we also lose the dependence of the thresholds w.r.t. the level of the chain, i.e. $\beta_j^j = \beta_j$, $i \in \mathbb{I}[1, n]$). We are hence not able to prove strong well-posedness for (1.5) as soon as $\beta \leq 1 - 1/(2\gamma)$, $\gamma = n + 1/2, n \in \mathbb{N}^*$. One may thus wonder if this factor 2 is the price to pay to pass from the weak to strong regularization phenomenon.

**Strong regularization** The first results on strong regularization go back to the pioneering works of Zvonkin [69] and Veretennikov [64] where (1.5) with $\mathcal{W} = W$ is shown to be well-posed in a strong sense as soon as $\beta \geq 0$ (the result holds therein for any bounded and measurable drift). This result has then been extended in the seminal work of Krylov and Röckner [40] for $L_q - L_p$ singular drifts (see also [67] and [24]). When $\mathcal{W}$ is a pure jump process ($\alpha \in (1, 2)$), Priola showed in [54] that strong well-posedness holds for any $\beta > 1 - \alpha/2$ for bounded Hölder drifts and Chen *et al.*
obtained in [18] the same condition for $\alpha \in (0, 1)$. When $\mathcal{W}$ is a fractional Brownian motion with Hurst parameter $H > 1/2$, it has been shown by Nualart and Ouknine [52] that (1.5) is well-posed as soon as $\beta > 1 - 1/(2H)$. This last result relies on the Girsanov transform and the Yamada-Watanabe Theorem to deduce that strong existence and uniqueness in law give strong uniqueness. Putting those results together and denoting by $\gamma$ the self similarity index of $\mathcal{W}$, we thus get (at least in the stable and fractional setting) that strong well-posedness holds if $\beta > 1 - 1/(2\gamma)$, $\gamma \in (0, 2]$ and that the “critical” case is attained for $\gamma = 1/2$. Again, a factor 2 is lost in comparison with (1.7) and these results exhibit the same type of thresholds as ours.

In the Markovian setting, a good manner to understand how such a factor (and thus threshold) appears consists in investigating the smoothing properties of the underlying PDE. It is indeed important to notice that most of the aforementioned results concerning strong uniqueness are based on the same technology: the Zvonkin transform of the SDE. This is precisely where PDEs come into play.

**A primer about strong regularization and the Zvonkin method.** Let us focus on the Markovian (stable) case. The main idea consists in rewriting the dynamics (1.5) as

$$X_t = u(t, X_t) + X_0 - u(0, X_0) - M_{0,t}(\alpha, u, X) + \mathcal{W}_t,$$  

with

$$M_{0,t}(\alpha, u, X) := \begin{cases} \int_0^t Du(s, X_s) \cdot d\mathcal{W}_s, & \text{if } \alpha = 2; \\ \int_0^t \int_{\mathbb{R}\setminus\{0\}} N(ds/dz)\{u(s, X_s + z) - u(s, X_s)\}, & \text{if } \alpha < 2, \end{cases}$$

where $\tilde{N}$ is the compensated Poisson measure associated with $\mathcal{W}$ and where $u$ is the solution of

$$\partial_t u(t, x) + (\text{sgn}(x)(|x|^\beta \wedge 1), Du(t, x)) + \frac{1}{2} \Delta^\frac{\gamma}{2} u(t, x) = \text{sgn}(x)(|x|^\beta \wedge 1), u(T, \cdot) = 0.$$  

Equation (1.8) then follows from the Itô formula provided $u$ is smooth enough. This transformation allows to get rid of the bad drift in the equation and to replace it by the solution of a parabolic PDE which benefits from the smoothing effect associated with the generator of the noise $\mathcal{W}$. The price to pay is that the diffusion matrix in the “martingale” part is now modified, as one adds a stochastic integral involving the derivative or a perturbation of $u$.

For this new equation to be well-posed in any dimension\(^3\), it is commonly assumed that the integrand in the stochastic integral $M(\alpha, u, X)$ must be Lipschitz continuous in the spatial variable (in order to apply a stability type argument based on martingale and Grönwall inequalities). This roughly means that for any $t$ in $[0, T]$, the gradient of $u(t, \cdot)$ must be Lipschitz continuous, uniformly in $t$, in the case $\alpha = 2$. In the

---

\(^3\) Note indeed that the scalar case induces very specific features, see e.g. [1, 4, 13, 31].
case $\alpha < 2$, it follows from the interpolation type Lemma 4.1 in [54] that a sufficient condition is that for any $t$ in $[0, T]$, $u(t, \cdot)$ must belong to $C^{1+\eta}$, where $C^{1+\eta}$ stands for the usual Hölder space see e.g. [41], with $\eta > \alpha/2$. This last condition comes from integrability purposes in order to compensate the lack of integrability of the Lévy measure associated with the pure jump process around 0 when applying martingale inequalities. From the corresponding parabolic bootstrap (Schauder estimates for (1.10), see e.g. [30, 39] for $\alpha = 2$ or [15, 49] in the pure jump case), we could expect at best that

$$\|u(t, \cdot)\|_{C^{\alpha+\beta}} \leq \|\text{sgn}(\cdot)(|\cdot|^\beta \wedge 1)\|_{C^\beta}.$$ 

Hence, we obtain that $\beta$ must satisfies

$$\alpha + \beta > 1 + \alpha/2 \iff \beta > 1 - \alpha/2.$$ 

In other words, denoting again by $\gamma$ the self similarity index of the noise, we get that strong well-posedness holds for

$$\beta > 1 - \frac{1}{2\gamma}.$$ 

These are precisely the thresholds obtained in the literature on strong regularization and we will say that they are thus (almost) sharp w.r.t. the methodology. Let us mention that in the critical case in (1.11) (with an equality therein) when the noise is a Brownian motion, the Lipschitz property of the gradient of the solution holds in a suitable $L_p$ space only (this is the so-called Calderón-Zygmund estimate for the case of a bounded and measurable drift and its ad hoc version from [42] in the $L_q - L_p$ framework of Krylov and Röckner). Such type of estimates require non-trivial techniques from harmonic analysis to handle singular integrals, which differ significantly from the Schauder type control adopted here.

The thresholds in $(T_\beta)$ are derived from the Zvonkin method. Let us bring to light how the Zvonkin transform naturally leads to the thresholds in $(T_\beta)$ for our simple system (1.4). The corresponding process $X$ writes

$$X_t = u(t, X_t) + x_0 - u(0, x_0) - \int_0^t Du(s, X_s) BdW_s + W_t.$$ 

where $u = (u_1, \ldots, u_n)^*$ and each $u_i$ solves

$$\partial_t u_i(t, x) + (Ax + P(x)) Du_i(t, x) + \frac{1}{2} \Delta_{x_i} u_i(t, x) = F_i(x).$$

In order to have a Lipschitz continuous in space integrand in the stochastic integral associated with the Zvonkin transform (1.12), we need, due to the particular structure of the embedding matrix $B$ (recall that $B = (1, 0_{1,n-1})^*$), the gradient $D_1 u$ to be Lipschitz in all the spatial directions. We thus ask each component of the function $u$ to have the same regularity (namely their gradient in the non-degenerate direction must be Lipschitz w.r.t. all variables) and this is the reason why the corresponding thresholds do not depend on the level of the chain, as opposed to the weak thresholds. Accordingly, from now on, we denote by $\beta_j$ the regularity index of any component $F_i, i$ in $[[1, n]]$ w.r.t. the $j^{th}$ variable, $j$ in $[[i, n]]$. 

\begin{figure}
\centering
\includegraphics[width=0.5\textwidth]{figure1.png}
\caption{A figure related to the text content.}
\end{figure}
Note that the main particularity of PDE (1.13) comes from the different scales at which each component of the PDE evolves and the unboundedness of the source terms \((F_i)_{i \in \mathbb{I}_{1,n}}\) coming from both the Zvonkin transform and the weak Hörmander setting. Especially, we cannot expect the solutions \((u_i)_{i \in \mathbb{I}_{1,n}}\) to be bounded. The associated parabolic bootstrap is thus more tricky than in (1.10). In the companion work [12], assuming that the source term in the right-hand side of (1.13) is bounded and \(\beta_j\)-Hölder w.r.t. the \(j^{\text{th}}\)-variable, we derived Schauder estimates and proved that for any \(i \in \mathbb{I}_{1,n}\), any \(j \in \mathbb{I}_{1,n}\), for fixed \((x_{1:j-1}, x_{j+1:n})\) in \(\mathbb{R}^{n-1}\) and \(t \in [0, T]\), the map

\[
(u_i)_{j}(t, \cdot) : \mathbb{R} \ni z_j \mapsto u_i(t, x_{1:j-1}, z_j, x_{j+1:n}) \in \mathbb{R}
\]

belongs to \(C^{2/(2j-1)+\beta_j}\) for any \(\beta_j\) in \((0, 1/(2j-1)]\), uniformly w.r.t. \(i\) and \((x_{1:j-1}, x_{j+1:n})\). This means that the smoothing effect of the hypoelliptic operator \(\partial_t + A_{x_1} + \langle A_x, D \rangle\) yields a regularity gain of order \(2/(2j-1)\) in the \(j^{\text{th}}\) variable. In other words, the smoothing effect decreases as one moves away from the source of noise. This particular feature is reminiscent from the weak Hörmander like structure of (1.13). Especially, one formally understands that, as \((u_i)_{j}(t, \cdot)\) belongs to \(C^{2/(2j-1)+\beta_j}\), \((D_1u_i)_{j}\) belongs to \(C^{1/(2j-1)+\beta_j}\), i.e. for homogeneity reasons, one differentiation w.r.t. the non-degenerate variable induces a loss of Hölder regularity of order \(1/(2j-1)\) w.r.t. the \(j^{\text{th}}\) degenerate variable. In the current work, we first manage to extend such type of estimates to unbounded sources. As we cannot expect anymore the solutions \((u_i)_{i \in \mathbb{I}_{1,n}}\) to be bounded, but to have linear growth (and consequently to have bounded gradients), we therefore specifically state the parabolic bootstrap in terms of usual Hölder spaces on the gradients. Namely, we manage below to prove that for any \(\beta_j\) in \(((2j-2)/(2j-1), 1)\), the map \((D_1u_i)_{j}(t, \cdot)\) belongs uniformly to \(C^{1/(2j-1)+\beta_j-\epsilon}\) for any \(0 < \epsilon \ll 1\) (see Sect. 1.5.1 for details). To obtain the Lipschitz control in all the spatial directions of the gradient \(D_1u\), we therefore need \(1/(2j-1) + \beta_j > 1 \Leftrightarrow \beta_j > (2j-2)/(2j-1)\), which is precisely the thresholds assumed in \((T^\beta)\).

Thus, one may conclude that the thresholds in \((T^\beta)\) are the one deriving from the Zvonkin method combined with the Schauder type approach. Let us eventually mention that other authors precisely recover those thresholds in the case \(n = 2\), see e.g. [10, 25, 66]. Note also that, therein, the critical case \((\beta_2 = 2/3)\) for the degenerate component is left open. As already mentioned, such framework requires to use significantly different techniques that are out of the scope of the present work as well.

Path by path regularization, associated thresholds and Zvonkin thresholds. To the best of our knowledge, the first result on path by path regularization goes back to the work of Davie [19] and has then been formalized through the works of Flandoli [28, 29]. The main particularity of that approach relies on the fact that the system (1.5) is considered “\(\omega\) by \(\omega\)”, so that it somehow goes beyond the probabilistic framework. This setting thus matches the rough path perspective for SDEs. Based on these considerations, Catellier and Gubinelli proposed in [17] a systematic study of fractional Brownian perturbations of ODEs, for any Hurst parameter \(H\) in \((0, 1)\). Therein, they prove in particular that (1.5) is well-posed as soon as \(\beta > 1 - 1/(2H)\). Therefore, the
authors obtained the same thresholds as the one required for strong well-posedness in
the literature although the approach, especially for [17], differs significantly from the
PDE based trick of the Zvonkin method. Furthermore, it seems that, modulo an addi-
tional work, the Zvonkin transform allows to recover, in the Markovian framework,
path by path results. We refer e.g. to Shaposhnikov [58], who revisited the result by
Davie, and to Priola [56, 57], who extended the Davie result to the $\alpha$-stable, $\alpha$ in $[1, 2)$
and the degenerate kinetic framework, for recent works in that direction.

From the previous considerations, it is in fact not clear that the Zvonkin thresholds
derive from the PDE approach. We rather feel that they are actually related to the type
of well-posedness considered and that the difference with the “weak thresholds” is
indeed a price to pay to pass from weak to strong or path by path uniqueness.

Conclusions and extension to “pure jump” noise In view of the previous discussions,
it appears that our thresholds are sharp w.r.t. the method employed, and almost sharp
w.r.t. the existing literature (at least the one we know) on regularization by noise. In
this perspective, our result roughly says (in particular) that a Brownian type noise $\mathcal{W}$
of self similarity index $\gamma = 1/2 + n, n$ in $\mathbb{N}$ restores strong well-posedness for ODEs
whose drift has regularity index $\beta > 1 - 1/(2\gamma)$. As the range $\gamma = H$ in $(0, 1)$ is
covered by [17] and the range corresponding formally to the case $H = 0$ has been
investigated recently in [36], one may wonder if the range $\gamma = 1/\alpha + n, n$ in $\mathbb{N}$ and
$\alpha$ in $(0, 2)$ could be attainable as well.

The main point to implement the Zvonkin transform consists in establishing
parabolic bootstrap results on the underlying PDE. To this end, the idea is to expand the
associated differential operator around the generator of a suitable proxy process. This
is the so-called parametrix type expansion that will be here performed at order one.
The proxy process needs somehow to resemble the initial process to be investigated
and to be well understood. For instance, for the considered example (1.4), the nat-
ural proxy process corresponds to $\bar{X}$ (degenerate Ornstein-Uhlenbeck type Gaussian
process) introduced after (1.4). By well understood, we mean that the proxy process
admits smooth marginal densities, which together with their derivatives, satisfy appro-
priate heat kernel estimates. These estimates turn out to be crucial as our methodology
relies on duality arguments between Besov spaces, the underlying Hölder spaces being
viewed as Besov spaces. Especially, because we will use the so-called thermic char-
acterization of Besov spaces (see e.g. Chapter 2.6.4 in [62]) with an underlying heat
kernel somehow compatible with the one of the proxy. This is precisely why we feel
that the methodology provided here is robust enough to handle the case of a large
class of symmetric non-degenerate pure jump noises “$\mathcal{W} = Z^\alpha$”, $\alpha$ in $(0, 2)$, up to a
modification of the associated proxy and heat kernels, see e.g. [33, 34, 46].

We then believe that the thresholds to get strong uniqueness in the strictly stable
case would write $\beta_j$ in $((1 + \alpha(j - 3/2))/(1 + \alpha(j - 1)), 1)$ which would extend to
the whole chain the result established in the recent work [37] for the kinetic case (i.e.
when $n = 2$).
1.4 Proof of the main result: Zvonkin Transform and smoothing properties of the PDE associated with (1.1)

We emphasize that under our assumptions, it follows from [14] that (1.1) is well-posed in the weak sense. Hence, from the Yamada-Watanabe theorem it is sufficient to prove that strong (or pathwise) uniqueness holds to prove strong well-posedness. As explained above, our main strategy rests upon the Zvonkin transform initiated by Zvonkin in [69] which heavily relies on the connection between SDEs and PDEs. We rewrite (1.1) in the shortened form

\[ dX_t = F(t, X_t)dt + B\sigma(t, X_t)dW_t, \]

where we recall that \( F = (F_1, \ldots, F_n) \) is an \( \mathbb{R}^{nd} \)-valued function and \( B \) is the embedding matrix from \( \mathbb{R}^d \) into \( \mathbb{R}^{nd} \), i.e. \( B = (I_{d,d}, 0_{d,d}, \ldots, 0_{d,d})^* = (I_{d,d}, 0_{d,(n-1)d})^* \), where \( I_{d,d} \) are \( 0_{d,d} \) respectively denote the identity matrix and the matrix with zero entries in \( \mathbb{R}^d \otimes \mathbb{R}^d \). For all \( \varphi \in C_0^2(\mathbb{R}^{nd}, \mathbb{R}) \) and \( (t, x) \in [0, T] \times \mathbb{R}^{nd} \) let

\[ L_t \varphi(x) = \langle F(t, x), D\varphi(t, x) \rangle + \frac{1}{2} \text{Tr} \left( a(t, x) D_x^2 \varphi(x) \right), \tag{1.15} \]

where \( a = \sigma \sigma^* \), denote the generator associated with (1.1). We then formally associate the SDE (1.1) with the following systems of PDEs:

\[
\begin{cases}
(\partial_t u_i + L_t u_i)(t, x) = F_i(t, x), \quad (t, x) \in [0, T) \times \mathbb{R}^{nd}, \\
u_i(T, x) = 0_d, \quad i \in [1, n].
\end{cases} \tag{1.16}
\]

Remark 2 Note that above we adopted the following convention for notational convenience: for each \( i \in [3, n] \), all \( t \in [0, T] \), \( x \in \mathbb{R}^{nd} \), \( F_i(t, x) := F_i(t, x_{i-1}, \ldots, x_n) \) i.e. the independence of each map \( F_i(t, \cdot) \) with respect to the first \( i - 2 \) components of the vector \( x = (x_1, \ldots, x_n) \in \mathbb{R}^{nd} \) is implicitly assumed.

Denote by \( U = (u_1, \ldots, u_n) \) its global solution. Let now \( (F^m)_{m \geq 0}, (a^m)_{m \geq 0} \) denote two sequences of mollified coefficients satisfying assumption (A) uniformly in \( m \) that are infinitely differentiable functions with bounded derivatives of all orders greater than 1 for \( F^m \), and converging in supremum norm to \( (F, a) \) (such sequences are easily obtained from [10]). Then, for each \( m \), the regularized systems of PDEs associated with (1.16) write:

\[
\begin{cases}
(\partial_t u_i^m + L_t u_i^m)(t, x) = F_i^m(t, x), \quad (t, x) \in [0, T) \times \mathbb{R}^{nd}, \\
u_i^m(T, x) = 0_d, \quad i \in [1, n].
\end{cases} \tag{1.17}
\]

where \( L_t^m \) is obtained from (1.15) replacing \( F \) by \( F^m \) and \( a \) by \( a^m \).

The above system (1.17) is well-posed and admits a unique smooth solution \( U^m = (u_1^m, \ldots, u_n^m) \). This can be derived from the Feynman-Kac representation formula, which holds independently of the degeneracy for smooth coefficients with linear
growth, and stochastic flow techniques (see e.g. Gikhman and Skorokhod [32] or Talay and Tubaro [63]). Hence, applying Itô’s formula, one easily deduces that

\[
\int_0^t F(s, X_s) ds = -U^m(0, x) + U^m(t, X_t) - \int_0^t DU^m(s, X_s) B\sigma(s, X_s) dW_s \\
+ \int_0^t R^m_s(X_s) ds,
\]

(1.18)

where

\[
R^m_s(X_s) := [F(s, X_s) - F^m(s, X_s)] - (L_s - L^m_s)U^m(s, X_s) \\
= [F(s, X_s) - F^m(s, X_s)] - (F - F^m)(s, X_s) \cdot DU^m(s, X_s) \\
- \frac{1}{2} \text{Tr}((a - a^m)(s, X_s) D^2 U^m(s, X_s)).
\]

This representation (1.18) is the Zvonkin Transform discussed above, up to a remainder. Then, the main idea consists in taking advantage of the regularization properties of the operator \(L^m\) (uniformly in \(m\)) and expect that the solutions \(U^m, m \geq 0\) will be smoother than the source term \(F\) so that the right-hand side of (1.18) is smoother than the integrand of the left-hand side of the considered equation. In other words, we are looking for a good regularization theory for the PDE (1.17) uniformly w.r.t. the mollification argument. This good regularization theory is summarized in the following crucial result whose proof is, in fact, the main subject of this work and is postponed to Section 2.

**Theorem 2** For \(T > 0\) small enough\(^4\), there exists a constant \(C_T := CT((A)) > 0\) satisfying \(C_T \to 0\) when \(T \to 0\) such that for every \(m \geq 0\), the solution \(U^m\) satisfies, with the notation of (1.3):

\[
\|DU^m\|_\infty + \|D(D_1 U^m)\|_\infty \leq C_T.
\]

**Remark 3** (On well-posedness of the initial PDE (1.16)). We also point out that, from the uniformity in \(m\) in the previous theorem, we could also derive some regularizing properties for the system (1.16) through appropriate compactness arguments. Indeed, as it will appear in the proof of this result, we are in fact able to control uniformly the Hölder moduli of the gradients and of the second order derivatives w.r.t. the non-degenerate direction (see Lemmas 11 and 13). These controls precisely allow to derive, through the Arzelà-Ascoli theorem, a well-posedness result for equation (1.16) under the sole assumption \((A)\) as well as the above gradient estimates.

Theorem 2, which is a consequence of Theorem 3 below, is the key to prove our main result for strong uniqueness.

\(^4\) By “small enough” we mean that there exists a time \(T > 0\) depending only on known parameters in \((A)\) s.t. for any \(T \leq T\) the statement of the theorem holds.
Proof of Theorem 1  Let now $X$ and $X'$ be two solutions of (1.1). Using the representation (1.18) to express the difference of the bad drifts in terms of the function $U^m$ and its derivative up to a remainder, we write:

$$X_t - X'_t = U^m(t, X_t) - U^m(t, X'_t) - \int_0^t [DU^m(s, X_s) B\sigma(s, X_s) - DU^m(s, X'_s) B\sigma(s, X'_s)] dW_s + \int_0^t \left[\mathcal{R}_s^m(X) - \mathcal{R}_s^m(X')\right] ds + \int_0^t B\left[\sigma(s, X_s) - \sigma(s, X'_s)\right] dW_s.$$

Take then the supremum in time of the square of the difference. Passing to the expectation, a convexity inequality then leads to the following estimate:

$$E\left[\sup_{t \leq T} |X_t - X'_t|^2\right] \leq 5\left(E\left[\sup_{t \leq T} |U^m(t, X_t) - U^m(t, X'_t)|^2\right]\right) + E\left[\int_0^T ds \left[\left|DU^m B\right|(s, X_s) - \left|DU^m B\right|(s, X'_s)\right]^2 \|\sigma\|_\infty^2\right] + E\left[\int_0^T ds \left[\|DU^m B\|_\infty + 1\right] \left[\left|\sigma(s, X_s) - \sigma(s, X'_s)\right|\right]^2 + 2T \|\mathcal{R}^m(\cdot)\|_\infty^2\right].$$

Note that thanks to the particular structure of $B$ one has $DU^m B = (D_1 U^m, 0, d, \ldots, 0)\ast$. Hence, thanks to Theorem 2 and Grönwall’s lemma, there exists $\bar{C}_T := \bar{C}_T(C_T, \sigma, n, d, T)$ satisfying $\bar{C}_T \to 0$ when $T$ goes to 0 such that

$$E\left[\sup_{t \leq T} |X_t - X'_t|^2\right] \leq \bar{C}_T E\left[\sup_{t \leq T} |X_t - X'_t|^2\right] + 10T \|\mathcal{R}^m(\cdot)\|_\infty^2. \quad (1.21)$$

Letting $m \to +\infty$, since $\|a - a^m\|_\infty$ and $\|F - F^m\|_\infty$ tend to 0, it readily follows from (1.19) and the bound (1.20) in Theorem 2 that $\|\mathcal{R}^m(\cdot)\|_\infty \to 0$. Hence, choosing $T$ small enough so that $\bar{C}_T \leq 1/2$, we deduce that strong uniqueness holds on a sufficiently small time interval. Iterating this procedure in time gives the result on $\mathbb{R}^+$ from usual Markov arguments involving the regular versions of conditional expectations, see e.g. [60].

1.5 Regularization properties of the underlying PDE (1.16): strategy of the proof and primer

As mentioned above, the regularization properties of the PDE (1.17) given by estimate (1.20) in Theorem 2 are the core of this work. Smoothing properties of linear partial
differential operators of second order with non-degenerate diffusion matrix have been widely studied in the literature for bounded H"older coefficients. In that setting, the estimates of Theorem 2 are well known, see e.g. the book of Friedman [30] or Bass [3]. For unbounded drift and source terms, those estimates have been recently established in [39]. In our case, the story is rather different since the diffusion matrix \( B \alpha \) of the system is totally degenerate in the directions 2 to \( n \). However, as we already emphasized, the non-degeneracy condition assumed on the family of Jacobian matrices \((Dx_i - F_i)_i \in \{2, n\}\) allows the noise to propagate in the indicated directions thanks to the drift. It can be viewed as a weak type of H"ormander condition. Under such a condition, the operator \( L^m \) with mollified coefficients is said to be hypoelliptic\(^5\) and it is well known that hypoelliptic differential operators also have some smoothing properties (see the seminal work of H"ormander [35] or, for a probabilistic viewpoint, the monograph of Stroock [59]). The tricky point in our weak H"ormander setting is that the pointwise gradient estimates (1.20) of Theorem 2 had, to the best of our knowledge, not been established yet, even though such a setting has already been considered by several authors (see e.g. Delarue and Menozzi [22] for density estimates, Menozzi [47, 48] and Priola [55] for the martingale problem and also Bramanti et al. [5, 6] for related \( L^p \) estimates and Bramanti and Zhu [8] for the VMO framework). We can mention the work of Lorenzi [43] which gives gradient estimates in the degenerate kinetic like case \((n = 2\) in our framework) when the diffusion coefficient is sufficiently smooth and the drift linear. We point out that our main estimate in Theorem 2 needs precisely to be uniform w.r.t. the mollification parameter and therefore does not depend on the smoothness of \( F^m, a^m \), but only on known parameters appearing in (A). Again, this is what would also allow to transfer those bounds to equation (1.16) from a suitable compactness argument, extending well known results for non-degenerate diffusions with H"older coefficients to the current degenerate setting.

To prove this result our main strategy rests upon the parametrix approach see e.g. the work of McKean and Singer [51] or the book of Friedman [30]. Roughly speaking, it is a perturbative argument consisting in expanding the operator \( L^m \) around a good proxy, usually denoted by \( \tilde{L}^m \) (we keep here the super-scripts in \( m \) to emphasize that the perturbative technique we perform will concern the system (1.17) with mollified coefficients). In our setting, the term good proxy relates to the fact that the operator \( \tilde{L}^m \) is the generator of the “closest” Gaussian approximation \( \tilde{X}^m \) of \( X^m \) which has generator \( L^m \). In our case, such a process is well known and is the linearized (with respect to the source of noise) version of (1.1) whose coefficients are frozen along the curve \((\theta^m_{s,t})_{s \in [t,T]}\) that solves the deterministic counterpart of (1.1) with mollified coefficients (i.e. with \( \sigma^m \equiv 0_{d,d} \)), namely \( \dot{\theta}^m_{s,t} = F^m(s, \theta^m_{s,t}) \). This proxy process \( \tilde{X}^m \) may be seen as a (non-linear) generalization of the so-called Kolmogorov example [38] and we refer the reader to the work of Delarue and Menozzi [22] and Menozzi [47] for more explanations. Having this proxy at hand, the parametrix procedure consists in deriving the desired estimates for the proxy and control the expansion error.

In [14], Chaudru de Raynal and Menozzi successfully used this approach in its backward form to prove weak well-posedness of (1.1) under less restrictive assumptions.

\(^5\) Pay attention that this is not the case for \( L \) whose coefficients do not have the required smoothness in \((T_\beta)\) to compute the corresponding Lie brackets.
(the critical thresholds for the Hölder exponents being smaller as indicated above). In that case, the curve along which the system is frozen for the proxy is the solution of the backward deterministic counterpart of (1.1)). This backward approach is very suitable when investigating the martingale problem associated with our main system since it allows to control subtly the expansion error associating precisely the coefficients $F_i$ with their corresponding differentiation operator $D_i$ and does not require any mollification of the coefficients. Unfortunately, when trying to obtain estimates on the derivatives of the solutions of the PDE, the backward approach is not convenient since the corresponding proxy does not provide an exact density and this fact does not allow to benefit from cancellation techniques which are very helpful in this context (see paragraph below).

Hence, our parametrix approach will be of forward\(^6\) form as done in the work of Chaudru de Raynal [10]. This is, in fact, a non-trivial generalization of the approach developed in the aforementioned paper where the strong well-posedness of (1.1) is obtained when $n = 2$. Indeed, the strategy used in [10] is not adapted to this general case because of some subtle phenomena appearing only when $n \geq 3$. In particular, the singularities appearing when considering the remainder term of the parametrix were in [10] equilibrated at hand through elementary cancellation arguments, whereas the current approach takes advantage of the full-force duality results between Besov spaces (see Sects. 1.5.2 and 2.2 below). This forward perturbative approach has also been successfully used in [12] to establish some weaker regularization properties of the PDE (1.16) through appropriate Schauder estimates.

### 1.5.1 Regularizing properties of the degenerate Ornstein-Uhlenbeck proxy

When exploiting such a forward parametrix approach, a good primer to understand what could be, at best, expected, consists in investigating the regularization properties of the proxy operator $\tilde{L}$. To be as succinct as possible, let us consider the case where $(\tilde{L}_t)_{t \geq 0}$ is the generator of a degenerate Ornstein-Uhlenbeck process $(\tilde{X}_t)_{t \geq 0}$ with dynamics:

$$d\tilde{X}_t = A_t \tilde{X}_t dt + BdW_t,$$

where $A_t$ is the $nd \times nd$ matrix with sub-diagonal blocks $(a_{i,i-1}(t))_{i \in [2,n]}$ of size $d \times d$ and $0_{d,d}$ elsewhere. In particular,

$$A_t = \begin{pmatrix}
0_{d,d} & \cdots & \cdots & \cdots & 0_{d,d} \\
0_{d,d} & a_{2,1}(t) & 0_{d,d} & \cdots & 0_{d,d} \\
0_{d,d} & a_{3,2}(t) & 0_{d,d} & \cdots & 0_{d,d} \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0_{d,d} & \cdots & 0_{d,d} & a_{n,n-1}(t) & 0_{d,d}
\end{pmatrix}.
$$

The entries $(a_{i,i-1}(t))_{i \in [2,n]}$ are uniformly in time non-degenerate elements of $\mathbb{R}^d \otimes \mathbb{R}^d$ (which expresses the weak Hörmander condition). The corresponding generator

\(^6\) Meaning that the freezing curve $\theta$ solves the corresponding ODE associated with (1.1) in a forward form.
\( \tilde{L}_t \) writes for any \( \varphi \in C_0^2(\mathbb{R}^{nd}, \mathbb{R}) \):

\[
\tilde{L}_t \varphi (x) = \langle A_t x, D \varphi (x) \rangle + \frac{1}{2} \Delta x_t \varphi (x).
\]

In such a case, each component \( \tilde{u}_i, i \in \llbracket 1, n \rrbracket \) of the solution \( \tilde{U} \) of the corresponding system of PDEs

\[
\begin{cases}
(\partial_t + \tilde{L}_t) \tilde{U}(t, x) = F(t, x), \; (t, x) \in [0, T) \times \mathbb{R}^{nd}, \\
\tilde{U}(T, x) = 0_{nd},
\end{cases}
\] (1.24)

where \( F \) is a non-linear (non-mollified) source satisfying (T\(_\beta\)), writes

\[
\tilde{u}_i (t, x) = - \int_t^T ds \int_{\mathbb{R}^{nd}} dy F_i (s, y) \tilde{p}(t, s, x, y).
\] (1.25)

Above, \( \tilde{p} \) stands for the transition density of the Gaussian process \((\tilde{X}_v)_{v \geq 0}\) with dynamics (1.22). Using the resolvent associated with \((A_v)_{v \in [\tau, \infty]}\), i.e. \( \partial_s \tilde{R}_{s,t} = A_s \tilde{R}_{s,t}, \tilde{R}_{t,t} = I_{nd, nd} \), the above equation can be explicitly integrated. Precisely, for a fixed starting point \( x \) at time \( t \):

\[
\tilde{X}_v = \tilde{R}_{v,t} x + \int_t^v \tilde{R}_{v,u} B dW_u.
\] (1.26)

Hence, at time \( s > t \) the covariance matrix of the random variable \( \tilde{X}_s \) writes as \( \tilde{K}_{s,t} := \int_t^s du \tilde{R}_{s,u} B B^* \tilde{R}_{s,u}^\dagger \).

From (1.26), the density at time \( v = s \) and at the spatial point \( y \) therefore writes:

\[
\tilde{p}(t, s, x, y) = \frac{1}{(2\pi)^{nd}} \frac{1}{\det(\tilde{K}_{s,t})^{1/2}} \exp \left( -\frac{1}{2} (\tilde{K}_{s,t})^{-1} (\tilde{R}_{s,t} x - y, \tilde{R}_{s,t} x - y) \right).
\] (1.27)

Note that the resolvent also appears in (1.26) and in the density. Since the drift in (1.22) is unbounded, the term \( \tilde{R}_{s,t} x \) actually corresponds to the transport of the initial condition \( x \) through the associated deterministic and linear differential system. It is well known, see e.g. [22] and Sect. 2.1 below, that the covariance \( \tilde{K}_{s,t} \) enjoys what we will call a good scaling property. Precisely, for a given \( T > 0 \) there exists \( C := C((A_v)_{v \in [0,T]}, T) \geq 1 \) s.t. for any \( \xi \in \mathbb{R}^{nd} \):

\[
C^{-1} (s - t)^{-1} |\mathbb{T}_{s-t} \xi|^2 \leq (\tilde{K}_{s,t} \xi, \xi) \leq C (s - t)^{-1} |\mathbb{T}_{s-t} \xi|^2,
\] (1.28)

where for any \( u > 0 \), we denote by \( \mathbb{T}_u \) the intrinsic scale matrix:

\[
\mathbb{T}_u = \begin{pmatrix}
    u I_{d,d} & 0_{d,d} & \cdots & 0_{d,d} \\
    0_{d,d} & u^2 I_{d,d} & 0_{d,d} & \cdots \\
    \vdots & \vdots & \ddots & \vdots \\
    0_{d,d} & \cdots & 0_{d,d} & u^n I_{d,d}
\end{pmatrix}.
\] (1.29)
Importantly, the *good scaling property* stated in (1.28) indicates that, for a given initial time $t$ and for each $i \in \{1, n\}$, each $\mathbb{R}^d$-valued component $\bar{X}_i^\theta$ has typical fluctuations of order $(s-t)^{i-1/2}$ which correspond to those of the $(i-1)^{\text{th}}$ iterated integrals of the Brownian motion. Accordingly, we derive that the frozen density $\bar{p}$ also satisfies the bound

$$
\bar{p}(t, s, x, y) \leq \frac{C}{(s-t)^{\frac{n+1}{2}}} \exp \left( -C^{-1}(s-t)\|\bar{R}_{s,t}^\theta (\bar{R}_{s,t} x - y)\|^2 \right)
$$

for some $\bar{C} := \bar{C}(C)$ such that $\int_{\mathbb{R}^n} dy \bar{p}_{C^{-1}}(t, s, x, y) = 1$. Similarly, the derivatives of $\bar{p}$ will be bounded by a density of the form $\bar{p}_{C^{-1}}$ up to an additional multiplicative contribution reflecting the time-singularities associated with the differentiation index. Precisely, there exists $\bar{C}$ s.t. for any $l \in \{1, n\}$, $r \in \{0, 1\}$:

$$
|D_{x^l} D_{x^r} \bar{p}(t, s, x, y)| \leq \frac{C}{(s-t)^{\frac{n+1}{2}+l^{-\frac{1}{2}}}} \exp \left( -C^{-1}(s-t)\|\bar{R}_{s,t}^\theta (\bar{R}_{s,t} x - y)\|^2 \right)
$$

\[\leq \frac{\bar{C}}{(s-t)^{l^{-\frac{1}{2}}}+\frac{1}{2}} \bar{p}_{C^{-1}}(t, s, x, y),\tag{1.30}\]

up to a modification of $\bar{C}$ for the last inequality. We refer to the proof of Proposition 5 for a complete version of this statement.

To prove estimate (1.20) of Theorem 2 for the current system (1.24), it follows from the specific structure of the matrix $B$ that we have to estimate for any $l \in \{1, n\}$ the quantities $D_{x^l} D_{x^l} \bar{u}_i(t, x)$, $r \in \{0, 1\}$. From (1.25), (1.30), we thus have

$$
|D_{x^l} D_{x^r} \bar{u}_i(t, x)| \leq \bar{C} \int_t^T ds \int_{\mathbb{R}^d} dy |F_i(s, y)| (s-t)^{-\frac{1}{2}} \bar{p}_{C^{-1}}(t, s, x, y).\tag{1.31}
$$

We now face two problems: first the $F_i$ are unbounded, second the above time singularity is, as is, not integrable. Let us consider the worst case i.e. when $r = 1$. To smoothen the time singularity, the main idea consists in using the regularity of the source term $F_i$ exploiting precisely the fact that, once integrated through the variables $y_i$ to $y_n$, the transition density $\bar{p}$ does not depend on the variable $x_i$ anymore. This is due to the structure of $A$ in (1.23), which in particular yields that the resolvent $(\bar{R}_{s,t})_{0 \leq t \leq s \leq T}$ is lower triangular. Precisely, denoting for conciseness by $\theta_{s,t}(x) = \bar{R}_{s,t} x$ (which is coherent with the notation below when handling non-linear flows), we write:

$$
\int_{\mathbb{R}^{(l-1)d}} dy_{1:l-1} F_i(t, y_1, \ldots, y_{l-1}, \theta_{s,t}^1(x), \ldots, \theta_{s,t}^n(x))
\times D_{x^l} D_{x^l} \int_{\mathbb{R}^{(n-(l-1))d}} dy_{l:n} \bar{p}(t, s, x, y) = 0.
$$
This is what will be called a *cancellation (or centering) argument* in the following. When using this property, we obtain that

\[
|D_{x_1}D_{x_l}\tilde{u}_i(t, x)| = \left| \int_t^T ds \int_{\mathbb{R}^d} dy \left( F_i(s, y) - F_i(t, y_1, \ldots, y_{l-1}, \theta^l_{s,t}(x), \ldots, \theta^n_{s,t}(x)) \right) D_{x_1}D_{x_l}\tilde{p}(t, s, x, y) \right|.
\]

We thus derive from (1.30):

\[
|D_{x_1}D_{x_l}\tilde{u}_i(t, x)| \leq \tilde{C} \int_t^T ds \int_{\mathbb{R}^d} dy \left| F_i(s, y) - F_i(t, y_1, \ldots, y_{l-1}, \theta^l_{s,t}(x), \ldots, \theta^n_{s,t}(x)) \right| \times (s - t)^{-\left(l - \frac{1}{2}\right)^2} \tilde{p}_{C-1}(t, s, x, y).
\]

Then, using the regularity assumed on $F_i$, which satisfies (T$\beta$), we get that for some constants $\tilde{C}$, $C$ (which possibly change from line to line)

\[
|D_{x_1}D_{x_l}\tilde{u}_i(t, x)| \leq \tilde{C} \int_t^T ds \int_{\mathbb{R}^d} dy \sum_{j=l}^n |y_j - \theta^l_{s,t}(x)|^{\beta_j} (s - t)^{-\left(l - \frac{1}{2}\right)^2} \tilde{p}_{C-1}(t, s, x, y)
\]

\[
\leq \tilde{C} \int_t^T ds \int_{\mathbb{R}^d} dy \sum_{j=l}^n \left| \frac{y_j - \theta^l_{s,t}(x)}{(s - t)^{j - \frac{1}{2}}} \right|^{\beta_j} (s - t)^{-\left(l - \frac{1}{2}\right)^2} \tilde{p}_{C-1}(t, s, x, y)
\]

\[
\leq \tilde{C} \int_t^T ds \sum_{j=l}^n (s - t)^{\beta_j(j - \frac{1}{2}) - \left(l - \frac{1}{2}\right)^2 - \frac{1}{2}}
\]

\[
\times \int_{\mathbb{R}^d} \frac{dy}{(s - t)^{n_d}} \exp \left( -\tilde{C}^{-1}(s - t)|T_{s,t}^{-1}(\theta{s,t}(x) - y)|^2 \right)
\]

\[
\leq \tilde{C} \int_t^T ds \sum_{j=l}^n (s - t)^{\beta_j(j - \frac{1}{2}) - \left(l - \frac{1}{2}\right)^2 - \frac{1}{2}} \int_{\mathbb{R}^d} \frac{dy}{(s - t)^{n_d}} \exp \left( -\tilde{C}^{-1}(s - t)|T_{s,t}^{-1}(\theta{s,t}(x) - y)|^2 \right)
\]

\[
\leq \tilde{C} \int_t^T ds \sum_{j=l}^n (s - t)^{-\left(l - \frac{1}{2}\right)^2 + \beta_j(j - \frac{1}{2})} \int_{\mathbb{R}^d} dy \tilde{p}_{C-1}(t, s, x, y)
\]
for some additional auxiliary constants \( \tilde{C}, \hat{C} \) in the last but one inequality. Namely, we are in the scalar situation \( z^{\beta_j/2} \exp(-z) \leq \exp(-c \beta_j z), \ c \beta_j < 1 \). The polynomial part is absorbed up to a slight deterioration of the concentration constant in the exponential.

Note that the above term is integrable only if for each \( j \in [1, n] \), \(- (l - 1/2) - 1/2 + \beta_j (j - 1/2) > -1 \iff \beta_j > (2l - 2)/(2j - 1) \). This condition actually holds if for any \( i \in [1, n] \), \( \beta_i > (2l - 2)/(2i - 1) \) which is exactly the infimum assumed in \( (T_\beta) \). As we can see, there is no hope to obtain better thresholds with such a strategy. This is the reason why we said that these thresholds are almost sharp for the approach used here.

### 1.5.2 Back to the perturbative analysis

Let us now briefly explain what happens when one wants to control the approximation error in the forward parametrix expansion. We now come back to our general setting and denote by \( \tilde{p}^m \) the transition density of a suitable Gaussian proxy process \( \tilde{X}^m \) with generator \( \tilde{L}^m \). Observe that equation (1.17) can be equivalently rewritten as:

\[
\begin{cases}
(\partial_t u_i^m + \tilde{L}_i^m u_i^m)(t, x) = F_i^m(t, x) - (L_i^m - \tilde{L}_i^m)u_i^m(t, x), \ (t, x) \in [0, T) \times \mathbb{R}^{nd}, \\
u_i^m(T, x) = 0_d, \quad i \in [1, n].
\end{cases}
\]

Since \( u_i^m(t, x) \) is smooth (see the arguments following (1.17)), so is the term \( (L_i^m - \tilde{L}_i^m)u_i^m(t, x) \) which appears above as part of the source.

Hence, we derive from the above equation and the Feynman-Kac formula the following representation. For each regularized component \( u_i^m, i \in [1, n] \) of our solution \( U^m \) of the systems (1.17) it holds that for any \( (t, x) \in [0, T] \times \mathbb{R}^{nd} \)

\[
u_i^m(t, x) = \int_0^T ds \int_{\mathbb{R}^{nd}} dy \left\{ -F_i^m(s, y) + (L_i^m - \tilde{L}_i^m)u_i^m(s, y) \right\} \tilde{p}^m(t, s, x, y). \tag{1.33}
\]

This corresponds to the so-called first order parametrix expansion of the initial equation (1.17) around the proxy generator \( \tilde{L}^m \). Above, the additional term in the right-hand side is, in comparison with (1.25), precisely the approximation error due to the parametrix expansion. It thus appears that the solution has an implicit representation which makes its derivatives themselves appear. Hence, when differentiating the above representation to derive the estimate (1.20) in Theorem 2, we obtain bounds that depend themselves on the derivatives of the solution. We then have to estimate each derivative appearing in the right-hand side and use a circular argument. Namely, when differentiating \( u_i^m(t, x) \), we will obtain the required estimate provided the multiplicative constants associated with the terms \( \|D u_i^m\|_\infty \) and \( \|D_i D u_i^m\|_\infty \), that will appear in the corresponding upper-bound for the above right-hand side, are small enough (see also Section 2 of [10] and Sect. 2.2 below for details).
Moreover, as we have already seen, in order to smoothen the time singularity appearing when we apply a cross differentiation operator in the $l^{th}$ and $1^{st}$ direction to the term $\int_{t-r}^{T} ds \int_{\mathbb{R}^d} dy (L_{s}^{m} - \tilde{L}_{s}^{m}) u_{i}^{m}(s, y) \tilde{p}^{m}(t, s, x, y)$ corresponding to the approximation error, we will have to center this term around the derivatives of the solution itself (in the sense given in the above discussion). This procedure allows us, thanks to Taylor expansions, to weaken the singularities and provides integrable (in time) terms. The dramatic point is that, when doing so, our bound involves the cross derivatives $D_{\ell}D_{j}u_{m,i}, \ell, j \in \{1, \ldots, n\}$ whose control in supremum norm is, as suggested by the discussion done in the explicit case of a simple degenerate Ornstein-Uhlenbeck process, definitely out of reach as soon as $\ell > 1$. In fact, as suggested by the results in [10], the only thing we could hope is that the gradient in the degenerate directions viewed as a function of the degenerate variables, i.e. $D_{2:n}u_{i}^{m}(t, x_{1}, \cdot) := \left( D_{2}u_{i}^{m}(t, x_{1}, \cdot), \ldots, D_{n}u_{i}^{m}(t, x_{1}, \cdot) \right)^{*}$ for any $(t, x_{1}) \in [0, T] \times \mathbb{R}^d$, belongs to an appropriate anisotropic Hölder space. Note importantly that such spaces can as well be viewed as particular cases of anisotropic Besov spaces with corresponding positive regularity indexes. The main idea is thus to use an integration by parts argument in order to rebalance one differentiation operator from the solution to the remaining terms coming from the differentiation of (1.33), which in particular involve the coefficients of the operator $L_{m} - \tilde{L}_{m}$ and contain the time singularities coming from the derivatives of the frozen Gaussian kernel $\tilde{p}^{m}$. As the coefficients of the operator $L_{m} - \tilde{L}_{m}$ are assumed to be only Hölder continuous, their generalized derivative should belong to some anisotropic Besov spaces of negative regularity index, strictly bigger than $-1$. To tackle this problem, our main idea, in order to balance the lack of differentiation property of the remaining terms, consists in putting precisely in duality the gradient $D_{2:n}u_{i}^{m}(t, x_{1}, \cdot)$, belonging to an anisotropic inhomogeneous Besov space with positive regularity exponent and the remaining terms, belonging to an anisotropic inhomogeneous Besov space with negative regularity exponent.

We refer to the proof of the main Theorem 2 in Sect. 2 for details and to Proposition 3.6 in the book of Lemarié-Rieusset [44] for duality results on Besov spaces. We are thus led to control on the one hand the Besov norm with positive exponent (or equivalently the Hölder norm) of the derivatives of the solution, see Lemma 11, and on the other hand the Besov norm with negative exponent of the remaining terms in (1.33) (involving the coefficients of the operator $L_{m} - \tilde{L}_{m}$), see Lemma 10. The first control (Hölder norm) is crucial and appears to be quite delicate. Indeed, due to the implicit representation (1.33), this estimate also involves Hölder norms of the full gradient $Du_{i}^{m}$. This again reflects the circular nature of the arguments needed to derive the result.

Let us close this discussion coming back to Remark 1. As we emphasized, in comparison with the non-degenerate result, Theorem 1 should hold assuming that the drift $F_{1}$ belongs to a suitable $L_q - L_\rho$ space w.r.t. time and the non-degenerate variable $x_{1}$. We are convinced that this is the case but we deliberately decide not to tackle this setting in order to keep this work shorter and more coherent. Indeed, in this case, the difficulty comes from the estimate on the second order derivative in the non-degenerate direction of the first component of the solution $U_{i}^{m}$, namely $D_{x_{1}}D_{x_{1}}u_{i}^{m}$ (which is a part of the main estimate (1.20) in Theorem 2). The point is to establish for this quantity an
Strong regularization by Brownian noise for a chain of ODEs

$L_q - L_p$ control. This cannot be derived from the previously described approach and requires harmonic analysis techniques (see also [42]). The main problem to establish the estimate is mainly due to the source term, which is actually $F_1$. To prove it, the main idea consists in exploiting the results of Menozzi [48] (where such an estimate is proved under the assumption that the drift is Lipschitz) through the tools developed in [14] (backward parametrix approach for drift $F$ whose first component may be in $L_q - L_p$ and the other ones in Hölder spaces). Then, the Zvonkin transform should also be tuned a little bit following the strategy developed by Veretennikov (see e.g. [64] and [25]). Such a program would surely toughen our paper without adding any surprising result and we prefer to focus on the novelty of the approach based on duality results for Besov spaces and the generalization of the strong uniqueness result to the whole chain (i.e. to any arbitrary $n \geq 1$) rather than drowning the reader into additional technical considerations.

Let us also say that the perturbative approach we develop in this work could a priori be used, associated with the techniques of [50], which addresses the non degenerate case for unbounded drifts, to derive Gaussian density bounds similar to (1.30) (see also (2.15) below) for the derivatives of the density of the process $X$ solving (1.1), i.e. not only for the proxy process, under some minimal smoothness assumptions on the coefficients. This will concern further research. Let us indicate that, when the coefficients in (1.1) are smooth such type of controls for the derivatives of the density have been obtained by Pigato [53] through Malliavin calculus techniques.

2 Perturbation techniques for the PDE: proof of Theorem 2

In order to keep the notations as clear as possible, we forget the superscript $m$ standing for the mollifying procedure and we suppose that the following assumptions hold:

Assumption (AM). We say that assumption (AM) holds if the assumptions gathered in (A) hold true and the coefficients $F$, $a$ are infinitely differentiable functions with bounded derivatives of all orders for $a$ and greater than 1 for the coefficient $F$.

About the constants. We importantly specify that all the constants appearing below do not depend on the (omitted) smoothing parameter $m$, but only on known parameters from assumption (A) and possibly on $T > 0$.

In the whole section, we consider a fixed final time $T > 0$ which is meant to be small, i.e. $T \ll 1$. Let us consider for this section a generic PDE with generator corresponding to (1.15) and scalar source $f$ having the same Hölder regularity than the drift terms in (1.1) (i.e. the scalar function $f$ below can be any of the entries of the $\mathbb{R}^d$-valued $(F_i)_{i \in \{1,n\}}$ in the dynamics (1.1)). Namely, we concentrate on

\[
\begin{cases}
(\partial_t u + L_t u)(t, x) = -f(t, x), & (t, x) \in [0, T) \times \mathbb{R}^{nd}, \\
u(T, x) = 0,
\end{cases}
\tag{2.1}
\]

where $(L_t)_{t \geq 0}$ is defined in (1.15) and stands for the generator associated with (1.1) when the coefficients are smooth.
The key result to prove strong uniqueness for the SDE (1.1) is actually the following theorem.

**Theorem 3** (Pointwise bounds for the derivatives of the PDE (2.1)) There exists $\gamma := \gamma((A)) > 0$ and $C := C((A)) > 0$ s.t.

$$\|Du\|_\infty + \|D(D_1u)\|_\infty \leq CT^\gamma,$$  \hspace{1cm} (2.2)

with obvious extension of the definition in (1.3) to the current scalar case.

The proof of Theorem 3 is performed in Sect. 2.2 through the forward parametrix approach consisting in considering a suitable proxy semi-group around which the initial solution of (2.1) can be expanded. To this end we first investigate in Sect. 2.1 below the linearized Gaussian process deriving from the dynamics in (1.1) which will provide the suitable model for the parametrix.

**Proof of Theorem 2** Theorem 2 is then a direct consequence of Theorem 3.

Precisely, recalling that $U = (u_1, \ldots, u_n)$ where for $i \in [1, n]$, $u_i$ solves (1.16), we derive that $u^j_i$, $j \in [1, d]$ solves (2.1) with $f(t, x) = -F^j_i(t, x)$. Theorem 2 thus follows in whole generality applying Theorem 3 to each component of the solution of the systems (1.16).

\[\square\]

**2.1 Gaussian proxy and associated controls**

**2.1.1 Linearization of the dynamics**

Fix some freezing points $(\tau, \xi) \in [0, T] \times \mathbb{R}^n$. For fixed initial conditions $(t, x) \in [0, T] \times \mathbb{R}^n$, a natural linearization associated with the mollified version of (1.1) writes

$$d\tilde{X}^{(\tau, \xi)}_v = [F(v, \theta_{v, \tau}(\xi)) + D\! F(v, \theta_{v, \tau}(\xi))(\tilde{X}^{(\tau, \xi)}_v - \theta_{v, \tau}(\xi))]dv$$

$$+ B\sigma(v, \theta_{v, \tau}(\xi))dW_v, \forall v \in [t, s],$$

$$\tilde{X}^{(\tau, \xi)}_t = x,$$  \hspace{1cm} (2.3)

where

$$\dot{\theta}_{v, \tau}(\xi) = F(v, \theta_{v, \tau}(\xi)), \quad v \in [0, T], \quad \theta_{\tau, \tau}(\xi) = \xi,$$  \hspace{1cm} (2.4)

and $D\! F(v, \cdot)$ denotes the subdiagonal of the Jacobian matrix $DF(v, \cdot)$. Namely, for $z \in \mathbb{R}^n$:

$$DF(v, z) = \begin{pmatrix} 0_{d,d} & \cdots & \cdots & \cdots & 0_{d,d} \\ Dz_1F_2(v, z) & 0_{d,d} & \cdots & \cdots & 0_{d,d} \\ \vdots & 0_{d,d} & \ddots & \cdots & \vdots \\ 0_{d,d} & \cdots & 0_{d,d} & Dz_{n-1}F_n(v, z_{n-1}, z_n) & 0_{d,d} \end{pmatrix}. $$

\[\square\] Springer
In the following, we will often refer to the Gaussian process \( \tilde{X}_v^{(t,x)} \) introduced in (2.3) as the proxy process. This terminology comes from the fact that it is a natural, well controlled object, meant to locally approximate the original dynamics in (1.1).

We explicitly integrate (2.3) to obtain for any \( v \in [t, s] \):

\[
\tilde{X}_v^{(t,x)}(t, x) = \tilde{R}^{(t,x)}(v, u) F(u, \theta_{u,\tau}(\xi)) - DF(u, \theta_{u,\tau}(\xi)) \theta_{u,\tau}(\xi)
+ \int_t^u \tilde{R}^{(t,x)}(v, u) B \sigma(u, \theta_{u,\tau}(\xi)) dW_u,
\]

where \( \tilde{R}^{(t,x)}(v, u) \) stands for the resolvent associated with the partial gradients \( DF(v, \theta_{u,\tau}(\xi)) \) which satisfies for \( v \in [t, s] \):

\[
\partial_v \tilde{R}^{(t,x)}(v, t) = DF(v, \theta_{v,\tau}(\xi)) \tilde{R}^{(t,x)}(v, t), \quad \tilde{R}^{(t,x)}(t, t) = I_{n \times n}.
\]

Note in particular that since the partial gradients are subdiagonal \( \det(\tilde{R}^{(t,x)}(v, t)) = 1 \).

Also, for \( v \in [t, s] \), we recall that \( m_{v,t}^{(t,x)}(x) \) stands for the mean of \( \tilde{X}_v^{(t,x)} \) and corresponds as well to the solution of (2.3) when \( \sigma = 0 \) and the starting point is \( x \).

Importantly, we point out that \( x \in \mathbb{R}^{nd} \mapsto m_{v,t}^{(t,x)}(x) \) is affine w.r.t. the starting point \( x \). Precisely, for \( x, x' \in \mathbb{R}^{nd} \):

\[
m_{v,t}^{(t,x)}(x + x') = \tilde{R}^{(t,x)}(v, t)x' + m_{v,t}^{(t,x)}(x).
\]

It is also useful to note that, since from (2.3)

\[
dm_{v,t}^{(t,x)}(x) = [F(v, \theta_{v,\tau}(\xi)) + DF(v, \theta_{v,\tau}(\xi))(m_{v,t}^{(t,x)}(x) - \theta_{v,\tau}(\xi))] dv,
\]

it holds from (2.4) that for \( (t, x) = (t, x) \):

\[
d\left(m_{v,t}^{(t,x)}(x) - \theta_{v,t}(x)\right)
= [F(v, \theta_{v,t}(x)) + DF(v, \theta_{v,t}(x))(m_{v,t}^{(t,x)}(x) - \theta_{v,t}(x)) - \theta_{v,t}(x))] dv
= [DF(v, \theta_{v,t}(x))(m_{v,t}^{(t,x)}(x) - \theta_{v,t}(x))] dv,
\]

\[
m_{v,t}^{(t,x)}(x) - \theta_{v,t}(x) = 0.
\]

Therefore, the Grönwall lemma yields:

\[
m_{v,t}^{(t,x)}(x)|_{(t,x)} = m_{v,t}^{(t,x)}(x) = \theta_{v,t}(x).
\]
Namely, when freezing the parameters at the initial time $t$ and the starting spatial point $x$, the linearized flow $m^{(t,x)}(x)$ and the non-linear one $\theta_{v,t}(x)$ coincide.

From the non-degeneracy of $\sigma$ and the Hörmander like condition, the Gaussian process defined by (2.5) admits a density $\tilde{p}^{(t,\xi)}(t, s, x, \cdot)$ which is suitably controlled (see Proposition 5 below and for instance [22], [14]).

We first give in the next proposition a key estimate on the covariance matrix associated with (2.5) and its properties w.r.t. a suitable scaling of the system.

**Proposition 4** (Good Scaling Properties of the Covariance Matrix) The covariance matrix of $\tilde{X}^{(t,\xi)}_v$ in (2.5) writes:

$$\tilde{K}^{(t,\xi)}_{v,t} := \int_t^v du \tilde{R}^{(t,\xi)}(v, u) Ba(u, \theta_{u,\tau}(\xi)) B^* \tilde{R}^{(t,\xi)}(v, u)^*.$$ 

Uniformly in $(\tau, \xi) \in [0, T] \times \mathbb{R}^{nd}$ and $s \in [0, T]$, it satisfies a good scaling property in the sense of Definition 3.2 in [22] (see also Proposition 3.4 of that reference). That is, for any fixed $T > 0$, there exists $C_{2.9} := C_{2.9}((A), T) \geq 1$ s.t. for all $0 \leq t < v \leq s \leq T$, for any $(\tau, \xi) \in [0, T] \times \mathbb{R}^{nd}$:

$$\forall \xi \in \mathbb{R}^{nd}, \quad C_{2.9}^{-1} (v-t)^{-1} |\mathbb{T}_{v-t} \xi|^2 \leq \langle \tilde{K}^{(t,\xi)}_{v,t} \xi, \xi \rangle \leq C_{2.9} (v-t)^{-1} |\mathbb{T}_{v-t} \xi|^2,$$ (2.9)

where we again use the notation introduced in (1.29) for the scaling matrix $\mathbb{T}_{v-t}$.

Under (A), Proposition 4 readily follows from Proposition 3.4 in [22] (see also the scaling Lemma 3.6 therein). A complete proof is provided in Appendix C.1.1 below.

**Remark 4** (On some important consequences of the good scaling property) We state here some rather direct yet important controls that follow from Proposition 4.

1. Setting for any $s \in (t, T]$,

$$\tilde{K}^{(t,\xi),s,t} := (s-t) \mathbb{T}_{s-t}^{-1} \tilde{K}^{(t,\xi)}_{s,t} \mathbb{T}_{s-t}^{-1},$$ (2.10)

it follows from (2.9) that for any $\xi \in \mathbb{R}^{nd}$,

$$C_{2.9}^{-1} |\xi|^2 \leq \langle \tilde{K}^{(t,\xi),s,t} \xi, \xi \rangle \leq C_{2.9} |\xi|^2 \Longrightarrow \exists \tilde{C} := \tilde{C}((A), T) \geq 1,$$

$$\tilde{C}^{-1} \leq \det(\tilde{K}^{(t,\xi),s,t}) \leq \tilde{C}.$$ (2.11)

2. Equations (2.10) and (2.11) then readily give that:

$$\tilde{C}^{-1} (s-t)^{nd} \leq \det(\tilde{K}^{(t,\xi)}_{s,t}) \leq \tilde{C} (s-t)^{nd},$$ (2.12)

as well as, for any $\xi \in \mathbb{R}^{nd}$,
\[ C_{2,9}^{-1} |\xi|^2 \leq \langle (\tilde{K}^{(s,t)}_{x})^{-1} \xi, \xi \rangle \leq C_{2,9} |\xi|^2 \Rightarrow \]
\[ C_{2,9}^{-1} (s-t) |T_{s-t}^{-1} \xi|^2 \leq \langle (\tilde{K}^{(s,t)}_{x})^{-1} \xi, \xi \rangle \leq C_{2,9} (s-t) |T_{s-t}^{-1} \xi|^2. \]

This last bound will be of particular relevance to control the tails of the Gaussian density of \( \tilde{X}^{(s,t)}_{x} \).

We now state some important density bounds for the linearized model.

**Proposition 5** (Density of the linearized dynamics) Under (A), we have that, for any \( s \in (t, T] \) the random variable \( \tilde{X}^{(s,t)}_{x} \) in (2.5) admits a Gaussian density \( \tilde{p}^{(s,t)}(t, s, x, \cdot) \) which writes for any \( y \in \mathbb{R}^{nd} \):

\[ \tilde{p}^{(s,t)}(t, s, x, y) := \frac{1}{(2\pi)^{\frac{nd}{2}} \det(\tilde{K}^{(s,t)}_{x})^\frac{1}{2}} \exp \left( -\frac{1}{2} \langle (\tilde{K}^{(s,t)}_{x})^{-1} (m^{(s,t)}_{x}(x) - y), m^{(s,t)}_{x}(x) - y \rangle \right), \]

with \( \tilde{K}^{(s,t)}_{x} \) as in Proposition 4.

Also, there exist \( C := C((A), T) > 0 \) and \( \tilde{C} := \tilde{C}(C) \) s.t. for all \( l \in \llbracket 1, n \rrbracket, q, r \in \{0, 1\}, j \in \llbracket 1, n \rrbracket, m \in \{0, 1\} \), we have:

\[ |D_{x}^{q} D_{y}^{r} m^{(s,t)}_{x}(t, s, x, y)| \leq \frac{C}{(s-t)^{\frac{3}{2}q + \frac{3}{2}r + \frac{3}{2}j + \frac{3}{2}m}} \exp \left( -C^{-1} (s-t) |T_{s-t}^{-1} (m^{(s,t)}_{x}(x) - y)|^2 \right) \]
\[ =: \tilde{C}(s-t)^{(l-\frac{1}{2})q + (l-\frac{1}{2})r + (l-\frac{1}{2})m} \tilde{p}^{(s,t)}(t, s, x, y). \]

**Proof** Expression (2.14) readily follows from (2.5). The control (2.15) is then a direct consequence of Proposition 4 for \( q = r = m = 0 \) (upper bound of the density).

Differentiating w.r.t. \( x \) recalling from (2.7) that \( x \mapsto m^{(s,t)}_{x}(x) \) is affine yields:

\[ D_{x} \tilde{p}^{(s,t)}(t, s, x, y) = \left[ (\tilde{K}^{(s,t)}_{x})^{-1} \right]^{*} \left[ \tilde{K}^{(s,t)}_{x} \right] \left[ (m^{(s,t)}_{x}(x) - y) \right] \tilde{p}^{(s,t)}(t, s, x, y), \]

where we denoted here for a vector \( z = (z_1, \ldots, z_n) \in \mathbb{R}^{nd}, |z|_j = z_j \) for readability. The point is now to use scaling arguments. We can first rewrite with the notations of (2.10):

\[ \left[ \tilde{R}^{(s,t)}_{x} \right]^{*} \left[ (\tilde{K}^{(s,t)}_{x})^{-1} \right] \left[ \tilde{R}^{(s,t)}_{x} \right] = (s-t) \left[ (\tilde{R}^{(s,t)}_{x}) \right] \left[ T_{s-t}^{-1} \right] \left[ (\tilde{K}^{(s,t)}_{x,s})^{-1} \right] \left[ T_{s-t}^{-1} \right]. \]
Introduce now the process $\hat{R}$ by:

$$\left[\hat{R}^{(\tau, \xi), s, t}\right]^* = T_{s-t}^{-1} \left[\hat{R}^{(\tau, \xi), s, t}\right]^* T_{s-t}. \quad (2.18)$$

It follows from equation (C.4) in the appendix that there exists $\hat{C}_1 > 0$ s.t.

$$|\left[\hat{R}^{(\tau, \xi), s, t}\right]^* \xi| \leq \hat{C}_1 |\xi|, \text{ for any } \xi \in \mathbb{R}^{nd}.$$

It also follows from (2.11) and (2.13) that there exists $C > 0$ s.t.

$$|\left(K^{(\tau, \xi), s, t}\right)^{-1} \xi| \leq \hat{C}_1 |\xi|, \text{ for any } \xi \in \mathbb{R}^{nd}.$$

Equations (2.16), (2.17) and (2.18) therefore yield:

$$|D_{x_j} \tilde{p}^{(\tau, \xi)}(t, s, x, y)|
\leq (s-t)^{-j + \frac{1}{2}} \left|\left[\hat{R}^{(\tau, \xi), s, t}\right]^* \left(K^{(\tau, \xi), s, t}\right)^{-1} \left((s-t)^{\frac{1}{2}} T_{s-t}^{-1} (m^{(\tau, \xi)}_{s, t}(x) - y)\right)\right| \times \hat{p}^{(\tau, \xi)}(t, s, x, y)
\leq C (s-t)^{-j + \frac{1}{2}} (s-t)^{\frac{1}{2}} |T_{s-t}^{-1} (m^{(\tau, \xi)}_{s, t}(x) - y)| \tilde{p}^{(\tau, \xi)}(t, s, x, y).$$

From the explicit expression (2.14), Proposition 4 and the above equation, we eventually derive:

$$|D_{x_j} \tilde{p}^{(\tau, \xi)}(t, s, x, y)|
\leq \frac{C}{(s-t)^{j - \frac{1}{2}} \left|T_{s-t}^{-1} (m^{(\tau, \xi)}_{s, t}(x) - y)\right|} \frac{1}{(s-t)^{\frac{n_d}{2}}}
\exp \left(-C^{-1} (s-t) |T_{s-t}^{-1} (m^{(\tau, \xi)}_{s, t}(x) - y)|^2\right)
\leq \frac{\hat{C}}{(s-t)^{j - \frac{1}{2}} \tilde{p}^{(\tau, \xi)}(t, s, x, y)},$$

which gives the statement for one partial derivative. The controls for the derivative w.r.t. $y_j, j \in \{1, n\}$ and for the higher order derivatives w.r.t. $x$ are obtained similarly (see e.g. the proof of Lemma 5.5 of [22] for the bounds on $D_{x_1}^2 \tilde{p}^{(\tau, \xi)}(t, s, x, y)$.)

**Remark 5** When the freezing couple $(\tau, \xi)$ corresponds to the couple $(t, x)$, where $t$ is the starting time and $x$ the starting position of $X^{(\tau, \xi)}$, we importantly derive from (2.8) that (2.15) can be specified as follows:

$$|D_{x_1}^q D_{y_j}^r D_{x_1}^m \tilde{p}^{(\tau, \xi)}(t, s, x, y)|_{(\tau, \xi) = (t, x)} \leq \frac{C}{(s-t)^{\frac{n_d}{2} + (q-\frac{1}{2})z + (r-\frac{1}{2})z + (j-\frac{1}{2})m}}
\times \exp \left(-C^{-1} (s-t) |T_{s-t}^{-1} (\theta_{s,t}(x) - y)|^2\right).$$

\(\square\) Springer
Now, let us state a useful control involving the previous Gaussian kernel which will be exploited in some cancellation techniques.

**Proposition 6** For all $k \in \mathbb{[1, n]}$, $0 \leq t \leq s \leq T$, $(x, \xi) \in \mathbb{R}^{nd} \times \mathbb{R}^{nd}$, and $z \in \mathbb{R}^{d}$ the following identity holds:

$$
\int_{\mathbb{R}^{nd}} dy D_{x_k} \tilde{p}^{(r, \xi)}(t, s, x, y)\left(z, (y - m_{s,t}^{(r, \xi)}(x))_k\right) = z. \quad (2.20)
$$

**Proof** From Proposition 5, we have $\int_{\mathbb{R}^{nd}} \tilde{p}^{(r, \xi)}(t, s, x, y)(y - m_{s,t}^{(r, \xi)}(x))_k dy = 0_d$. Differentiating this expression w.r.t. $x_k$ and using the Leibniz formula (recalling as well the identity (2.7) which yields $D_{x_k}[m_{s,t}^{(r, \xi)}(x)]_k = (\tilde{R}^{(r, \xi)}(s, t))_{k,k} = I_{d,d}$) gives (2.20). $\square$

### 2.1.2 Density and associated inhomogeneous semi-group: regularization properties

For fixed $(t, x) \in [0, T] \times \mathbb{R}^{nd}$, we give in this section some important properties concerning the regularization effects in time of the density $\tilde{p}^{(r, \xi)}(t, s, x, \cdot)$, $s \in (t, T]$ and the associated semi-group. From now on, for notational simplicity, we will write with a slight abuse of notation $\tilde{p}^{\xi}(t, s, x, y) := \tilde{p}^{(r, \xi)}(t, s, x, y)$, i.e. we omit the freezing parameter in time when it corresponds to the considered starting time.

**Density.** The following result will be thoroughly used to derive some key bounds of the sensitivity analysis performed in Sect. 2.2.

**Lemma 7** (Regularization effects for the frozen density). There exists $C := C((A), T)$ s.t. for all $\gamma_1, \gamma_2 \in (0, 2], \ell \in \{0, 1\}$, $l \in \mathbb{[1, n]}$, $j, k \in \mathbb{[1, n]}$,

$$
\int_{\mathbb{R}^{nd}} dy \left| D_{x_k} D_{x_1}^{\ell} \tilde{p}^{\xi}(s, t, x, y) \bigg|_{\xi = x} \right| \left| (y - \theta_{s,t}(x))_j \right|^{\gamma_1} \left| (y - \theta_{s,t}(x))_k \right|^{\gamma_2} \leq C(s - t)^{-\frac{\ell}{2} - \frac{k}{2} - \frac{l}{2} + \gamma_1 \left( j - \frac{1}{2} \right) + \gamma_2 \left( k - \frac{1}{2} \right)}.
$$

**Proof** Equation (2.21) actually follows from (2.19) observing that:

$$
\int_{\mathbb{R}^{nd}} dy \left| D_{x_k} D_{x_1}^{\ell} \tilde{p}^{\xi}(s, t, x, y) \bigg|_{\xi = x} \right| \left| (y - \theta_{s,t}(x))_j \right|^{\gamma_1} \left| (y - \theta_{s,t}(x))_k \right|^{\gamma_2} \leq \int_{\mathbb{R}^{nd}} dy \frac{C}{(s - t)^{\frac{n^2}{2} + (l - \frac{1}{2}) + \frac{\ell}{2}}} \exp\left(-C^{-1}(s - t)\left| \Pi_{s,t}^{l-1}(\theta_{s,t}(x) - y) \right|^{\gamma_1} \left| (\theta_{s,t}(x) - y)_j \right|^{\gamma_2} \left( s - t \right)^{\gamma_1 \left( j - \frac{1}{2} \right) + \gamma_2 \left( k - \frac{1}{2} \right)} \right)
$$

$$
\times \left( \frac{\left| (\theta_{s,t}(x) - y)_j \right|^{\gamma_2} \left( s - t \right)^{\gamma_1 \left( j - \frac{1}{2} \right)}}{(s - t)^{j - \frac{1}{2}}} \right)^{\gamma_1} \left( \frac{\left| (\theta_{s,t}(x) - y)_k \right|^{\gamma_2} \left( s - t \right)^{\gamma_1 \left( k - \frac{1}{2} \right)}}{(s - t)^{k - \frac{1}{2}}} \right)^{\gamma_2}.
$$

\( \square \) Springer
where we normalized the \( j \)th and \( k \)th entries at their corresponding time scales in order to absorb them in the off-diagonal bound of the density. Namely, since:

\[
\left( \frac{|(\theta_{s,t}(x) - y)_j|}{(s-t)^{j-\frac{1}{2}}} \right)^{\gamma_1} \left( \frac{|(\theta_{s,t}(x) - y)_k|}{(s-t)^{k-\frac{1}{2}}} \right)^{\gamma_2} \leq |(s-t)^{\frac{1}{2}}(\mathbb{T}_{s-t}^{-1}(\theta_{s,t}(x)) - y)|^{\gamma_1 + \gamma_2},
\]

we eventually derive from (2.22) that up to a modification of \( C \), which can be chosen uniformly with respect to \( \gamma_1, \gamma_2 \in (0, 2] \):

\[
\int_{\mathbb{R}^{nd}} dy \left| D_{\xi}^l D_{x_1}^k \tilde{P}^\xi(s, t, x, y) \right|_{\xi=x} \left| (y - \theta_{s,t}(x))_j \right|^{\gamma_1} \left| (y - \theta_{s,t}(x))_k \right|^{\gamma_2} \leq \int_{\mathbb{R}^{nd}} dy \frac{C}{(s-t)^{\frac{n^2 d}{2} + (l-\frac{1}{2}) + \frac{1}{2}}} \exp \left( -C^{-1}(s-t) |\mathbb{T}_{s-t}^{-1}(\theta_{s,t}(x) - y)|^2 \right) \\
\times (s-t)^{\gamma_1(j-\frac{1}{2}) + \gamma_2(k-\frac{1}{2})} \leq C(s-t)^{-\frac{n}{2} - \frac{1}{2}}(s-t)^{\gamma_1(j-\frac{1}{2}) + \gamma_2(k-\frac{1}{2})},
\]

which precisely gives (2.21).

\[\square\]

**Semi-group.** Fix \( t \in [0, T] \), \( \xi \in \mathbb{R}^{nd} \). With the notations of the previous paragraph, we introduce the following inhomogeneous semi-group associated with (2.3) for \( \tau = t \). Namely, for all \( s \in (t, T] \), \( g \in B_{\text{lin}}(\mathbb{R}^{nd}, \mathbb{R}) \) (space of measurable functions with linear growth), \( x \in \mathbb{R}^{nd} \):

\[
\tilde{P}_{s,t}^\xi g(x) := \int_{\mathbb{R}^{nd}} dy \tilde{P}^{(t, \xi)}(s, t, x, y) g(y).
\]

One can derive from Proposition 5 the following important centering and regularization result.

**Lemma 8** (Centering and Regularization effects for the inhomogeneous semi-group). Let \( f : \mathbb{R}^{nd} \to \mathbb{R} \) be a \( \vartheta \)-Hölder continuous functions where \( \vartheta := (\vartheta_1, \ldots, \vartheta_n) \in (0, 1]^n \) is a multi-index and for \( i \in [1, n] \), \( \vartheta_i \) stands for the Hölder regularity of \( f \) in the variable \( x_i \).

- **Centering arguments.** For all \( l \in [1, n] \), \( k \leq l \), \( 0 \leq t < s \leq T \), \( x, \xi \in \mathbb{R}^{nd} \) and any \( a \in \mathbb{R}^{nd} \), it holds that:

\[
D_{x_l} \tilde{P}_{s,t}^\xi \left( f(\cdot, k-1, a_{k:n}) \right)(x) = 0.
\]

(2.24)

- As particular cases of the previous property, we have that there exists \( C := C((A), T) \) such that for all \( l \in [1, n] \), \( x, \xi \in \mathbb{R}^{nd} \):

\[
|D_{x_l} D_{x_l} \tilde{P}_{s,t}^\xi f(x)| \leq C \sum_{j=l}^n |f_j(s, \cdot)|_{\vartheta_j} (s-t)^{-l+\vartheta_j(j-\frac{1}{2})}.
\]
\[ |D_{x_l} \tilde{P}^\xi_{s,t} f(x)| \leq C \sum_{j=l}^{n} |f_j(s, \cdot)| \theta_j (s-t)^{-\left(\frac{l-1}{2}\right)} + \theta_j (j-\frac{1}{2}), \tag{2.25} \]

with the notations introduced in (1.2).

**Proof** Centering arguments like (2.24) will be a crucial tool in the analysis below. To justify such an identity, write:

\[
\tilde{P}^\xi_{s,t}\left( f(\cdot, k-1, a_{k:n}) \right)(x) = \int_{\mathbb{R}^nd} dy \tilde{p}^\xi_{s,t}(t, s, x, y) f(y_{1:k-1}, a_{k:n}) \\
= \int_{\mathbb{R}^nd} dy \tilde{p}^\xi_{s,t}(t, s, x, y + m_{s,t}^{(t, \xi)}(x)) f(y_{1:k-1} + m_{s,t}^{(t, \xi)}(x))_{1:k-1}, a_{k:n}).
\]

Note now from (2.14) that \( \tilde{p}^\xi_{s,t}(t, s, x, y) \) does not depend on \( x \). Observe as well that \( (m_{s,t}^{(t, \xi)}(x))_{1:k-1} \) does not depend on \( x_l \). Thus, the right-hand side of the above equality does not depend on \( x_l \), which implies (2.24).

Let us now prove (2.25). The idea is to use first a centering argument w.r.t. the variables \( l \) to \( n \). Namely, from (2.24) and Proposition 5, it holds that for any \( a \in \mathbb{R}^{nd} \),

\[
\left| D_{x_l} D_{x_l} \left[ \tilde{P}^\xi_{s,t} f(s, \cdot) \right](x) \right| = \left| D_{x_l} D_{x_l} \left[ \tilde{P}^\xi_{s,t} \left( f(s, \cdot) - f(s, \cdot, l-1, a_{l:n}) \right) \right](x) \right| \\
\leq \tilde{C} (s-t)^{-\left(\frac{l-1}{2}\right)-1/2} \int_{\mathbb{R}^nd} dy \tilde{p}^\xi_{s,t}(t, s, y) \left| f(s, y) - f(s, y_{1:l-1}, a_{l:n}) \right| \\
\leq \tilde{C} (s-t)^{-\left(\frac{l-1}{2}\right)-1/2} \sum_{j=l}^{n} |f_j(s, \cdot)| \theta_j \int_{\mathbb{R}^nd} dy \tilde{p}^\xi_{s,t-1}(t, s, y) \left| y_j - a_j \right| \theta_j.
\]

Taking now \( a = m_{s,t}^{(t, \xi)}(x) \), we derive from (2.15) that:

\[
\left| D_{x_l} D_{x_l} \left[ \tilde{P}^\xi_{s,t} f(s, \cdot) \right](x) \right| \leq \tilde{C} (s-t)^{-\left(\frac{l-1}{2}\right)-1/2} \sum_{j=l}^{n} |f_j(s, \cdot)| \theta_j \\
\times \int_{\mathbb{R}^nd} dy \tilde{p}^\xi_{s,t}(t, s, y) \left| y - m_{s,t}^{(t, \xi)}(x) \right| \theta_j \\
\leq C \sum_{j=l}^{n} |f_j(s, \cdot)| \theta_j (s-t)^{-l+\theta_j (j-\frac{1}{2})},
\]

which gives the first bound in (2.25). The control for \( |D_{x_l} \tilde{P}^\xi_{s,t} f(x)| \) is derived similarly. \( \square \)

We state in the lemma below a useful control to obtain through Lemma 8 some smoothing effects for the degenerate part of the operator. The statement readily follows from \((H_\beta)\) and \((H_\eta)\).
Lemma 9 From the smoothness assumption on the drift coefficient in \((H_\beta)\), there exists \(C := C((A))\) s.t. for all \(\ell \in \llbracket 2, n \rrbracket\), \(k \in \llbracket \ell, n + 1 \rrbracket\), and for any \((s, y, a) \in [0, T] \times (\mathbb{R}^n)^2\):

\[
\left| \left( \begin{array}{c} F_s(s, y_{\ell-1:k-1}, a_{k:n}) - F_s(s, a_{\ell-1:n}) - D_{\ell-1}F_s(s, a_{\ell-1:n})(y - a)_{\ell-1} \end{array} \right) \right| \\
\leq C \left\{ \sum_{j=\ell}^{k-1} \left[ |(F_s)_j(s, \cdot)|_{\beta_j} |(y - a)_j|_{\beta_j} \right] + \left[ |(D_{\ell-1}F_s)_\ell(s, \cdot)|_{\eta} \right] |(y - a)_{\ell-1}|^{1+\eta} \right\},
\]

with the convention that for \(k = n + 1\), \(F_s(s, y_{\ell-1:k-1}, a_{k:n}) = F_s(s, y_{\ell-1:n})\).

### 2.2 Control of the sensitivities: proof of Theorem 3

To prove Theorem 3, the idea is to expand the solution of the PDE with regularized coefficients around a suitable proxy, as explained in Sect. 1.5 (see the connection between Eqs. (1.17) and (1.33)). The proxy used here is the Gaussian process introduced in Sect. 2.1 for a suitable freezing parameter \(\xi\) to be specified later on and whose generator is given, for all \(\varphi \in C^2_0(\mathbb{R}^n, \mathbb{R})\) and \((t, x) \in [0, T] \times \mathbb{R}^n\), by

\[
\bar{L}_t^\xi \varphi(x) := \left( F(t, \theta_{s,t}(\xi)) + DF(t, \theta_{s,t}(\xi))(x - \theta_{s,t}(\xi)) + \frac{1}{2} \text{Tr}(\sigma^s \sigma^t (t, \theta_{s,t}(\xi)) D_\xi^2 \varphi(x)).
\]

Then, the Duhamel formula (or first order parametrix expansion) yields:

\[
u(t, x) = \int_t^T ds \left[ \bar{P}^\xi_{s,t} \varphi(s, \cdot) \right](x) + \int_t^T ds \left[ \bar{P}^\xi_{s,t} \left( (L_s - \bar{L}_s^\xi)u \right) (s, \cdot) \right](x)
\]

\[
= \int_t^T ds \left[ \bar{P}^\xi_{s,t} \varphi(s, \cdot) \right](x) + \int_t^T ds \left[ \bar{P}^\xi_{s,t} \left( (F_1(s, \cdot) - F_1(s, \theta_{s,t}(\xi)), D_1u(s, \cdot)) \right) \right](x)
\]

\[
+ \int_t^T ds \left[ \bar{P}^\xi_{s,t} \left( \frac{1}{2} \text{Tr} \left[ a(s, \cdot) - a(s, \theta_{s,t}(\xi))D_1^2 u(s, \cdot) \right] \right) \right](x)
\]

\[
+ \int_t^T ds \left[ \bar{P}^\xi_{s,t} \left( \sum_{i=2}^n (F_i(s, \cdot) - F_i(s, \theta_{s,t}(\xi))) \right) \right](x)
\]

\[
- D_{i-1}F_i(s, \theta_{s,t}(\xi))(\cdot - \theta_{s,t}(\xi))_{i-1}, D_i u(s, \cdot) \right) \right](x).
\]

for any \(\xi\) in \(\mathbb{R}^n\), according to the notations introduced in Remark 2 for the entries \((F_i)_{i \in \llbracket 1, n \rrbracket}\) of \(F\).

To establish (2.2) we need to differentiate the above expression w.r.t. \((x_i)_{i \in \llbracket 1, n \rrbracket}\) and then w.r.t. \(x_1\). Differentiating first this expression w.r.t. \(x_i, l \in \llbracket 1, n \rrbracket\) we obtain:

\[
\sum_{i=2}^n \left( F_i(s, \theta_{s,t}(\xi)) \right)_{i-1} (\cdot - \theta_{s,t}(\xi))_{i-1} \right) \right](x).
\]
concerning the boundedness of the gradient $D$ of (2.27). We then only concentrate on this term and omit the proof of the statement (2.27). The term $H_i^\xi(s, x)$ gathers the sensitivities w.r.t. $x_1$ of the source term and the non-degenerate part of the difference of the operators, whereas $I_i^\xi(s, x)$ precisely gathers the sensitivities w.r.t. $x_1$ of the degenerate part of the difference of the operators. We will now start from the representation (2.27) which we will again differentiate w.r.t. the non-degenerate variable $x_1$ in order to prove the estimates of Theorem 3 concerning the second order derivatives which are the trickiest ones. Indeed, as it has been succinctly explained in Sect. 1.5, when differentiating the kernel associated with the frozen semigroup defined by (2.23) we generate an a priori non integrable time singularity which then needs to be smoothened by using, among others, tools developed in Lemma 8 (centering or cancellation arguments). The worst case then corresponds to the higher order of differentiation, namely $D_{x_1}(D_{x_1}u(t, x))$ which, as suggested by Proposition 5, generates a time singularity of order 1/2 + (l − 1/2) in the time integrand of the r.h.s. of (2.27). We then only concentrate on this term and omit the proof of the statement concerning the boundedness of the gradient $D_{x_1}u(t, x)$ which could be shown more directly.

The proof will be divided into two parts: we first handle the source and non-degenerate part of the operator (i.e. the estimate for $D_{x_1}H_i^\xi(s, x)$) and then the degenerate part (i.e. the estimate for $D_{x_1}I_i^\xi(s, x)$) which is a bit more involved.

Source term and non-degenerate part of the operator: estimates for $(D_{x_1}H_i^\xi(s, x))_{\xi=x}$. We first focus on the source term and the derivatives w.r.t. the non-degenerate variable $x_1$, three first terms in the r.h.s. denoted by $H_{i,1}^\xi(s, x)$, $H_{i,2}^\xi(s, x)$, $H_{i,3}^\xi(s, x)$ in (2.27), taking $\xi = x$ after the additional differentiation.
For each \( l \in \{1, n\} \), one readily derives from (2.25) in Lemma 8 that for the source term\(^7\):

\[
\left| D_{x_l} H_{l,1}^{\xi} (s, x) \right|_{\xi = x} = \left| D_{x_l} D_{x_l} \left[ \tilde{P}_{s,t}^{\xi} f (s, \cdot) \right] (x) \right|_{\xi = x} \\
\leq C \sum_{j=1}^{n} [ f_j (s, \cdot) ] \beta_j (s-t)^{-l+\beta_j (j-\frac{1}{2})}. \tag{2.28}
\]

Those terms are integrable in time as soon as

\[
\beta_j \left( j - \frac{1}{2} \right) - l > -1, \quad j \in \{l, n\} \iff \beta_j \in \left( \frac{2j - 2}{2j - 1}, 1 \right]. \tag{2.29}
\]

Similarly, from (2.26), (2.27), for the drift associated with the non-degenerate part, we first rewrite from the centering properties of Lemma 8:

\[
\left| D_{x_l} H_{l,2}^{\xi} (s, x) \right|_{\xi = x} = \left| D_{x_l} D_{x_l} \left[ \tilde{P}_{s,t}^{\xi} \left( \left[ F_1 (s, \cdot) - F_1 (s, \theta_{s,t}^{\xi} (\xi)) \right), D_1 u(s, \cdot) \right) \right] (x) \right|_{\xi = x} \\
\leq \left| D_{x_l} D_{x_l} \left[ \tilde{P}_{s,t}^{\xi} \left( \left[ F_1 (s, \cdot) - F_1 (s, \cdot; l-1), \theta_{s,t}^{\xi, n} (\xi)) \right), D_1 u(s, \cdot) \right) \right] (x) \right|_{\xi = x} \\
+ \left| D_{x_l} D_{x_l} \left[ \tilde{P}_{s,t}^{\xi} \left( \left[ F_1 (s, \cdot; l-1), \theta_{s,t}^{\xi, n} (\xi)) \right) - F_1 (s, \theta_{s,t}^{\xi} (\xi))(D_1 u(s, \cdot) \\
- D_1 u(s, \cdot; l-1), \theta_{s,t}^{\xi, n} (\xi)) \right) \right] (x) \right|_{\xi = x}.
\]

Expanding the semi-group yields

\[
\left| D_{x_l} H_{l,2}^{\xi} (s, x) \right|_{\xi = x} \\
\leq \int_{\mathbb{R}^d} d\gamma | D_{x_l} D_{x_l} \tilde{P}_{s,t}^{\xi} (t, s, x, y) |_{\xi = x} | F_1 (s, y) - F_1 (s, y_1;l-1, \theta_{s,t}^{\xi, n} (x)) | | D_1 u(s, y) | \\
+ \int_{\mathbb{R}^d} d\gamma | D_{x_l} D_{x_l} \tilde{P}_{s,t}^{\xi} (t, s, x, y) |_{\xi = x} \\
\left[ F_1 (s, y_1;l-1, \theta_{s,t}^{\xi, n} (x)) - F_1 (s, \theta_{s,t}^{\xi} (x)) \right] | D_1 u(s, y) - D_1 u(s, y_1;l-1, \theta_{s,t}^{\xi, n} (x)) |.
\]

Hence,

\(^7\)Observe that for this contribution, from (2.25) the bound would hold for any freezing parameter \( \xi \). We choose here to take \( \xi = x \) for the compatibility with the other terms \( D_{x_l} H_{l,k}^{\xi} (s, x) |_{k \in \{2,3\}} \) for which this specific choice is indeed needed.
Strong regularization by Brownian noise for a chain of ODEs

\[ |D_{x_1} H^\xi_{l,2}(s, x)|_{\xi=x} \leq \left( \sum_{j=l}^{n} \| D_1 u \|_\infty ((F_1)_j (s, \cdot))^\beta_j \right) \int_{\mathbb{R}^n} dy |D_{x_1} \tilde{P}^\xi (s, t, x, y)|_{\xi=x} |(\theta_{s,t}(x) - y)_j|^{\beta_j} + \sum_{j=l}^{n-1} \sum_{k=1}^{l-1} \| D_1 D_j u \|_\infty ((F_1)_k (s, \cdot))^\beta_k \int_{\mathbb{R}^n} dy |D_{x_1} \tilde{P}^\xi (s, t, x, y)|_{\xi=x} |(\theta_{s,t}(x) - y)_k|^{\beta_k} |(\theta_{s,t}(x) - y)_j| \right). \]

We then conclude from equation (2.21) of Lemma 7, taking \( \gamma_1 = \beta_j, \gamma_2 = 0 \) for the first terms and \( \gamma_1 = 1, \gamma_2 = \beta_k \) for the second ones, that

\[ |D_{x_1} H^\xi_{l,2}(s, x)|_{\xi=x} \leq C \left( \sum_{j=l}^{n} \| D_1 u \|_\infty ((F_1)_j (s, \cdot))^\beta_j (s-t)^{\beta_j \left( j - \frac{1}{2} \right)} + \sum_{j=l}^{n-1} \sum_{k=1}^{l-1} \| D_1 D_j u \|_\infty ((F_1)_k (s, \cdot))^\beta_k (s-t)^{\beta_k \left( k - \frac{1}{2} \right) + \left( j - \frac{1}{2} \right)} \right), \]

leading precisely to the same integrability thresholds of equation (2.29) and assumption (T_\beta) (as for the source term). The idea behind this control is crucial. We first handle, with the sole Hölder properties of the drift and the supremum norm of \( D_1 u \), the variables which are at a good smoothing scale w.r.t. the induced singularity. For the remaining term, which exhibits for the drift non-sufficient smoothing effects, we then additionally exploit a cancellation argument involving the gradient of the solution itself, which consequently makes the cross derivatives appear.

Eventually, we get for the diffusive part:

\[ |D_{x_1} H^\xi_{l,3}(s, x)|_{\xi=x} = |D_{x_1} D_{x_1} \left[ \tilde{P}^\xi (s, t) \left( \frac{1}{2} \text{Tr} \left[ (a(s, \cdot) - a(s, \theta_{s,t}(\xi))) D^2_1 u(s, \cdot) \right] \right) \right](x)|_{\xi=x} \leq |D_{x_1} D_{x_1} \left[ \tilde{P}^\xi (s, t) \left( \frac{1}{2} \text{Tr} \left[ (a(s, \cdot) - a(s, \cdot, \cdot, \cdot)_{s,t}(\xi)) D^2_1 u(s, \cdot) \right] \right) \right](x)|_{\xi=x} + |D_{x_1} D_{x_1} \left[ \tilde{P}^\xi (s, t) \left( \frac{1}{2} \text{Tr} \left[ (a(s, \cdot, \cdot, \cdot)_{s,t}(\xi)) - a(s, \theta_{s,t}(\xi))) D^2_1 u(s, \cdot) \right] \right) \right](x)|_{\xi=x} = |D_{x_1} H^\xi_{l,31}(s, x)|_{\xi=x} + |D_{x_1} H^\xi_{l,32}(s, x)|_{\xi=x}. \]

The term \( |D_{x_1} H^\xi_{l,31}(s, x)|_{\xi=x} \) is already centered at the appropriate scales, i.e. from variables \( l \) to \( n \). Recalling that \( a \) is Lipschitz continuous, write:
\[|D_{x_1}H_{t,31}^{\xi}(s,x)|_{\xi=x} = \int_{R^d} dy D_{x_1} D_{x_1} \tilde{p}_t^{\xi}(t,s,x,y) \left( \frac{1}{2} \text{Tr} \left[ (a(s,y) - a(s,y_{1:l-1},\theta_{s,t}^{l,n}(\xi))) D_1^{2} u(s,y) \right] \right) \bigg|_{\xi=x} \]

\[\leq \frac{1}{2} \int_{R^d} dy |D_{x_1} D_{x_1} \tilde{p}_t^{\xi}(t,s,x,y)|_{\xi=x} \left| \left( (a(s,y) - a(s,y_{1:l-1},\theta_{s,t}^{l,n}(\xi))) \right) \right| |D_1^{2} u(s,y)| \]

\[\leq \frac{1}{2} \left[ (a(s,\cdot)) \right]_1 |D_1^{2} u|_{\infty} \int_{R^d} dy |D_{x_1} D_{x_1} \tilde{p}_t^{\xi}(t,s,x,y)|_{\xi=x} |y_{l:n} - \theta_{s,t}^{l,n}(x)|. \]

We can apply (2.21) of Lemma 7 with \(\gamma_1 = 1, \gamma_2 = 0\) to any of the entries \(|y_j - \theta_{s,t}^{l,n}(x)|, j \in [l,n]\) to obtain:

\[|D_{x_1}H_{t,31}^{\xi}(s,x)|_{\xi=x} \leq C |D_1^{2} u|_{\infty} \left[ (a(s,\cdot)) \right]_1 \sum_{j=l}^{n} (s-t)^{-l \left( j-\frac{1}{2} \right)} \leq C |D_1^{2} u|_{\infty} \left[ (a(s,\cdot)) \right]_1 (s-t)^{-\frac{1}{2}}, \]

which does not give a critical contribution w.r.t. the previously exhibited thresholds in (2.29) and (T_\beta). For the contribution \(|D_{x_1}H_{t,32}^{\xi}(s,x)|_{\xi=x}\) we use, in the same spirit as for \(|D_{x_1}H_{t,31}^{\xi}(s,x)|_{\xi=x}\), a centering argument and an integration by parts to obtain:

\[|D_{x_1}H_{t,32}^{\xi}(s,x)|_{\xi=x} = \left| D_{x_1} D_{x_1} \left[ \tilde{p}_t^{\xi}\left( \frac{1}{2} \text{Tr} \left[ \left( (a(s,\cdot,1:l-1,\theta_{s,t}^{l,n}(\xi))) - a(s,\theta_{s,t}(\xi)) \right) \left( D_1^{2} u(s,\cdot) - D_1^{2} u(s,1:l-1,\theta_{s,t}^{l,n}(\xi)) \right) \right] \right)(x) \right| \bigg|_{\xi=x} \]

\[= \frac{1}{2} D_{x_1} D_{x_1} \left( \sum_{j=1}^{d} \int_{R^d} dy \left( \partial_{y}^{j} \left( \tilde{p}_t^{\xi}(t,s,x,y) \left( a_j(s,1:l-1,\theta_{s,t}^{l,n}(\xi)) - a_j(s,\theta_{s,t}(\xi)) \right) \right) \right) \left( D_1^{2} u(s,y) - D_1^{2} u(s,1:l-1,\theta_{s,t}^{l,n}(\xi)) \right) \right) \bigg|_{\xi=x}, \]

where \(y_1 = (y^1, \ldots, y^n), \partial_{y}^{j}\) denotes the derivative w.r.t. the \(j\)th scalar entry of the non-degenerate variable \(y_1\) and \(a_j\) denotes the \(j\)th row of the diffusion matrix \(a\). We therefore derive from the proof of Lemma 7, (2.15) and the smoothness of \(a\) (by the Rademacher theorem, \(a\) is differentiable almost everywhere):

\[|D_{x_1}H_{t,32}(s,x)|_{\xi=x} \leq C |D_{x_1} D_{x_1} u(s,\cdot)|_{\infty} \left( \frac{1}{(s-t)^{l-\frac{1}{2}}} \right) \left| y_{1:l} - \theta_{s,t}(x) \right|_{l:n} \leq C |D_{x_1} D_{x_1} u(s,\cdot)|_{\infty} (s-t)^{-\frac{1}{2}} \cdot \]

\[\square \text{ Springer} \]
With the notations of (2.27), plugging (2.32), (2.33) into (2.31) and together with (2.30), (2.28), we eventually derive that there exists $\delta := \delta((A)) > 0$:

$$
\left| \int_t^T ds D_{x_1} H^\xi_{i_1}(s, x) \right|_{\xi=x} \leq CT^\delta (\| D_1 u \|_\infty + \| D D_1 u \|_\infty + 1). \tag{2.34}
$$

**Degenerate part of the operator: estimates for** $(D_{x_1} I^\xi_{i_1} (s, x))_{\xi=x}$. These are the most delicate terms to handle. Restarting from (2.27), we first write for any $l \in \llbracket 1, n \rrbracket$:

$$
\begin{align*}
|D_{x_1} I^\xi_{i_1} (s, x)|_{\xi=x} &= |D_{x_1} D_{x_2} \left[ \tilde{P}^\xi_{s,t} \left( \sum_{i=2}^n \left( (F_i(s, \cdot) - F_i(s, \theta_{s,t}(\xi)) - D_{i-1} F_i(s, \theta_{s,t}(\xi))(\cdot - \theta_{s,t}(\xi))_{i-1}, D_1 u(s, \cdot) \right) \right) \right](x)|_{\xi=x} \\
&+ |D_{x_1} D_{x_2} \left[ \tilde{P}^\xi_{s,t} \left( \sum_{i=2}^l \left( (F_i(s, \cdot) - F_i(s, \theta_{1:1\ldots:1-1, i-1, \theta_{s,t}(\xi)}(s, \theta_{s,t}(\xi))) - D_{i-1} F_i(s, \theta_{s,t}(\xi))(\cdot - \theta_{s,t}(\xi))_{i-1}, D_1 u(s, \cdot) \right) \right) \right](x)|_{\xi=x} \\
&+ |D_{x_1} D_{x_2} \left[ \tilde{P}^\xi_{s,t} \left( \sum_{i=1}^l \left( (F_i(s, \cdot) - F_i(s, \theta_{1:1\ldots:1-1, i-1, \theta_{s,t}(\xi)}(s, \theta_{s,t}(\xi))) - F_i(s, \theta_{s,t}(\xi)) \right) - D_{i-1} F_i(s, \theta_{s,t}(\xi))(\cdot - \theta_{s,t}(\xi))_{i-1}, D_1 u(s, \cdot) \right) \right](x)|_{\xi=x} \\
&=: |D_{x_1} I^\xi_{i_1,1} (s, x)|_{\xi=x} + |D_{x_1} I^\xi_{i_1,2} (s, x)|_{\xi=x} + |D_{x_1} I^\xi_{i_1,3} (s, x)|_{\xi=x}.
\end{align*}
\tag{2.35}
$$

- **Control of** $|D_{x_1} I^\xi_{i_1,1} (s, x)|_{\xi=x}$ and $|D_{x_1} I^\xi_{i_1,2} (s, x)|_{\xi=x}$. We emphasize that the integrands $|D_{x_1} I^\xi_{i_1,1} (s, x)|_{\xi=x}$ and $|D_{x_1} I^\xi_{i_1,2} (s, x)|_{\xi=x}$ are already designed to smoothen the time singularities generated by the cross differentiation of the inhomogeneous semi-group w.r.t. the variables $x_i$ and $x_1$. Indeed, write:
where \( D_{l+1:n}u := (D_{l+1}u, \ldots, D_n u) \). Applying now Lemma 9 with \( a = \theta_{s,t}(x) \), \( \ell = i, k = n + 1 \) yields:

\[
\sum_{i = l+1}^{n} \left| F_i(s,y) - F_i(s,\theta_{s,t}(x)) - D_{i-1} F_i(s,\theta_{s,t}(x))(y - \theta_{s,t}(x))_{i-1} \right|
\]

\[
\leq C \sum_{j = l+1}^{n} \left( \left( (F_{l+1:j}(s,\cdot))_j \right)_{\beta_j} |(\theta_{s,t}(x) - y)_j|^{\beta_j} + |(D_{j-1} F_j(s,\cdot))_{j-1}||\theta_{s,t}(x) - y|_{j-1}^{1+\eta} \right),
\]

using the notation \( \left( (F_{l+1:j}(s,\cdot))_j \right)_{\beta_j} := \max_{j \in [l+1,j]} \left( (F_i)_j(s,\cdot) \right)_{\beta_j} \) for the last inequality. We finally derive from Lemma 7:

\[
|D_{x_1}I_{l,1}^\xi(s,x)|_{\xi = x}
\]

\[
\leq C \| D_{l+1:n} u \|_\infty \sum_{j = l+1}^{n} \left( \left( (F_{l+1:j}(s,\cdot))_j \right)_{\beta_j} (s-t)^{-l + \beta_j \left( j - \frac{1}{2} \right)} + |(D_{j-1} F_j(s,\cdot))_{j-1}||\theta_{s,t}(x) - y|^{1+\eta}_{j-1} \right) \right) ,
\]

(2.36)

Since in the above contribution \( j \geq l+1 \), we have on the one hand \(-l + (1 + \eta)(j - 3/2) \geq -l + (1 + \eta)(l + 1 - 3/2) \geq -1/2 \). On the other hand, from assumption (T_{\beta}), \( \beta_j \geq (2j-2)/(2j-1) \) and \(-l + \beta_j(j - 1/2) \geq -l + (2j-2)/(2j-1) \times (j-1/2) = -l + j - 1 \geq 0 \). Thus, none of the associated exponent is critical. We have either integrable singularities or no singularities at all.

Write now for the other term,

\[
|D_{x_1}I_{l,2}^\xi(s,x)|_{\xi = x}
\]

\[
= \left| D_{x_1} D_{x_2} \left[ \tilde{p}_{s,t}^\xi \left( \sum_{i=2}^{l} (F_i(s,\cdot) - F_i(s,\cdot|_{i-1},\theta_{s,t}(\xi)), D_i u(s,\cdot)) \right) \right](x) \right|_{\xi = x}
\]

\( \odot \) Springer
\[ \leq \sum_{l=2}^{n} \int_{\mathbb{R}^{nd}} dy |D_{X_1} D_{X_1} \tilde{p}^\xi(s, t, x, y)|_{\xi=x} |F_i(s, y) - F_i(s, y_1; \theta^{l,n}(\xi))| |D_i u(s, y)| \]

\[ \leq \| D_{2;l} u \|_{\infty} \sum_{l=2}^{n} \sum_{j=1}^{n} \int_{\mathbb{R}^{nd}} dy |D_{X_1} D_{X_1} \tilde{p}^\xi(s, t, x, y)|_{\xi=x} [(F_i)_{j}(s, \cdot) \beta_j |(\theta_{s,t}(x) - y)_{j} \beta_j] \]

\[ \leq C \| D_{2;l} u \|_{\infty} \sum_{j=1}^{n} [(F_{2;j})_{j}(s, \cdot) \beta_j \int_{\mathbb{R}^{nd}} dy |D_{X_1} D_{X_1} \tilde{p}^\xi(s, t, x, y)|_{\xi=x} |(\theta_{s,t}(x) - y)_{j} \beta_j]. \]

Hence, Lemma 7 also yields

\[ |D_{X_1} I_{l,2}^\xi(s, x)|_{\xi=x} \leq C \| D_{2;l} u \|_{\infty} \sum_{j=1}^{n} [(F_{2;j})_{j}(s, \cdot) \beta_j (s-t)^{-l+\beta_j (j-\frac{1}{2})}, (2.37) \]

and those terms are again integrable as soon as the thresholds of \((T_{\beta})\) hold.

- **Control of** \(|D_{X_1} I_{l,3}^\xi(s, x)|_{\xi=x}\). It hence remains to control the terms in \(|D_{X_1} I_{l,3}^\xi(s, x)|_{\xi=x}\) for \(l \in [2, n]\) \(^8\). These terms are the tricky ones since they are, *a priori*, not designed to smooth the time singularities generated by the cross differentiation. Observe indeed that, if one tries to reproduce the above calculations, we obtain from Lemmas 7 and 9, that

\[ |D_{X_1} I_{l,3}^\xi(s, x)|_{\xi=x} \]

\[ = |D_{X_1} D_{X_1} \left[ \tilde{p}^\xi(s, t) \left( \sum_{i=2}^{l} (F_i(s, \cdot; \theta^{l,n}(\xi)) - F_i(s, \theta_{s,t}(\xi)) \right. \right. \]

\[ \left. \left. - D_{i-1} F_i(s, \theta_{s,t}(\xi))(\cdot - \theta_{s,t}(\xi))_{i-1}, D_i u(s, \cdot)) \right) \right](x)|_{\xi=x} \]

\[ \leq C \| D_{2;l} u \|_{\infty} (s-t)^{-l} \]

\[ \times \left( \sum_{j=2}^{n} [(F_{2;j})_{j}(s, \cdot) \beta_j (s-t)^{(j-\frac{1}{2})\beta_j} + [(D_{j-1} F_j)_{j-1}(s, \cdot) \eta(s-t)^{(1+\eta)(j-\frac{1}{2})}] \right) \]

\[ \leq C \| D_{2;l} u \|_{\infty} (s-t)^{-l} \left( (s-t)^{\frac{2}{(1+\eta)(j-\frac{1}{2})}} + (s-t)^{\frac{1}{(1+\eta)}} \right), \]

up to a modification of \(C\) and recalling that \(T\) is small. This leads, as soon as \(l \geq 2\), to a time singularity which is not integrable. Indeed, \((1 + \eta)/2 < 1\) (recall that \(\eta\) is meant to be small). To overcome this problem, the idea consists in writing, thanks to the cancellation properties of Lemma 8,

\[ |D_{X_1} I_{l,3}^\xi(s, x)|_{\xi=x} \]

\(^8\) Since \(I_{l,3}^\xi(s, x) = 0\) for all \(x \in \mathbb{R}^{nd}\), the contribution is clearly 0 for \(l = 1\).
\[
= \left| D_{x_1} D_{x_l} \left[ \hat{P}_{s,t}^i \left( \sum_{i=2}^{l} (F_i(s, \cdot, l-1, \theta_{s,i}^l(x))) \right. \right.
\left. \left. - F_i(s, \theta_{s,t}(x)) \right) - D_{l-1} F_i(s, \theta_{s,t}(x)) \left( \theta_{s,i}^l(x) \right) \right) \right] (x) \right|_{x=x} (2.39)
\]

and to take advantage of the additional smoothing effect from the solution of the regularized PDE itself through the above contribution \(D_i u(s, \cdot) - D_i u(s, \cdot, l-1, \theta^l_{s,i}(x))\), \(i \in [2, l]\). This was the strategy implemented in [10] which, unfortunately, cannot be repeated as this in our general framework. Roughly speaking, to control for any \(k \in [l, n]\), for fixed \((y_{1:k-1}, y_{k+1:n}) \in \mathbb{R}^{(n-1)d}\) and any \(0 \leq s < t \leq T\), the \(\alpha^l_k\)-Hölder modulus of the partial application (see (1.14)) \(y_k \mapsto (D_i u)_k(s, y_k)\), we need to control the \(\alpha^l_k\)-Hölder modulus of the map \(y_k \mapsto (D_i \tilde{G}^s f)_k(s, y_k)\), where \(\tilde{G}^s f(s, y) := \int T dv \tilde{P}_{v,s} \hat{G}^s f(v, y)\) is the Green kernel involving the source \(f\) in the indicated equation. When doing so, we are faced with a situation where we should only consider indices \(k \leq i\) in order to apply cancellation arguments. The only corresponding index in (2.39) is then \(i = k = l\). When \(i = l\) and \(k\) lies in \([l+1, n]\), we can handle the corresponding terms by interpolating the \(\alpha^l_k\)-Hölder modulus of the partial application \((D_i u)_k\) from the \(\alpha^l_k\)-Hölder modulus of the partial application \((D_i u)_l\), see Lemma 13 below. Otherwise, i.e. when \(i < l\), we expand the gradients in (2.39) with the Taylor formula. Namely, using the Fubini theorem we write:

\[
\left| D_{x_1} I_{l,3}^i (s, x) \right|_{x=x} \leq \sum_{i=2}^{l-1} \sum_{k=1}^{n} \int T d \lambda \left| D_{x_1} D_{x_l} \hat{P}_{s,t}^i \left( \left( F_i(s, \cdot, l-1, \theta^l_{s,i}(x)) - F_i(s, \theta_{s,t}(x)) \right) - D_{l-1} F_i(s, \theta_{s,t}(x)) \left( \theta_{s,i}^l(x) \right) \right) \right] (x) \right|_{x=x} \]

\[
+ \left| D_{x_1} D_{x_l} \hat{P}_{s,t}^i \left( \left( F_i(s, \cdot, l-1, \theta^l_{s,i}(x)) \right. \right. \left. \left. - F_i(s, \theta_{s,t}(x)) \right) - D_{l-1} F_i(s, \theta_{s,t}(x)) \left( \theta_{s,i}^l(x) \right) \right) \right] (x) \right|_{x=x} \]

\[
= \left| D_{x_1} I_{l,31}^i (s, x) \right|_{x=x} + \left| D_{x_1} I_{l,32}^i (s, x) \right|_{x=x}. (2.40)
\]

As already underlined, the term \(\left| D_{x_1} I_{l,32}^i (s, x) \right|_{x=x}\) is handled thanks to interpolation type argument. It thus remains to deal with the term \(\left| D_{x_1} I_{l,31}^i (s, x) \right|_{x=x}\). Our main idea consists first in switching the differential operators acting on the map \(x\) therein and then in rebalancing the differential operator w.r.t. the “less degenerate” direction (namely
\( D_l \) on the remaining terms in the integrand through an integration by parts. To do so, we introduce now for all \( i \in [2, l-1] \), \( k \in [l, n] \), \((y_{1:i-1}, y_{i+1:n}) \in \mathbb{R}^{(n-1)d} \), \((t, x) \in [0, T] \times \mathbb{R}^{nd} \), \( s \in (t, T) \) and \( y_i \in \mathbb{R}^d \), the function:

\[
\Psi_{l,(l,1),k}^{(s,y_{1:i-1},y_{i+1:n}),(t,\bar{x})}(y_i) := \left[ D_{x_i} D_{x_i} \tilde{p}^k(t, s, x, y) \otimes \left( (F_i(s, y_{1:l-1}, \bar{\theta}_{s:l,t}^{l:n}(\bar{x})) - F_i(s, \theta_{s:l,t}(\bar{x}))) - D_{l-1} F_i(s, \theta_{s:l,t}(\bar{x}))(y - \theta_{s:l,t}(\bar{x}))_{l-1} \right) ((y - \theta_{s:l,t}(\bar{x}))_k) \right]_{\xi = x}, (2.41)
\]

where the subscript \((l, 1)\) in \( \Psi_{l,(l,1),k}^{(s,y_{1:i-1},y_{i+1:n}),(t,\bar{x})} \) is here to indicate the differentiation w.r.t. \( D_{x_i} D_{x_i} \) acting on the frozen density. Pay attention that the above function is \((\mathbb{R}^d)^{\otimes 4}\)-valued. With these notations at hand, we write for \( |D_{x_i} I_{l,31}^{\xi}(s, x)|_{\xi = x} \):

\[
|D_{x_i} I_{l,31}^{\xi}(s, x)|_{\xi = x} \leq \sum_{i=2}^{l-1} \sum_{k=l}^{n} \left| \int_0^1 d\lambda \int_{\mathbb{R}^{(n-1)d}} d(y_{1:i-1}, y_{i+1:n}) \int_{\mathbb{R}^d} dy_i \left\{ \Psi_{l,(l,1),k}^{(s,y_{1:i-1},y_{i+1:n}),(t,\bar{x})}(y_i) : D_{y_i} D_{y_i} u(s, y_{1:l-1}, \theta_{s:l,t}^{l:n}(x) + \lambda (y_{l:n} - \theta_{s:l,t}^{l:n}(x))) \right\} \right|, (2.42)
\]

where “:” stands for the double tensor contraction. We now use the Schwarz theorem to exchange the order of the differentiation operators acting on the PDE solution and then integrate by parts to obtain

\[
|D_{x_i} I_{l,31}^{\xi}(s, x)|_{\xi = x} \leq \sum_{i=2}^{l-1} \sum_{k=l}^{n} \left| \int_0^1 d\lambda \int_{\mathbb{R}^{(n-1)d}} d(y_{1:i-1}, y_{i+1:n}) \int_{\mathbb{R}^d} dy_i \left\{ D_{y_i} \left( \Psi_{l,(l,1),k}^{(s,y_{1:i-1},y_{i+1:n}),(t,\bar{x})}(y_i) : D_{y_i} u(s, y_{1:l-1}, \theta_{s:l,t}^{l:n}(x) + \lambda (y_{l:n} - \theta_{s:l,t}^{l:n}(x))) \right) \right\} \right|.
\]

Let us now explain why such an expression is well designed. From the very definition of the \( \Psi \) term in (2.41), one can see that the additional contribution \((\cdot - \theta_{s,l,t}(x))_k\) is, thanks to Lemma 7, precisely designed to smoothen the time singularity coming from the differentiation w.r.t. the variables \( x \) of the semigroup (recall that \( k \leq l \)). Also, we have from Lemma 9 that the contribution of the transport term (degenerate part of the operator) therein is, up to a multiplicative constant, bounded by \( \sum_{j=i}^n (\cdot \).
Lemma 10
Let \( l \in \llbracket 2, n \rrbracket \), \( i \in \llbracket 2, l - 1 \rrbracket \) and \( k \in \llbracket l, n \rrbracket \). Let \( \Psi^{(s,y_{i-1},y_{i+1:n}),k}_{i,n}(t, \mathbf{x}) : \mathbb{R}^d \to (\mathbb{R}^d)^{\otimes 4} \) be the function defined by (2.41). There exist \( C := C((\mathbf{A}), T) > 0 \), \( \tilde{C} := \tilde{C}(C) \), \( \alpha_i := 1/(2i - 1) \), \( \gamma_i := 1/2 + \eta(i - 3/2) \) such that

\[
\left\| D_l \Psi^{(s,y_{i-1},y_{i+1:n}),k}_{i,n}(t, \mathbf{x}) \right\|_{B_{1,1}^{\alpha_i}} \leq \tilde{C} (s - t)^{-\frac{3}{2} + \gamma_i} \tilde{q}_{C_i}(t, s, \mathbf{x}, (y_{1:i-1}, y_{i+1:n})),
\]

\( \odot \) Springer
where \( c = C^{-1} \) and with the notations of Proposition 5, exploiting as well equation (2.19) of Remark 5,

\[
\hat{q}_{c,i}(t, s, x, (y_{1:i-1}, y_{i+1:n})) := \int_{\mathbb{R}^d} dy_i \hat{p}_c^x(t, s, x, y) = \prod_{j \in \{1,n\}, j \neq i} \mathcal{N}_{c(s-t)^{2j-1}}((\theta_{s,t}(x) - y)_j), \quad (2.44)
\]

denoting for \( a > 0 \), \( z \in \mathbb{R}^d \), by \( \mathcal{N}_a(z) = (2\pi a)^{-d/2} \exp\left(-|z|^2/(2a)\right) \) the standard Gaussian density on \( \mathbb{R}^d \) with covariance matrix \( aI_d \).

**Lemma 11** Let \( u \) be the solution of (2.1). There exists \( C := C((A)) > 0 \) such that for all \( i \leq k \in \{2, n\}^2 \) and \( \alpha_i = 1/(2i - 1) \),

\[
\sup_{y_j, j \in \{1,n\}, j \neq i} \left\| D_k u(s, y_{1:i-1}, \cdot, y_{i+1:n}) \right\|_{B^{\alpha_i} \infty} \leq C(\| Du \|_\infty + \| D D_1 u \|_\infty). \quad (2.45)
\]

To control the term \( |D_{s_i} I_{l, 32}^{x}(s, x)|_{\xi = x} \) in (2.40), we will need the following auxiliary lemma whose proof is postponed to Appendix D. It can be seen as a complement to Lemma 11 since it precisely characterizes the Hölder regularity of \( (D_l u)_k(s, \cdot) \) \( k \in \{l+1, n\} \) i.e. for the variables that are strictly greater than the current differentiation index \( l \).

**Lemma 12** (Hölder moduli of the gradients strictly after the differentiation index). There is a constant \( C > 0 \) s.t. for all \( l \in \{2, n\}, k \in \{l+1, n\} \), \( s \in [t, T] \):

\[
[(D_l u)_k(s, \cdot)]_{\xi_l} \leq C(\| Du \|_\infty + [(D_l u)_l(s, \cdot)]_{\alpha_l}),
\]

where \( \xi_l := 1/2l \) and \( \alpha_l = 1/(2l - 1) \) as in Lemma 11.

As a consequence of Lemma 12 we readily get:

**Lemma 13** (Hölder moduli of the gradients after the differentiation index). There is a constant \( C > 0 \) s.t. for all \( l \in \{2, n\}, (y, \xi) \in \mathbb{R}^{nd} \times \mathbb{R}^{nd}, s \in [t, T] \):

\[
|D_l u(s, y) - D_l u(s, y_{1:l-1}, \theta_{s,t}^{l:n}(\xi))| \leq C(\| Du \|_\infty + [(D_l u)_l(s, \cdot)]_{\alpha_l}) \left( |(\theta_{s,t}(\xi) - y)_l|^{\alpha_l} + \sum_{k=l+1}^n |(\theta_{s,t}(\xi) - y)_k|^{\xi_l} \right),
\]

with \( \xi_l, \alpha_l \) as in Lemma 12.

Now, from Proposition 5, Lemma 9 and Lemma 13,

\[
|D_{s_i} I_{l, 32}^{x}(s, x)|_{\xi = x} \leq \frac{\tilde{C}[(D_{l-1} F_l)_{l-1}(s, \cdot)]_{\eta}}{(s - t)^l} \int_{\mathbb{R}^{nd}} dy \hat{p}_c^{x}(t, s, x, y)|(\theta_{s,t}(x) - y)_{l-1}|^{1+\eta}
\]
\[
\times (\| Du\|_\infty + [(D_t u)_l(s, \cdot)]_{a_l}) (\| (\theta_{s,t}(x) - y)_l \|^a_l + \sum_{k=l+1}^n \| (\theta_{s,t}(x) - y)_k \|^\zeta_l)
\]
\[
\leq C (\| Du\|_\infty + [(D_t u)_l(s, \cdot)]_{a_l})
\times \left( \sum_{k=l}^n (s - t)^{-l+(1+\gamma)\left(l - \frac{3}{2}\right) + a_l \left(l - \frac{1}{2}\right) + \zeta_l (k - \frac{1}{2})} \right),
\]  
\[
(2.46)
\]

using as well Lemma 7 for the last inequality. Hence, the above equation yields:

\[
\| D_{x_l} I_{l,32}^t(s, x) \|_{\xi = x} \leq C (\| Du\|_\infty + \| (D_t u)_l(s, \cdot) \|_{B^a_l,\infty})
\times \sum_{k=l}^n (s - t)^{-l+(1+\gamma)\left(l - \frac{3}{2}\right) + a_l \left(l - \frac{1}{2}\right) + \zeta_l (k - \frac{1}{2})} \|_{l \in \{l+1, n\}}.
\]  
\[
(2.47)
\]

Observe now that the time exponents in the previous r.h.s. are integrable. In particular, for each \( l \in \{2, n\} \)

\[
-l + (1 + \gamma)\left(l - \frac{3}{2}\right) + \frac{1}{2} \geq -3 + \eta\left(l - \frac{3}{2}\right) + \eta\left(l - \frac{3}{2}\right) + \frac{1}{2} = -\frac{3}{2} + \gamma_l,
\]  
\[
(2.48)
\]

with the notations of Lemma 10. Also, for \( k \in \{l+1, n\} \):

\[
-l + (1 + \gamma)\left(l - \frac{3}{2}\right) + \zeta_l \left(k - \frac{1}{2}\right)
\]

\[
= -3 + \eta\left(l - \frac{3}{2}\right) + \frac{1}{2} \left(k - \frac{1}{2}\right) \geq -3 + \eta\left(l - \frac{3}{2}\right) + \frac{1}{2} \frac{l + \frac{1}{2}}{l}
\]

\[
\geq -\frac{3}{2} + \eta\left(l - \frac{3}{2}\right) + \frac{1}{2} = -\frac{3}{2} + \gamma_l.
\]  
\[
(2.49)
\]

Observe carefully from the above computations that the exponent \( \zeta_l \) precisely allows to recover a smoothing effect in time of order strictly greater than \( 1/2 \). This perfectly fits the smoothing effects observed for the other contributions.

We can then deduce from (2.40), (2.43), Lemma 10, (2.47) and Lemma 11 that

\[
\| D_{x_l} I_{l,3}^t(s, x) \|_{\xi = x} \leq C \sum_{i=2}^l (s - t)^{-\frac{3}{2} + \gamma_i} \left(\| Du\|_\infty + \| D_1 Du\|_\infty \right),
\]  
\[
(2.50)
\]

which are integrable terms since \( \gamma_i > 1/2 \). With the notations of (2.27), (2.35), we eventually derive from (2.50), (2.37), (2.36) that there exists \( \gamma := \gamma((\mathcal{A})) > 0 \) such that:

\[
\left| \int_t^T ds D_{x_l} I_{l,3}^t(s, x) \right|_{\xi = x} \leq C T^\gamma \left(\| Du\|_\infty + \| D_1 Du\|_\infty \right).
\]  
\[
(2.51)
\]
Final proof of Theorem 3. Bringing together (2.51) and (2.34) yields for all \( l \in \{1, n\} \) and \((t, x) \in [0, T]\):

\[
|D_{x_l} D_{x_1} u(t, x)| \leq C(T^{\gamma} + T^\delta)(\|Du\|_\infty + \|DD_1u\|_\infty + 1).
\] (2.52)

It is clear that the previous analysis can be reproduced without differentiating w.r.t. \( x_1 \), leading to improved singularity exponents (see also the proof of Lemma 11 which somehow exactly explicit these computations). We therefore get:

\[
|D_{x_l} u(t, x)| \leq C(T^{\gamma'} + T^{\delta'})(\|Du\|_\infty + \|DD_1u\|_\infty + 1),
\] (2.53)

for some positive exponents \( \gamma', \delta' \) (with \( \gamma' > \gamma, \delta' > \delta \)).

Taking the time-space supremum in the l.h.s of (2.52) and (2.53), recalling as well that \( T \) is meant to be small, i.e. s.t. \( 4CT^{\delta v} \leq 1/2 \), we derive:

\[
\|Du\|_\infty + \|DD_1u\|_\infty \leq 2C(T^{\gamma'} + T^\delta).
\]

This concludes the proof.

\[ \square \]

3 Estimates in Besov norm

This section is dedicated to the proofs of the main technical results needed to obtain Theorem 3. Namely, we prove the Besov estimates of Lemmas 10 and 11. We first start by recalling some definitions/characterizations of Besov spaces from Section 2.6.4 of Triebel [62]. For \( \alpha \in \mathbb{R}, q \in (0, +\infty], p \in (0, \infty], B_{p,q}^\alpha(\mathbb{R}^d) : = \{ f \in S'(\mathbb{R}^d) : \| f \|_{\mathcal{H}^\alpha_{p,q}} < +\infty \} \) where \( S(\mathbb{R}^d) \) stands for the Schwartz class and

\[
\| f \|_{\mathcal{H}^\alpha_{p,q}} : = \| \varphi(D)f \|_{L^p(\mathbb{R}^d)} + \left( \int_0^1 \frac{dv}{v} v^{(m-\frac{d}{2})q} \| \partial_v h_v*f \|_{L^q(\mathbb{R}^d)} \right)^{\frac{1}{q}},
\] (3.1)

with \( \varphi \in C_0^\infty(\mathbb{R}^d) \) (smooth function with compact support) s.t. \( \varphi(0) \neq 0, \varphi(D)f : = \hat{\varphi}\hat{f} \) where \( \hat{\varphi} \) and \( \hat{(\varphi f)} \) respectively denote the Fourier transform of \( f \) and the inverse Fourier transform of \( \varphi f \). The parameter \( m \) is an integer s.t. \( m > \alpha / 2 \) and for \( v > 0, z \in \mathbb{R}^d, h_v(z) : = \frac{1}{(2\pi v)^{d/2}} \exp \left( -|z|^2 / 2v \right) \) is the usual heat kernel of \( \mathbb{R}^d \). We point out that the quantities in (3.1) are well defined for \( q < \infty \). The modifications for \( q = +\infty \) are obvious and can be written passing to the limit.

Observe that the quantity \( \| f \|_{\mathcal{H}^\alpha_{p,q}} \), where the subscript \( \mathcal{H} \) stands to indicate the dependence on the heat-kernel, depends on the considered function \( \varphi \) and the chosen \( m \in \mathbb{N} \). It also defines a quasi-norm on \( B_{p,q}^\alpha(\mathbb{R}^d) \). The previous definition of \( B_{p,q}^\alpha(\mathbb{R}^d) \) is known as the thermic characterization of Besov spaces and is particularly well adapted to our current framework. By abuse of notation we will write as soon as this quantity is finite \( \| f \|_{\mathcal{H}^\alpha_{p,q}} =: \| f \|_{B_{p,q}^\alpha} \).

\[ \square \] Springer
3.1 Proof of Lemma 10

We will here exploit the thermic characterization of Besov spaces (see Chapter 2.6.4 in [62]) which is also recalled above.

From (3.1), we are thus led to estimate, for any \( l \in \{2, n\} \), \( i \in \{2, l - 1\} \) and \( k \in \{l, n\} \):

\[
\| \varphi(D) D_i \Psi^{(s, y_{i-1}, y_{i+1,n}), (t, x)}_{i, (l, 1), k} \|_{L^1(\mathbb{R}^d, \mathbb{R})} + \int_0^1 \frac{dv}{v} \left[ 1 + a_v \| \partial_v h_v \right. \left. \star D_i \Psi^{(s, y_{i-1}, y_{i+1,n}), (t, x)}_{i, (l, 1), k} \|_{L^1(\mathbb{R}^d, \mathbb{R})},
\]

for a \( C^\infty \) compactly supported function \( \varphi \) s.t. \( \varphi(0) \neq 0 \). From the definition of \( \Psi^{(s, y_{i-1}, y_{i+1,n}), (t, x)}_{i, (l, 1), k} \) in (2.41), using Lemmas 7 and 9, we easily deduce

\[
\| \varphi(D) D_i \Psi^{(s, y_{i-1}, y_{i+1,n}), (t, x)}_{i, (l, 1), k} \|_{L^1(\mathbb{R}^d, \mathbb{R})} \leq \int_{\mathbb{R}^d} d\mathbf{z} \left| \int_{\mathbb{R}^d} d\mathbf{y} D_1 D_1 \varphi(\mathbf{z} - \mathbf{y}) \cdot \Psi^{(s, y_{i-1}, y_{i+1,n}), (t, x)}_{i, (l, 1), k} (\mathbf{y}) \right| \leq \tilde{C} (s - t)^{-1/2} \left( (s - t)^{3/2} + (s - t)^{(1 + \eta)} \right) \hat{q}_{c, i} (t, s, \mathbf{x}, (y_{i-1}, y_{i+1,n})).
\]

(3.2)

In the above expression we used from condition (T_\beta) that \( j \in \{1, n\} \mapsto \beta_j (j - \frac{1}{2}) \) is increasing. Indeed, for \( j > i \), so that \( j \geq i + 1 \),

\[
\beta_j \left( j - \frac{1}{2} \right) \geq j - 1 \geq i + 1 - 1 = i \geq i - \frac{1}{2} \geq \beta_i \left( i - \frac{1}{2} \right).
\]

This is why the term \( (t - s)^{3/2} \), giving the smallest smoothing effect, appeared in the previous equation. Let us now focus on the second term in the above definition. We split therein the time integral into two parts writing:

\[
\int_0^{(s - t)^{\rho_i}} \frac{dv}{v} \left[ 1 + a_v \| \partial_v h_v \star D_i \Psi^{(s, y_{i-1}, y_{i+1,n}), (t, x)}_{i, (l, 1), k} \|_{L^1(\mathbb{R}^d, \mathbb{R})},
\]

\[
= \text{Lower} \left[ D_i \Psi^{(s, y_{i-1}, y_{i+1,n}), (t, x)}_{i, (l, 1), k} \right] + \text{Upper} \left[ D_i \Psi^{(s, y_{i-1}, y_{i+1,n}), (t, x)}_{i, (l, 1), k} \right],
\]

(3.4)

for a parameter \( \rho_i > 0 \) to be specified. The term \text{Upper} corresponding to the upper-part of the integral w.r.t. \( v \) does not involve singularities. We will use this fact to calibrate...
the associated parameter $\rho_i$ in order to match the integrability constraint
\[
\text{Upper}
\left[ D_i \Psi_{i,(l,1),k}^{(s,y_{1;-1},y_{1+1};n)}(t,x) \right] \leq \frac{\tilde{C}}{(s-t)^{1+\frac{1}{2}-\gamma}} \hat{q}_{c\setminus i}(t,s,x,(y_{1;-1},y_{1+1;1})) ,
\]
(3.5)

where $\hat{q}_{c\setminus i}$ has been defined in (2.44) and $\gamma_i > 1/2$ in order to obtain a time integrable singularity. For this term, we will only use crude upper-bounds on the derivatives of the heat-kernel and the coefficients satisfying $(T_\beta)$. On the other hand, the contribution Lower in (3.4) precisely contains the singularities w.r.t. $v$. It is therefore crucial to use there suitable cancellation tools. The point will then be to prove that the associated estimates are compatible with the upper-bound in equation (3.5).

We now write:
\[
\text{Upper}
\left[ D_i \Psi_{i,(l,1),k}^{(s,y_{1;-1},y_{1+1};n)}(t,x) \right]
= \int_{(s-t)^{\rho_i}} d\nu \frac{a_i}{v^T} \|\partial_v h_v \ast D_i \Psi_{i,(l,1),k}^{(s,y_{1;-1},y_{1+1};n)}(t,x)\|_{L^1(\mathbb{R}^d,\mathbb{R})} \\
= \int_{(s-t)^{\rho_i}} d\nu \frac{a_i}{v^T} \int_{\mathbb{R}^d} dz \left| \int_{\mathbb{R}^d} dy_i D_i \partial_v h_v(z - y_i) \Psi_{i,(l,1),k}^{(s,y_{1;-1},y_{1+1};n)}(t,x)(y_i) \right|.
\]

Recall from the definition in (2.41) that $\Psi_{i,(l,1),k}^{(s,y_{1;-1},y_{1+1};n)}(y_i)$ is $(\mathbb{R}^d)^4$-valued. To proceed with the computations we assume w.l.o.g. for the rest of the proof that $d = 1$ to avoid tensor notations for simplicity.

Writing explicitly the function $\Psi_{i,(l,1),k}^{(s,y_{1;-1},y_{1+1};n)}(y_i)$ leads to:
\[
\text{Upper}
\left[ D_i \Psi_{i,(l,1),k}^{(s,y_{1;-1},y_{1+1};n)}(t,x) \right]
\leq \int_{(s-t)^{\rho_i}} d\nu \frac{a_i}{v^T} \int_{\mathbb{R}} dz \left| \int_{\mathbb{R}} dy_i D_i \partial_v h_v(z - y_i) \left( D_{x_1} D_{x_1} \hat{p}^5(t,s,x,y) \right) \left( (\theta_{s,t}(\xi) - y) \right) \\
\left( F_i(s,y_{1;-1},\theta_{s,t}^m(\xi)) - F_i(s,\theta_{s,t}(\xi)) - D_{i-1} F_i(s,\theta_{s,t}(\xi))(y - \theta_{s,t}(\xi))_{i-1} \right) \right|_{\xi = x}.
\]

From Lemma 9 and Proposition 5, we derive that there exist $C := C((A), T) > 0$, $\tilde{C} := \tilde{C}(C)$ such that introducing $\hat{q}_c(t,s,x,y) = \hat{p}_c^x(t,s,x,y)$, $c = C^{-1}$:
\[
\text{Upper}
\left[ D_i \Psi_{i,(l,1),k}^{(s,y_{1;-1},y_{1+1};n)}(t,x) \right]
\leq \tilde{C} \int_{(s-t)^{\rho_i}} d\nu \frac{a_i}{v^T} \int_{\mathbb{R}} dz \int_{\mathbb{R}} dy_i \frac{h(x_1)}{v^T} \frac{\hat{q}_c(t,s,x,y)}{(s-t)^{1+\frac{1}{2}+\frac{1}{2}}} |(\theta_{s,t}(x) - y)| \\
\times \left\{ \sum_{j=1}^{l-1} \left| (\theta_{s,t}(x) - y) \right|^{\rho_j} \right\} + \left| (\theta_{s,t}(x) - y)_{i-1} \right|^{1+\eta}.
\]
\[
\begin{align*}
\leq & \quad \tilde{C} \int_{(s-t)^{\rho_i}}^1 dvv^{\alpha_i} \int_{\mathbb{R}} dz \int_{\mathbb{R}} dy_i \frac{h_{cv}(z-y_i)}{|v|^{\frac{1}{2}}} \tilde{q}_c(t, s, x, y) \\
& \quad \times \left\{ \sum_{j=i}^{l-1} (s-t)^{\beta_j(j-\frac{1}{2})} + (s-t)^{(1+\eta)(i-\frac{3}{2})} \right\} \\
\leq & \quad \tilde{C} \tilde{q}_{c\setminus i}(t, s, x, (y_{1:i-1}, y_{i+1:n}))(s-t)^{-\frac{1}{2}} \int_{(s-t)^{\rho_i}}^1 dvv^{\alpha_i} (s-t)^{-\frac{1}{2}} \\
& \quad \times \left\{ \sum_{j=i}^{l-1} (s-t)^{\beta_j(j-\frac{1}{2})} + (s-t)^{(1+\eta)(i-\frac{3}{2})} \right\} \\
\leq & \quad \tilde{C} \tilde{q}_{c\setminus i}(t, s, x, (y_{1:i-1}, y_{i+1:n}))(s-t)^{-\frac{1}{2}} \int_{(s-t)^{\rho_i}}^1 dvv^{\alpha_i} (s-t)^{-\frac{1}{2}} \\
& \quad \times \left\{ \sum_{j=i}^{l-1} (s-t)^{\beta_j(j-\frac{1}{2})} + (s-t)^{(1+\eta)(i-\frac{3}{2})} \right\} \\
\leq & \quad \tilde{C} \tilde{q}_{c\setminus i}(t, s, x, (y_{1:i-1}, y_{i+1:n}))(s-t)^{-\frac{1}{2}} \int_{(s-t)^{\rho_i}}^1 dvv^{\alpha_i} (s-t)^{-\frac{1}{2}} \\
& \quad \times \left\{ \sum_{j=i}^{l-1} (s-t)^{\beta_j(j-\frac{1}{2})} + (s-t)^{(1+\eta)(i-\frac{3}{2})} \right\} \\
& \quad \leq \tilde{C} \tilde{q}_{c\setminus i}(t, s, x, (y_{1:i-1}, y_{i+1:n}))(s-t)^{-\frac{1}{2}} \int_{(s-t)^{\rho_i}}^1 dvv^{\alpha_i} (s-t)^{-\frac{1}{2}} \\
& \quad \times \left\{ \sum_{j=i}^{l-1} (s-t)^{\beta_j(j-\frac{1}{2})} + (s-t)^{(1+\eta)(i-\frac{3}{2})} \right\} \\
& \quad \leq \tilde{C} \tilde{q}_{c\setminus i}(t, s, x, (y_{1:i-1}, y_{i+1:n}))(s-t)^{-\frac{1}{2}} \int_{(s-t)^{\rho_i}}^1 dvv^{\alpha_i} (s-t)^{-\frac{1}{2}} \\
& \quad \times \left\{ \sum_{j=i}^{l-1} (s-t)^{\beta_j(j-\frac{1}{2})} + (s-t)^{(1+\eta)(i-\frac{3}{2})} \right\}
\end{align*}
\]

(recalling from (3.3) that the lower bound of \(\beta_j(j-1/2)\) is increasing for the last inequality (recall indeed that we assumed that \(\beta_j \in \left( \frac{j-2}{j-1}, 1 \right] \), up to modification of \(C, \tilde{C} \) in the previous inequalities. We also used that \(\beta_i(i-1/2) \wedge (1+\eta)(i-3/2) = (1+\eta)(i-3/2)\) for the last inequality. Indeed \(\beta_i(i-1/2) > i-1\) and \(\eta\) is meant to be small.

It therefore remains to investigate the time integral part for \(v \in [0, (s-t)^{\rho_i}]\) in the thermic characterisation of the Besov norm, see (3.4). We point out that for this term it is absolutely essential to get rid of the exponent \(v^{-3/2}\) coming from the upper-bound of the thermic heat-kernel, i.e. \(D_i \partial_v h_v(z-y_i)\). In order to get an integrable singularity in \(v\), we need to decrease the crude upper-bound on \(D_i \partial_v h_v(z-y_i)\). This is done through cancellation techniques exploiting the smoothness properties of \(\Psi_{i, (l, l), k}(s, y_{1:i-1}, y_{i+1:n}, (t, x))\).

To investigate \textbf{Lower} \([D_i \Psi_{i, (l, l), k}(s, y_{1:i-1}, y_{i+1:n}, (t, x))\]) let us first recall from the definition in (2.41) that for each \(k \in [l, n]\):

\[
\begin{align*}
\Psi_{i, (l, l), k}(s, y_{1:i-1}, y_{i+1:n}, (t, x)) & \quad = \quad D_{x_j} D_{x_i} \tilde{p}^\xi(t, s, x, y) \\
& \quad \times \left[ \left( F_i(s, y_{1:i-1}, (\theta_{s,i}^j(\xi)) - F_i(s, \theta_{s,i}(\xi)) \right)
\right.
\end{align*}
\]

\[
- D_{i-1} F_i(s, \theta_{s,i}(\xi))(y - \theta_{s,i}(\xi))_{i-1} \left] \right( y - \theta_{s,i}(\xi) \right) k \end{align*}
\]

\[
\xi = x \quad (3.7)
\]

Let us now specify the dependence w.r.t. \(y_i\) of the previous expression in function of the considered indexes \(l \in [2, n], i \in [2, l-1], k \in [l, n]\). This will be useful to develop corresponding adapted cancellation arguments.
Observe first that the dependence in $y_i$ in (3.7) appears for all $l \in \mathbb{N}$, $i \in \mathbb{N}$, $k \in (l, n]$ through the term $D_{x_i}D_{x_j} \tilde{p}^\xi(t, s, x, y)$.

For the term into brackets, since $i \leq l - 1$, then $k > i$ and the only bracket term containing $y_i$ is the one associated with $F_i(s, y_1:i−1, \theta_{s,t}^{l,n}(\xi))$.  

With the notations of (3.4), we write:

\[
\begin{align*}
\text{Lower} & \left[ D_l \Psi_{i,(l,1),k}^{(s,y_1:i−1,y_{i+1:n}),(t,x)} \right] \\
& = \int_0^{(s−t)\rho_l} dv \int \mathbb{R} \left[ dy_i D_i \partial_y h_v(z − y_i) \right] \\
& \quad \cdot \left( \Psi_{i,(l,1),k}^{(s,y_1:i−1,y_{i+1:n}),(t,x)}(y_i) − \Psi_{i,(l,1),k}^{(s,y_1:i−1,y_{i+1:n}),(t,x)}(z) \right) \\
& = \int_0^{(s−t)\rho_l} dv \mathbb{R} \left[ \mathcal{T}_{1,i,(l,1),k}^{(s,y_1:i−1,y_{i+1:n}),(t,x)}(v, z) \right],
\end{align*}
\]

where:

\[
\begin{align*}
\mathcal{T}_{1,i,(l,1),k}^{(s,y_1:i−1,y_{i+1:n}),(t,x)}(v, z) & \\
& \quad := \int \mathbb{R} \left[ dy_i D_i \partial_y h_v(z − y_i) \left( D_{x_i}D_{x_j} \tilde{p}^\xi(t, s, x, y) \right) \\
& \quad \times \left( F_i(s, y_1:i−1, \theta_{s,t}^{l,n}(\xi)) − F_i(s, y_1:i−1, z, y_{i+1:j−1}, \theta_{s,t}^{l,n}(\xi)) \right)(y − \theta_{s,t}(\xi)) \right] \bigg|_{\xi = x}.
\end{align*}
\]

with a slight abuse of notation when $i = l − 1$ and

\[
\begin{align*}
\mathcal{T}_{2,i,(l,1),k}^{(s,y_1:i−1,y_{i+1:n}),(t,x)}(v, z) & \\
& \quad := \int \mathbb{R} \left[ dy_i D_i \partial_y h_v(z − y_i) \left( D_{x_i}D_{x_j} \tilde{p}^\xi(t, s, x, y) − D_{x_i}D_{x_j} \tilde{p}^\xi(t, s, x, y_{1:i−1}, z, y_{i+1:n}) \right) \\
& \quad \times \left( F_i(s, y_1:i−1, z, y_{i+1:j−1}, \theta_{s,t}^{l,n}(\xi)) − F_i(s, \theta_{s,t}(\xi)) − D_{l−1}F_i(s, \theta_{s,t}(\xi))(y − \theta_{s,t}(\xi)) \right) \right] \bigg|_{\xi = x}.
\end{align*}
\]

Write now from (3.10), Proposition 5 (see in particular (2.15) for the sensitivities in both the $x$ and $y$ variables) and Lemma 9:

\[
\begin{align*}
| \mathcal{T}_{2,i,(l,1),k}^{(s,y_1:i−1,y_{i+1:n}),(t,x)}(v, z) | & \\
& \leq C \int \mathbb{R} \left[ dy_i \frac{h_v(z − y_i)}{v^{\frac{3}{2}}} \right] \int_0^1 d\lambda.
\end{align*}
\]
\[ \dot{q}_e(t, s, x, y_{1:i-1}, z + \lambda(y_i - z), y_{i+1:n}) \]
\[ \times \left( \left| \frac{F_i(s, y_{1:i-1}, z, y_{i+1:d-1}, \theta_{s,t}(x)) - F_j(s, \theta_{s,t}(x))}{s - t} \right| \right) \]
\[ \leq \tilde{C} \int_{\mathbb{R}} dy_i \frac{h_{cv}(z - y_i)}{v} \int_0^1 d\lambda \dot{q}_e(t, s, x, y_{1:i-1}, z + \lambda(y_i - z), y_{i+1:n}) \]
\[ \times \left\{ |z - (\theta_{s,t}(x))_i|^{\beta_i} + \sum_{j=i+1}^{l-1} (s - t)^{\beta_j(j-\frac{1}{2})} + (s - t)^{(1+\eta)(i-\frac{3}{2})} \right\} , \]

where for the second inequality, we used that for \( k \geq l > i, \) \(|(y - \theta_{s,t}(x))_k| (s - t)^{-(l-1/2)} \leq |(y - \theta_{s,t}(x))_k| (s - t)^{-(k-1/2)}\) which can be absorbed by the \( k^{th} \) variables of \( \dot{q}_e. \)

Writing now for any \( \lambda \in [0, 1], \)
\[ |z - \theta_{s,t}(x)_i| \leq \lambda |z - y_i| + |z + \lambda(y_i - z) - (\theta_{s,t}(x))_i|, \]
we thus derive
\[ |\mathcal{F}^{(s, y_{1:i-1}, y_{i+1:n})_{2,i,l}, k}(t, x)(v, z)| \]
\[ \leq \tilde{C} \int_{\mathbb{R}} dy_i \frac{h_{cv}(z - y_i)}{v} \int_0^1 d\lambda \dot{q}_e(t, s, x, y_{1:i-1}, z + \lambda(y_i - z), y_{i+1:n}) \]
\[ \times \left( v^{i/2} (s - t)^{-\left(i - \frac{1}{2}\right) - \frac{1}{2}} + (s - t)^{\left(i - \frac{1}{2}\right)^{\beta_i(i - \frac{1}{2})}} \right) \]
\[ + \sum_{j=i+1}^{l-1} (s - t)^{\left(i - \frac{1}{2}\right) - \frac{1}{2} + \beta_j(j - \frac{1}{2})} + (s - t)^{\left(i - \frac{1}{2}\right) - \frac{1}{2} + (1+\eta)(i - \frac{3}{2})} \right) \]
\[ \leq \tilde{C} \dot{q}_{e \setminus i}(t, s, x, y_{1:i-1}, y_{i+1:n}) \]
\[ \times \int_0^1 d\lambda \int_{\mathbb{R}} dy_i h_{cv}(z - y_i) \mathcal{N}_{c(s-t)^{2i-1}}(z + \lambda(y_i - z) - (\theta_{s,t}(x))_i) \]
\[ \times v^{-1} \left( v^{i/2} (s - t)^{-\left(i - \frac{1}{2}\right) - \frac{1}{2}} + (s - t)^{\left(i - \frac{1}{2}\right) - \frac{1}{2} + (1+\eta)(i - \frac{3}{2})} \right) , \quad (3.11) \]
recalling for the last inequality that for any \( j \) in \([i, l-1],\) \( \beta_j(j - 1/2) > j - 1 > i - 3/2 \)
and \( \eta \) is supposed to be a small parameter.

Let us now consider the contribution defined in (3.7),
\[ |\mathcal{F}^{(s, y_{1:i-1}, y_{i+1:n})_{1,i,l}, k}(t, x)(v, z)| \]
\[ \leq \tilde{C} \int_{\mathbb{R}} dy_i \frac{h_{cv}(z - y_i)}{v^{i/2}} \dot{q}_e(t, s, x, y) \times (s - t)^{\left(i - \frac{1}{2}\right) + \frac{1}{2}} |z - y_i|^{\beta_i} |(y - \theta_{s,t}(x))_k| \]
\[ \Box \text{ Springer} \]
\[
\leq \tilde{C} \int_{\mathbb{R}} d\bf{y}_t \frac{h_{cv}(z - \bf{y}_t) \hat{q}_c(t, s, \bf{x}, \bf{y})}{\nu - \nu^2} \frac{1}{(s - t)^{1/2}}. 
\] (3.12)

From (3.11) and (3.12) we derive, with the notation introduced in (2.44):

\[
\int_{\mathbb{R}} dz \left| \mathcal{F}_1(t, s, \bf{x}, (\bf{y}_1:i-1, \bf{y}_i+1:2)) + \mathcal{F}_2(t, s, \bf{x}) \right| (v, z) \\
\leq \tilde{C} \hat{q}_{c|\{i}(t, s, \bf{x}, (\bf{y}_1:i-1, \bf{y}_i+1:2)) \left( \frac{1}{\nu - \nu^2} \frac{1}{(s - t)^{1/2}} + \frac{1}{(s - t)^i} + \frac{1}{(s - t)^{i-(1+\eta)}(i^{3/2})} \right) \\
\times \int_{0}^{1} d\lambda \int_{\mathbb{R}} dz \int_{\mathbb{R}} d\bf{y}_t h_{cv}(z - \bf{y}_t) \mathcal{N}_{c(s-t)^{2i-1}}(z + \lambda(y_i - z) - (\theta_{s,t}(\bf{x}))_i) \\
\leq \tilde{C} \hat{q}_{c|\{i}(t, s, \bf{x}, (\bf{y}_1:i-1, \bf{y}_i+1:2)) \left( \frac{1}{\nu - \nu^2} \frac{1}{(s - t)^{1/2}} + \frac{1}{(s - t)^i} + \frac{1}{(s - t)^{i-(1+\eta)}(i^{3/2})} \right), 
\]

using the change of variable \((w_1, w_2) = (z - y_i, z + \lambda(y_i - z) - (\theta_{s,t}(\bf{x}))_i)\) for the last inequality.

One gets from the definition in (3.4):

\[
\text{Lower} \left[ D_t \Psi_{i,\{i}(s, \bf{y}_1:i-1, \bf{y}_i+1:2), (t, \bf{x}) \right] \\
\leq \tilde{C} \hat{q}_{c|\{i}(t, s, \bf{x}, (\bf{y}_1:i-1, \bf{y}_i+1:2)) \int_{0}^{(s-t)^{\rho_i}} d\nu \frac{a_i}{\nu^{1/2}} \\
\times \left( \frac{1}{(s - t)^{1/2}} + \frac{1}{(s - t)^i} + \frac{1}{(s - t)^{i-(1+\eta)}(i^{3/2})} \right) \\
\leq \tilde{C} \hat{q}_{c|\{i}(t, s, \bf{x}, (\bf{y}_1:i-1, \bf{y}_i+1:2)) \\
\times \left( (s - t)^{\rho_i} a_i + \frac{1}{2} - \frac{1}{2} + (s - t)^{\rho_i} a_i + \frac{\beta_i - 1}{2} - i + (s - t)^{\rho_i} a_i + \frac{\beta_i - 1}{2} + (s - t)^{\rho_i} a_i + \frac{\beta_i - 1}{2} + (s - t)^{\rho_i} a_i + \frac{\beta_i - 1}{2} + (s - t)^{\rho_i} a_i + \frac{\beta_i - 1}{2} + \eta \left( i - \frac{3}{2} \right) \right), 
\] (3.13)

where we have assumed that \(\alpha_i\) is s.t. \(\alpha_i + \beta_i > 1\) to guarantee the integrability of the first above integrand. From (3.1) together with (3.2) and (3.4), combining the inequalities (3.13) and (3.6) we derive:

\[
\left\| D_t \Psi_{i,\{i}(s, \bf{y}_1:i-1, \bf{y}_i+1:2), (t, \bf{x}) \right\|_{B_1} \leq \tilde{C} \hat{q}_{c|\{i}(t, s, \bf{x}, (\bf{y}_1:i-1, \bf{y}_i+1:2)) (s - t)^{-1+\lambda}, 
\]

where

\[
\lambda = \min \left( \frac{1}{2} + \frac{\alpha_i}{2} \right) \rho_i + \frac{1}{2} + (1 + \eta) \left( i - \frac{3}{2} \right), \rho_i \frac{\alpha_i + \beta_i - 1}{2} + \frac{1}{2}, \\
\rho_i \frac{\alpha_i + \beta_i}{2} - i + 1, \rho_i \frac{\alpha_i - 1}{2} + \eta \left( i - \frac{3}{2} \right) \right) . 
\] (3.14)
To maximize (in $\rho_i$ for fixed $\alpha_i$) the minimum of the first and last term above, one has to take:

$$
\left[ -\frac{1}{2} + \frac{\alpha_i}{2} \right] \rho_i + \frac{1}{2} + (1 + \eta) \left( i - \frac{3}{2} \right) = \rho_i \frac{\alpha_i}{2} - \frac{1}{2} + \eta \left( i - \frac{3}{2} \right),
$$

which implies $\rho_i = 2i - 1$. It is immediate to see that by definition of $\beta_i$, $\eta$, for such choice of $\rho_i$ we also have:

$$
\rho_i \frac{\alpha_i + \beta_i - 1}{2} + \frac{1}{2} = \rho_i \frac{\alpha_i + \beta_i}{2} - i + 1 > \rho_i \frac{\alpha_i}{2} - \frac{1}{2} + \eta \left( i - \frac{3}{2} \right).
$$

Thus,

$$
\lambda = \left( i - \frac{1}{2} \right) \alpha_i - \frac{1}{2} + \eta \left( i - \frac{3}{2} \right).
$$

To ensure the condition $\alpha_i + \beta_i > 1$ it suffices to take $\alpha_i = 1 - \frac{2i - 2}{2i - 1} = \frac{1}{2i - 1}$, which finally gives:

$$
\lambda = \eta \left( i - \frac{3}{2} \right).
$$

Recalling (3.14), we get the desired bound of Lemma 10. \qed

### 3.2 Proof of Lemma 11

We now tackle the Besov estimate of the derivative of the solution of (1.16). Fix $t \in [0, T]$ and $(x_{1; i - 1}, x_{i + 1}; n) \in \mathbb{R}^{(n-1)d}$.

From the thermic characterization of Besov spaces recalled in equation (3.1), we actually have to control:

$$
\|D_k u(t, x_{1; i - 1}, \cdots, x_{i + 1}; n)\|_{B_{\infty, \infty}^{\alpha_i}} = \|D_k u(t, x_{1; i - 1}, \cdots, x_{i + 1}; n)\|_{C^{\alpha_i}} := \|D_k u(t, x_{1; i - 1}, \cdots, x_{i + 1}; n)\|_{\infty} + \|D_k u(t, x_{1; i - 1}, \cdots, x_{i + 1}; n)\|_{\infty} \alpha_i. \tag{3.15}
$$

Observe first that $\|D_k u(t, x_{1; i - 1}, \cdots, x_{i + 1}; n)\|_{\infty} \leq \|Du\|_{\infty}$. Hence, we can focus on the Hölder modulus in (3.15). For given $x_i, z \in \mathbb{R}^d$ we want to control the difference:

$$
D_{x_i} u(t, x_{1; i - 1}, x, x_{i + 1}; n) - D_{x_i} u(t, x_{1; i - 1}, z, x_{i + 1}; n) \tag{3.16}
$$

through the expansion of the gradients given by (2.27). Two cases then arise: the system is globally in the off-diagonal regime, i.e. the spatial points $x_i$ and $z$ are far w.r.t. their corresponding time scale (there exists $c_0$ such that $c_0^{i-1/2}|x_i - z| \geq (T - t)^{i-1/2}$ or equivalently $c_0|x_i - z|^2/(2i - 1) \geq (T - t)$); the system is globally in the diagonal...
regime, i.e. the spatial points $x_i$ and $z$ are close w.r.t. their corresponding time scale $(c_0|x_i - z|^2/(2i - 1) < (T - t))$.

Since in the global off-diagonal regime the spatial points are far, it is not expectable to control suitably the expansion of the gradients around their difference.

In this case, it is in fact more natural to expand each gradient term thanks to (2.27) taking as freezing point the associated spatial argument, i.e. $\xi = x$ for the first gradient and denoting by $x' = (x_{1:i-1}, z, x_{i+1:n})$, $\xi' = x'$ for the second one. This allows to take advantage of the underlying smoothing properties in time of the gradient (cf. Sect. 3.2.1 below). On the other hand, in the global diagonal regime (when $c_0|x_i - z|^2/(2i - 1) < (T - t)$), we are again faced with a regime dichotomy. Note indeed that, expanding the gradients in (3.16) from (2.27), we have to deal with a time integral associated with the source term, i.e. the spatial points $x_i$ and $z$ are close w.r.t. their corresponding time scale $(c_0|x_i - z|^2/(2i - 1) < (T - t))$.

Concerning the local diagonal regime, the proximity of the spatial points suggests to expand the gradients through a Taylor expansion. Starting again from their corresponding representation of (2.27), it is natural to consider similar spatial freezing points. Such a strategy indeed yields to only consider spatial sensitivities of the underlying Gaussian proxy (see Sect. 3.2.2). Observe that keeping the two distinct freezing points would lead to investigate the full sensitivity between two different proxys, including the sensitivity of the corresponding covariance matrix and generator. Such an investigation appears to be quite involved. Furthermore, we did not succeed to make it work.

With our approach, we are led to expand one of the gradients in (3.16) around two different freezing points. Such a strategy was already used in the companion paper [12] and leads to consider an additional boundary term arising precisely from the change of freezing point (see Sect. 3.2.3). Namely, we will expand the term $D_{\xi}u(t, x_{1:i-1}, x_i, x_{i+1:n})$ with (2.27) taking $\xi = x$, whereas we will expand differently the contribution $D_{\xi'}u(t, x_{1:i-1}, z, x_{i+1:n})$, depending on the considered (local) regime (off-diagonal or diagonal) for the current running time. For parameters $\xi', \xi'' \in \mathbb{R}^d, t_0 \in (s, T]$ to be fixed later on, we actually have:

$$ u(t, x') = \int_t^T ds \left( \mathbb{I}_{s \leq t_0} \left[ \tilde{P}_{s,t}^{\xi'} f(s, \cdot) \right](x') + \mathbb{I}_{s > t_0} \left[ \tilde{P}_{s,t}^{\xi''} f(s, \cdot) \right](x') \right) $$

$$ + \left[ \tilde{P}_{t_0,t}^{\xi'} u(t_0, \cdot) \right](x') - \left[ \tilde{P}_{t_0,t}^{\xi''} u(t_0, \cdot) \right](x') $$

Note that in this case we have that for any $s$ in $(t, T]$ the off-diagonal regime holds.
\[ + \int_t^T ds \left( \mathbb{I}_{s \leq t_0} \left[ \tilde{P}_s^{\xi'}(L_s - \tilde{L}_s^{\xi'})u(s, \cdot) \right](\mathbf{x}') \right. \\
\left. + \mathbb{I}_{s > t_0} \left[ \tilde{P}_s^{\xi} (L_s - \tilde{L}_s^{\xi}) u(s, \cdot) \right](\mathbf{x}') \right) \]  
\hspace{1cm} (3.17)

We refer to Appendix B below for a proof of expansion (3.17) (see also Section 2.4 of the “Detailed guide to the proof” in [12]). We again emphasize that in comparison with (2.27), the additional discontinuity term \[ \tilde{P}_t^{\xi}(L_t - \tilde{L}_t^{\xi})u(t_0, \cdot) \] and the dichotomy in the time integrals in the r.h.s. of the above equation are the price to pay to consider different freezing points associated with the corresponding local off-diagonal and diagonal regimes.

From now on, we will assume w.l.o.g. that \( c_0 \) is a constant meant to be small (see Sect. 3.2.3, Lemma 16 and its proof). We also suppose that \(|\mathbf{x}' - \mathbf{x}| \leq 1\) since otherwise the global off-diagonal regime holds, and the analysis of Sect. 3.2.1 applies.

Starting from (3.16), we expand \( D_{x_k}u(t, \mathbf{x}) = D_{x_k}u(t, \mathbf{x}_{i;i-1}, \mathbf{x}_i, \mathbf{x}_{i+1:n}) \) using (2.27) with \( \xi = \mathbf{x} \) and we expand \( D_{x_k}u(t, \mathbf{x}_{i;i-1}, z, \mathbf{x}_{i+1:n}) \) by differentiating (3.17) w.r.t. \( x_k \) and setting then \( \xi' = \mathbf{x}', \xi = \mathbf{x} \) and \( t_0 = t + c_0|\mathbf{x}_i - z|^{2/(2i-1)} \) (characteristic cutting time for the change of regime). We thus rewrite with the previous notations:

\[
\begin{aligned}
D_{x_k}u(t, \mathbf{x}) &- D_{x_k}u(t, \mathbf{x}') =: \left( D_{x_k}u(t, \mathbf{x}) - D_{x_k}u(t, \mathbf{x}') \right)|_{S_l} \\
\quad &+ \left( D_{x_k}u(t, \mathbf{x}) - D_{x_k}u(t, \mathbf{x}') \right)|_{S_l^c} \\
\quad &- D_{x_k}u(t, \mathbf{x}')|_{\partial S_l},
\end{aligned}
\]

(3.18)

where, with the notation of (2.27):

\[
\begin{aligned}
\left( D_{x_k}u(t, \mathbf{x}) - D_{x_k}u(t, \mathbf{x}') \right)|_{S_l} &= \int_{S_l} ds \left\{ [H_k^\xi(s, \mathbf{x}) - H_k^\xi(s, \mathbf{x}')] + [I_k^\xi(s, \mathbf{x}) - I_k^\xi(s, \mathbf{x}')] \right\}|_{(\xi, \xi')=(\mathbf{x}, \mathbf{x}')}, \\
\quad &|_{(\xi, \xi')=(\mathbf{x}, \mathbf{x}')},
\end{aligned}
\]

(3.19)

corresponds to the difference of the previous expansions on the off-diagonal regime,

\[
\begin{aligned}
\left( D_{x_k}u(t, \mathbf{x}) - D_{x_k}u(t, \mathbf{x}') \right)|_{S_l^c} &= \int_{S_l^c} ds \left\{ [H_k^\xi(s, \mathbf{x}) - H_k^\xi(s, \mathbf{x}')] + [I_k^\xi(s, \mathbf{x}) - I_k^\xi(s, \mathbf{x}')] \right\}|_{(\xi, \xi')=(\mathbf{x}, \mathbf{x}')}, \\
\quad &|_{(\xi, \xi')=(\mathbf{x}, \mathbf{x}')},
\end{aligned}
\]

(3.20)

is the contribution of the diagonal regime and

\[
D_{x_k}u(t, \mathbf{x}')|_{\partial S_l} := \left\{ D_{x_k} \tilde{P}_{t_0}^{\xi'}u(t_0, \mathbf{x}') - D_{x_k} \tilde{P}_{t_0}^{\xi}u(t_0, \mathbf{x}') \right\}|_{(\xi, \xi')=(\mathbf{x}', \mathbf{x})},
\]

(3.21)

is the resulting boundary term. This last term, arising from the change of freezing point, is particularly delicate to analyze.
3.2.1 Off-diagonal estimates: control of \((3.19)\)

On the time set \(S_i\), we cannot expect some regularization from the difference of the transition densities so that we bluntly estimate the terms appearing in \((3.19)\), writing:

\[
|D_{x_i} u(t, x) - D_{x_i} u(t, x')|_{S_i} \leq \int_{S_i} ds \left( |H^\xi_k(s, x)|_{x = x'} + |H^\xi_k(s, x')|_{x' = x'} + |I^\xi_k(s, x)|_{x = x'} + |I^\xi_k(s, x')|_{x' = x'} \right)
\]

(3.22)

Those terms can then be handled following the previous analysis performed in Theorem 3 and Lemma 10, observing here that, w.r.t. the previous proofs, the above terms are not differentiated w.r.t. \(x_1\). This improves the exponents of the time singularities of 1/2. Similarly to (2.28), (2.30), (2.32) and (2.33) this therefore yields for the terms \(H^\xi_k(s, x)\):

\[
|H^\xi_k(s, x)|_{x = x} \leq C(s - t)^{-1 + \delta_k} \left(1 + \|D u\|_\infty + \|D D_1 u\|_\infty\right),
\]

(3.23)

with \(1 > \delta_k \geq 1/2 + \eta/2 > 1/2\).

Reproducing the arguments that led to equations (2.36), (2.37), (2.43)-(2.47) and the statement of Lemma 10, exploiting again that there is now no differentiation w.r.t. \(x_1\) we get with the notations of (1.2):

\[
|I^\xi_k(s, x)|_{x = x} \leq C(s - t)^{-1 + \inf_{j \in \{2, k\} \backslash \gamma_j} \|D u\|_\infty + \sup_{\hat{s} \in [0, T]} \sup_{j \in \{2, k\}} \|[D_k u]_{\hat{s}}(\cdot)\|_{\alpha_j}).
\]

(3.24)

Similar bounds hold for \(H^\xi_k(s, x')\) and \(I^\xi_k(s, x')\). Hence, from (3.23) and (3.24),

\[
\int_{S_i} ds \left( |H^\xi_k(s, x)|_{x = x'} + |H^\xi_k(s, x')|_{x' = x'} + |I^\xi_k(s, x)|_{x = x'} + |I^\xi_k(s, x')|_{x' = x'} \right)
\]

\[
\leq C \left( \int_{t}^{(t+c_0|x_i-z|^{2\gamma})/T} (s - t)^{-1 + \delta_k} \left(\|D u\|_\infty + \|D D_1 u\|_\infty\right) + (s - t)^{-1 + \inf_{j \in \{2, k\} \backslash \gamma_j} \|D u\|_\infty + \sup_{\hat{s} \in [0, T]} \sup_{j \in \{2, k\}} \|[D_k u]_{\hat{s}}(\cdot)\|_{\alpha_j}) \right)
\]

\[
\leq C \left( |x_i - z|^{2\delta_k} \left(\|D u\|_\infty + \|D D_1 u\|_\infty\right) + |x_i - z|^{2inf_{j \in \{2, k\} \backslash \gamma_j} \|D u\|_\infty + \sup_{\hat{s} \in [0, T]} \sup_{j \in \{2, k\}} \|[D_k u]_{\hat{s}}(\cdot)\|_{\alpha_j}) \right)
\]

\(\square\) Springer
\[ C|\mathbf{x}_i - z|^{\alpha_1} T^\delta \left( \|D u\|_\infty + \|D D_1 u\|_\infty + \sup_{j \in [1, k]} \sup_{t \in [0, T]} \left( (D_k u)_{j}(\tilde{s}, \cdot) \right)_{|\tilde{\alpha}_j|} \right), \]

(3.25)

for some \( \delta' := \delta'(A) > 0 \), recalling for the last inequality that \( \gamma_j = 1/2 + \eta(j - 3/2) \) so that \( 2\gamma_j/(2i - 1) = [1 + 2\eta(j - 3/2)]/(2i - 1) > 1/(2i - 1) = \alpha_i \) (see also the statements of Lemmas 10 and 11) and similarly for the contributions involving \( \delta_k > 1/2 \).

We eventually get from (3.25) and (3.22):

\[ |D_{x_k} u(t, \mathbf{x}) - D_{x_k} u(t, \mathbf{x}')|_{S_1^c} \leq C |\mathbf{x}_i - z|^{\alpha_1} T^\delta' \left( \|D u\|_\infty + \|D D_1 u\|_\infty + \sup_{j \in [1, k]} \sup_{t \in [0, T]} \left( (D_k u)_{j}(\tilde{s}, \cdot) \right)_{|\tilde{\alpha}_j|} \right). \]

(3.26)

### 3.2.2 Diagonal estimates: control of the term (3.20)

We consider here the difference \( D_{x_k} u(t, \mathbf{x}) - D_{x_k} u(t, \mathbf{x}') \) of (3.16) on \( S_1^c \). In that case the points \( \mathbf{x} \) and \( \mathbf{x}' \) are close w.r.t. the characteristic time scale of the \( i \)th variable and the main idea consists in controlling the difference between the frozen densities at \( \xi = \tilde{\xi}' = \mathbf{x} \) with starting points \( \mathbf{x} \) and \( \mathbf{x}' \) respectively. We first state the following lemma whose proof is postponed to Appendix C.2.

**Lemma 14** (Hölder controls for the difference of the gradients of the frozen density (diagonal regime)). Assume that \( \mathbf{x}, \mathbf{x}' \) only differ in their \( i \)th component and that \( s \in S_1^c \). Then there exists \( C \geq 1 \) s.t. for all \( k \in [1, n] \), \( \beta \in (0, 1) \),

\[ |D_{x_k} D_{y_1}^\beta \hat{p}^\xi (t, s, \mathbf{x}, \mathbf{y}) - D_{x_k} D_{y_1}^\beta \hat{p}^\xi (t, s, \mathbf{x}', \mathbf{y})| \leq \frac{C |(\mathbf{x}' - \mathbf{x})_i|^{\alpha_1} \tilde{\rho}^\xi}{(s - t)_1 + (k - 1) + \frac{2}{z}} \]

\[ (s - t)_{\frac{1}{2} + (k - 1) + \frac{2}{z}} \hat{p}^\xi_{c-1} (t, s, \mathbf{x}, \mathbf{y}). \]

(3.27)

Write now from (3.20), recalling that \( \tilde{\xi}' = \tilde{\xi} \),

\[ |D_{x_k} u(t, \mathbf{x}) - D_{x_k} u(t, \mathbf{x}')|_{S_1^c} := \left| \int_{S_1^c} ds \left[ H^\xi_k (s, \mathbf{x}) - H^\xi_k (s, \mathbf{x}') \right] \right| \xi = \mathbf{x} \]

(3.28)

and let us discuss how the terms \( H^\xi_k (s, \mathbf{x}) - H^\xi_k (s, \mathbf{x}') \), \( I^\xi_k (s, \mathbf{x}) - I^\xi_k (s, \mathbf{x}') \) in the above equation can be handled.

We first focus on the term \( I^\xi_k (s, \mathbf{x}) - I^\xi_k (s, \mathbf{x}') \) in (3.28). This contribution, associated with the degenerate components of perturbed operator, is again the most delicate to handle. From the definitions in (2.35) we are led to control the sum \( \sum_{\ell=1}^3 |I_{k, \ell}^\xi (s, \mathbf{x}) -
For the terms $I_{k,1}^{\xi}(s, x) - I_{k,1}^{\xi}(s, x')$, $I_{k,2}^{\xi}(s, x) - I_{k,2}^{\xi}(s, x')$ we are going to reproduce the analysis leading to (2.36), (2.37). Observe first that the above terms do not involve $D_{x_1}$, therefore we gain a singularity of order $1/2$ w.r.t. the indicated equations (2.36), (2.37). On the other hand, the difference of the derivatives of the frozen densities w.r.t. $x_k$ can be handled with (3.27) (for $\beta = 0$). This leads to:

$$
\left| \sum_{\ell=1}^{2} \int_{T} ds \left( I_{k,\ell}^{\xi}(s, x) - I_{k,\ell}^{\xi}(s, x') \right) \right|_{\xi=x} \leq C |(x' - x)_i|^{\alpha_i} \|D u\|_{\infty} \sum_{j=k}^{n} \int_{T} ds (s - t)^{-\frac{1}{2} - (k - \frac{1}{2})} \times \left( (s - t)^{\beta_j} (j^{-\frac{1}{2}}) + (s - t)^{(1 + \eta)(j^{-\frac{1}{2}})} \right),
$$

changing the summation variables from (2.36) for notational simplicity.

From (T$_{\beta}$), which yields that $\beta_j (j - 1/2) > j - 1$, we derive

$$
\left| \sum_{\ell=1}^{2} \int_{T} ds \left( I_{k,\ell}^{\xi}(s, x) - I_{k,\ell}^{\xi}(s, x') \right) \right|_{\xi=x} \leq C |(x' - x)_i|^{\alpha_i} \|D u\|_{\infty} T^{\delta},
$$

(3.29)

for some $\delta > 0$.

From the previous analysis it is therefore sufficient to focus on the tricky term, namely $I_{k,3}^{\xi}(s, x)$ introduced in (2.35). Exploiting as well Lemma 8 for a centering argument w.r.t. the $k$th variable, recall $i \leq k$, we write:

$$
I_{k,3}^{\xi}(s, x) - I_{k,3}^{\xi}(s, x')
= \sum_{\ell=2}^{k} \int_{\mathbb{R}^{nd}} dy \left( (F_\ell(s, y_{1:k-1}, \theta_{s,t}^{k:n}(\xi)) - F_\ell(s, \theta_{s,t}(\xi)) - D_\ell F_\ell(s, \theta_{s,t}(\xi))(y - \theta_{s,t}(\xi))_{\ell-1},

\begin{align*}
&\left( D_{y_{1:k-1}} F_\ell(s, y_{1:k-1}, \theta_{s,t}^{k:n}(\xi)) - D_{y_{1:k-1}} u(s, y_{1:k-1}, \theta_{s,t}^{k:n}(\xi)) \right) \\
&\quad (D_{x_k} \tilde{p}_\xi(t, s, x, y) - D_{x_k} \tilde{p}_\xi(t, s, x', y)).
\end{align*}
$$

Let us reproduce now the arguments used in Sect. 2.2 to handle $I_{k,3}^{\xi}$ (see e.g. the computations from equation (2.39) to (2.40)). For $\ell \in \llbracket 2, k - 1 \rrbracket$, expanding with the Taylor formula the difference $\left( D_{y_{1:k-1}} F_\ell(s, y_{1:k-1}, \theta_{s,t}^{k:n}(\xi)) \right)$, using the Schwarz theorem to exchange the order of differentiations$^{10}$, and integrating by parts

$^{10}$ Recall indeed that what we are able to control is precisely the Hölder moduli of the derivatives $D_{y_m} u(s, \cdot)$ w.r.t. the variables $\ell \leq m$. 

---

$\text{Springer}$
as well, we obtain:

\[ I_{k,3}^s(s, x) - I_{k,3}^s(s, x') = \sum_{\ell=2}^{k-1} \sum_{m=k}^{n} \int_0^1 d\lambda \int_{\mathbb{R}^d} dy D_{y\ell} \left[ \left( F_\ell(s, y_{1:k-1}, \theta_{s,t}^{kn}(\xi)) - F_\ell(s, \theta_{s,t}(\xi)) \right)(y - \theta_{s,t}(\xi))_{\ell=1} \right] \]

\[ =: \Delta_{k,31}^s(s, x, x') + \Delta_{k,32}^s(s, x, x'). \]  

Let us now modify slightly the definition of (2.41) and introduce: for all \( \ell \in [2, k-1] \), \( m \in [k, n] \), \((y_{1:\ell-1}, y_{\ell+1:n}) \in \mathbb{R}^{(n-1)d}\), \((t, z, x) \in [0, T] \times (\mathbb{R}^d)^2\), \( s \in (t, T] \) and \( y_i \in \mathbb{R}^d\):

\[ \Psi_x^{s,y_{1:\ell-1},y_{\ell+1:n}}(t, z, x) (y_\ell) \]

\[ = D_{x\ell} \cdot \frac{d}{d\ell} \left[ \left( F_\ell(s, y_{1:k-1}, \theta_{s,t}^{kn}(\xi)) - F_\ell(s, \theta_{s,t}(\xi)) \right)(y - \theta_{s,t}(\xi))_{\ell=1} \right] ((y - \theta_{s,t}(\xi))_{m=1})^{\ell} \]  

Write now with the notations of (3.31):

\[ (\Delta_{k,31}^s(s, x, x'))_{\xi=x} = \sum_{\ell=2}^{k-1} \sum_{m=k}^{n} \int_0^1 d\lambda \int_{\mathbb{R}^d} dy D_{y\ell} \left[ \Psi_x^{s,y_{1:\ell-1},y_{\ell+1:n}}(t, x, x) (y_\ell) \right] \]

\[ =: \sum_{\ell=2}^{k-1} \sum_{m=k}^{n} \int_0^1 d\lambda \int_{\mathbb{R}^d} dy D_{y\ell} \left[ \Psi_y^{s,y_{1:\ell-1},y_{\ell+1:n}}(t, x', x) (y_\ell) \right] - \Psi_x^{s,y_{1:\ell-1},y_{\ell+1:n}}(t, x', x) (y_\ell) \]

\[ = D_{y\ell} u(s, y_{1:k-1}, \theta_{s,t}^{kn}(x) + \lambda (y - \theta_{s,t}(x))_{k:n}). \]
Thus, we derive similarly to (2.43):

$$\begin{align*}
|\Delta_{k,31}^\xi (s, x, x')|_{\xi = x}
\leq \sum_{k-1} \sum_{n} \int_{\mathbb{R}^{(n-1)d}} d(y_{1:1}, y_{1:n}) \\
\left\{ \left\| D_{\ell} \left[ \Psi_{\ell,k,m}^{(s,y_{1:1},y_{1:n}), (t,x,x')}, (t,x,x') \right] \right\|_{B^{-\alpha_{\ell}}_{1,1}} \times \sup_{\xi, j \in [1,n], j \neq \ell} \left\| D_{\ell} u(s, z_{1:1}, \cdot, z_{1:n}) \right\|_{B_{\infty,\infty}^{\alpha_{\ell}}} \right\},
(3.32)
\end{align*}$$

where $\int_{\mathbb{R}^{(n-1)d}} d(y_{1:1}, y_{1:n})$ means that we integrate over $y_{1:1}$ and $y_{1:n}$. To conclude, we need the following appropriate version of Lemma 10 to handle the Besov norm with negative exponent in the above r.h.s. Its proof is postponed to the next section.

**Lemma 15** Let $k \in \{2, n\}$, $\ell \in \{2, k-1\}$ and $m \in \{k, n\}$ and let $\Psi_{\ell,k,m}^{(s,y_{1:1},y_{1:n}), (t,x,x)} : \mathbb{R}^{d} \to \mathbb{R}^{d}$ be the function defined by (3.31). There exist $C := C((A)) > 0$ and $\gamma_{\ell} := \gamma_{\ell}((A)) := 1/2 + \eta(\ell - 3/2) > 1/2$ such that

$$\begin{align*}
\left\| D_{\ell} \left[ \Psi_{\ell,k,m}^{(s,y_{1:1},y_{1:n}), (t,x,x)} - \Psi_{\ell,k,m}^{(s,y_{1:1},y_{1:n}), (t,x,x)} \right] \right\|_{B^{-\alpha_{\ell}}_{1,1}}
\leq C \hat{q}_{\ell}(t, s, x, (y_{1:1}, y_{1:n}))(s - t)^{-3/2 + \gamma_{\ell}} |(x - x')_{i}|^{\alpha_{\ell}},
\end{align*}$$

with $\hat{q}_{\ell}(t, s, x, (y_{1:1}, y_{1:n}))$ as in (2.44).

We derive from Lemma 15 and (3.32) that:

$$\begin{align*}
\int_{S_{i}} ds |\Delta_{k,31}^\xi (s, x, x')|_{\xi = x} \\
\leq C \left( \sum_{k-1} \sum_{n} \sup_{\ell = 2, m = k, s \in [0,T]} \left\| (D_{\ell} u)_{t} (s, \cdot) \right\|_{B^{\alpha_{\ell}}_{\infty,\infty}} \int_{t} ds (s - t)^{-3/2 + \gamma_{\ell}} ||(x - x')_{i}|^{\alpha_{\ell}} \right) \\
\leq C T^{\delta} \sup_{m \in [k,n], \ell \in [1,k], s \in [0,T]} \left\| (D_{\ell} u)_{t} (s, \cdot) \right\|_{B^{\alpha_{\ell}}_{\infty,\infty}} ||(x - x')_{i}|^{\alpha_{\ell}},
(3.33)
\end{align*}$$

for some $\delta := \eta/2 > 0$. On the other hand, from the definition in (3.30), using as well (3.27) with $\xi = x$ and $\beta = 0$, Lemmas 9, 13 and 7, we get similarly to (2.46):

$$\begin{align*}
|\Delta_{k,32}^\xi (s, x, x')|_{\xi = x} \\
\leq \frac{C[(D_{\ell-1} F_{k-1}) (s, \cdot)]_{\eta}}{(s - t)^{(k-1) + \frac{k}{2} + \frac{1}{2}}}(x - x')_{i}|^{\alpha_{\ell}} \int_{\mathbb{R}^{d}} \hat{p}_{\ell,1}^{x} (t, s, x, y) |(\theta_{s,t}(x) - y)|_{k-1}^{1+\eta}.
\end{align*}$$
\[ \times (\|Du\|_\infty + [(D_ku)_k(s,\cdot)]_{\alpha_k}) (|\theta_{s,t}(x) - y_k|^{\alpha_k} + \sum_{m=k+1}^n |\theta_{s,t}(x) - y_m|^{\alpha_k}) \]

\[ \leq C(\|Du\|_\infty + [(D_ku)_k(s,\cdot)]_{\alpha_k}) (|x - x'|_i)^{\alpha_i} \]

\[ \times (\sum_{m=k}^n (s - t)^{-\frac{\nu}{2} - \gamma_k + \frac{\nu}{2} + \frac{\nu}{2} - \frac{\nu}{2} - \frac{\nu}{2} + \frac{\nu}{2}} + \frac{\nu}{2} + \frac{\nu}{2})_{m=k+1,\alpha_k}). \quad (3.34) \]

From (3.34) and recalling (2.48) and (2.49) we finally get:

\[ \int_{S_i}^T ds |\Delta_{k,32}^\xi (s, x, x')|_{\xi = x} \]

\[ \leq C(\|Du\|_\infty + [(D_ku)_k(s,\cdot)]_{\alpha_k}) (|x - x'|_i)^{\alpha_i} \]

\[ \sum_{m=k}^n \int_T^{T'} ds (s - t)^{-\frac{\nu}{2} - \gamma_k} \]

\[ \leq C T^\delta (\|Du\|_\infty + [(D_ku)_k(s,\cdot)]_{\alpha_k}) (|x - x'|_i)^{\alpha_i}. \quad (3.35) \]

Combining the estimates (3.33) and (3.35) together with (3.29) we eventually derive

\[ \left| \int_{S_i}^T ds \left[ I_k^\xi (s, x) - I_k^\xi (s, x') \right] \right|_{\xi = x} \]

\[ \leq T^\delta C (|x' - x|_i)^{\alpha_i} \|Du\|_\infty + \sup_{m \in [k,n], \ell \in [1,k], s \in [0,T]} \| (D_mu)_\ell (s,\cdot) \|_{\delta_t} \right). \quad (3.36) \]

Now, the term \( H_k^\xi (s, x) - H_k^\xi (s, x') \) in (3.28) (non-degenerate variables) can be handled reproducing the same previous arguments for \( I_k^\xi (s, x) - I_k^\xi (s, x') \), exploiting Lemma 14 and following the computations performed for \( H_k \) in the proof of Theorem 3 (see e.g. (2.34)). From the definition of \( S_i^\xi \) we obtain:

\[ \left| \int_{T' + t}^T ds \left( H_k^\xi (s, x) - H_k^\xi (s, x') \right) \right|_{\xi = x} \]

\[ \leq C (|x' - x|_i)^{\alpha_i} (\|DD_1u\|_\infty + \|Du\|_\infty) \int_t^{T'} ds (s - t)^{-\frac{\nu}{2} - \delta_k} \]

\[ \leq C (|x' - x|_i)^{\alpha_i} (\|DD_1u\|_\infty + \|Du\|_\infty) T^\delta, \quad (3.37) \]

for some \( \delta > 0 \) recalling for the last inequality that, from the bound following (3.23), \( \delta_k \geq 1/2 + \eta/2 > 1/2 \). The arguments needed to control this term are actually those already exploited in [10] when \( n = 2 \).
We now split the above contribution into three terms:

\[ |D_{x_k} u(t,x) - D_{x_k} u(t,x')| \big| D_{x_i} \]

\[ \leq C T^{\frac{1}{2}} |(x' - x)_i| a_i \left( \| D D_1 u \|_\infty + \| D u \|_\infty + \sup_{2 \leq \ell \leq m \leq n, s \in [0,T]} \| (D_{m} u)_{\ell} (s, \cdot) \|_{B_{\infty, \infty}^\alpha} \right). \]

(3.38)

### 3.2.3 Discontinuity term associated with the regime time change: control of the term (3.21)

We here aim at handling

\[ D_{x_k} u(t,x') \big|_{\partial S_i} = \left\{ D_{x_k} \tilde{p}_{t_0,t}^{x'} u(t_0, x') - D_{x_k} \tilde{p}_{t_0,t}^{x'} u(t_0, x') \right\}_{(x', x')} \]

which we will actually handle like the off-diagonal components.

We mention that for the analysis below, similarly to what we already did for the density at the beginning of Sect. 2.1.2, we will drop the time argument in the freezing point for the mean \( m_{v_t}^x(x) \) of the frozen process \( \tilde{X}_{v_t} \) introduced in equation (2.5).

Namely, since we always consider \( \tau = t \) we simply write \( m_{v_t}^x(x) := m_{v_t}^{x, \xi} \) for a given spatial freezing point \( \xi \in \mathbb{R}^{nd} \).

Recall here that the transition time \( t_0 = t + c_0 |(x - x')_i|^{2/(3-1)} \) which is again set after differentiation w.r.t. From Lemma 8 (cancellation argument), we write:

\[
D_{x_k} \tilde{p}_{t_0,t}^{x'} u(t_0, x') - D_{x_k} \tilde{p}_{t_0,t}^{x'} u(t_0, x')
\]

\[
= \int_{\mathbb{R}^{nd}} dy D_{x_k} \tilde{p}_{t_0,t}^{x'} (t, t_0, x', y) [u(t_0, y) - u(t_0, y_1:k-1, (\theta_{t_0,t}(x'))_{k:n})]
\]

\[
- \int_{\mathbb{R}^{nd}} dy D_{x_k} \tilde{p}_{t_0,t}^{x'} (t, t_0, x', y) [u(t_0, y) - u(t_0, y_1:k-1, (m_{t_0,t}^{x'})(k, (\theta_{t_0,t}(x'))_{k+1:n})].
\]

We now split the above contribution into three terms:

\[
D_{x_k} \tilde{p}_{t_0,t}^{x'} u(t_0, x') - D_{x_k} \tilde{p}_{t_0,t}^{x'} u(t_0, x') = \left( B_1^{x_1, \xi} + B_2^{x_1, \xi} + B_3^{x_1, \xi} \right)(t_0, x'), \quad (3.39)
\]

where

\[
B_1^{x_1, \xi} (t_0, x') := \int_{\mathbb{R}^{nd}} dy D_{x_k} \tilde{p}_{t_0,t}^{x'} (t, t_0, x', y) [u(t_0, y) - u(t_0, y_1:k, (\theta_{t_0,t}(x'))_{k+1:n})]
\]

\[
- \int_{\mathbb{R}^{nd}} dy D_{x_k} \tilde{p}_{t_0,t}^{x'} (t, t_0, x', y) [u(t_0, y) - u(t_0, y_1:k, (\theta_{t_0,t}(x'))_{k+1:n})]. \quad (3.40)
\]

\[
B_2^{x_1, \xi} (t_0, x') := \left[ \int_{\mathbb{R}^{nd}} dy D_{x_k} \tilde{p}_{t_0,t}^{x'} (t, t_0, x', y)
\]

\[
[u(t_0, y_1:k, (\theta_{t_0,t}(x'))_{k+1:n}) - u(t_0, y_1:k-1, (\theta_{t_0,t}(x'))_{k:n})].
\]
From Proposition 5 (see also Lemma 7), we thus derive:

\[
- \langle D_k u(t_0, y_{1:k-1}, (\theta_{t_0,t}(x'))_{k:n}), (y - \theta_{t_0,t}(x'))_k \rangle] \\
- \left[ \int_{\mathbb{R}^d} dy D_{x_k} \bar{p}^{\bar{e}}(t, t_0, x', y) \right.
\]

\[
[u(t_0, y_{1:k-1}, (\theta_{t_0,t}(x'))_{k+1:n}) - u(t_0, y_{1:k-1}, (\mathbf{m}_{t_0,t}(x'))_k, \theta_{t_0,t}(x')_{k+1:n}) \\
- \langle D_k u(t_0, y_{1:k-1}, (\mathbf{m}_{t_0,t}(x'))_k, \theta_{t_0,t}(x')_{k+1:n}), (y - \mathbf{m}_{t_0,t}(x'))_k \rangle] \right] \\
\}

\begin{equation}
(3.41)
\end{equation}

\[
\mathcal{B}_{2}^{k',\bar{e}'}(t_0, x') := \left\{ \int_{\mathbb{R}^d} dy D_{x_k} \bar{p}^{\bar{e}}(t, t_0, x', y) [D_k u(t_0, y_{1:k-1}, (\theta_{t_0,t}(x'))_{k:n}), (y - \theta_{t_0,t}(x'))_k] \\
- \int_{\mathbb{R}^d} dy D_{x_k} \bar{p}^{\bar{e}}(t, t_0, x', y) [D_k u(t_0, y_{1:k-1}, (\mathbf{m}_{t_0,t}(x'))_k, \theta_{t_0,t}(x')_{k+1:n}), (y - \mathbf{m}_{t_0,t}(x'))_k] \right\}
\end{equation}

\begin{equation}
(3.42)
\end{equation}

We now exploit the Hölder regularity of \( D_k u \) w.r.t. the \( k \)th variable to control the terms in \( \mathcal{B}_{2}^{k',\bar{e}'}(t_0, x') \) defined in (3.41). Let us first write from the previous decomposition:

\[
| \mathcal{B}_{2,1}^{k',\bar{e}'}(t_0, x') | \\
:= \left| \int_{\mathbb{R}^d} dy D_{x_k} \bar{p}^{\bar{e}}(t, t_0, x', y) \left[u(t_0, y_{1:k-1}, (\theta_{t_0,t}(x'))_{k+1:n}) - u(t_0, y_{1:k-1}, (\theta_{t_0,t}(x'))_{k:n}) \right.ight. \\
- \langle D_k u(t_0, y_{1:k-1}, (\theta_{t_0,t}(x'))_{k+1:n}), (y - \theta_{t_0,t}(x'))_k \rangle \left. \right] \right| \\
\leq \int_{\mathbb{R}^d} dy D_{x_k} \bar{p}^{\bar{e}}(t, t_0, x', y) \left[ |D_k u(t_0, y_{1:k-1}, \cdot, (\theta_{t_0,t}(x'))_{k+1:n})| |y - \theta_{t_0,t}(x')|_k \right]^{1+\alpha_k}.
\]

From Proposition 5 (see also Lemma 7), we thus derive:

\[
| \mathcal{B}_{2,1}^{k',\bar{e}'}(t_0, x') |_{(\xi',\bar{e}')=(x',x)} \\
\leq C(t_0 - t)^{\alpha_k(k - \frac{1}{2})} \sup_{z_j, j \in [1:n], j \neq k} |D_k u(t_0, z_{1:k}, \cdot, z_{k+1:n})|_{\alpha_k} \\
\int_{\mathbb{R}^{(n-1)d}} dy_{1:k-1} dy_{k+1:n} \hat{q}_{C^{-1}}(t, s, x, (y_{1:k-1}, y_{k+1:n})) \\
= C(t_0 - t)^{\frac{1}{2}} \left[ |D_k u|^k(t_0, \cdot) \right]_{\alpha_k},
\]

(3.43)

recalling from Lemma 10 that \( \alpha_k = 1/(2k - 1) \). The same arguments readily give:

\[
| \mathcal{B}_{2,2}^{k',\bar{e}'}(t_0, x') | \\
:= \left| \int_{\mathbb{R}^d} dy D_{x_k} \bar{p}^{\bar{e}}(t, t_0, x', y) \left[u(t_0, y_{1:k-1}, (\theta_{t_0,t}(x'))_{k+1:n}) - u(t_0, y_{1:k-1}, (\mathbf{m}_{t_0,t}(x'))_k, \theta_{t_0,t}(x')_{k+1:n}) \right. \right. \\
- \langle D_k u(t_0, y_{1:k-1}, (\mathbf{m}_{t_0,t}(x'))_k, \theta_{t_0,t}(x')_{k+1:n}), (y - \mathbf{m}_{t_0,t}(x'))_k \rangle \left. \right] \right| \\
= C(t_0 - t)^{\frac{1}{2}} \left[ |D_k u|^k(t_0, \cdot) \right]_{\alpha_k},
\]
\[-\langle D_ku(t_0, y_{1:k-1}, (m^x_{l_0,t}(x'))_k, \theta_{l_0,t}(x')_{k+1:n}), (y - m^x_{l_0,t}(x'))_k \rangle \right| \\
\leq C(0 - t)^{\frac{1}{2}} |(D_ku)_k(t_0, \cdot)|_{\alpha k}. \quad (3.44)\]

Observe also that for this term we do not need to exploit the specific choice \((\xi', \tilde{\xi}') = (x', x)\) for the freezing point.

Let us now deal with the contribution \(B_1^{\xi', \tilde{\xi}'}(t_0, x')\) in (3.40). Observe from this definition that this term is non-zero if and only if \(k < n\). Write then

\[
\left| B_1^{\xi', \tilde{\xi}'}(t_0, x') \right|_{(\xi', \tilde{\xi}')=(x', x)} \\
\leq \int_{\mathbb{R}^d} dy |D_x \tilde{p}^{\xi'}(t, t_0, x', y)|_{\xi'=x'} \left| u(t_0, y) - u(t_0, y_{1:k}, (\theta_{l_0,t}(x'))_{k+1:n}) \right| \\
+ \int_{\mathbb{R}^d} dy |D_x \tilde{p}^{\xi'}(t, t_0, x', y)|_{\xi'=x'} \left| u(t_0, y) - u(t_0, y_{1:k}, (\theta_{l_0,t}(x'))_{k+1:n}) \right| \\
\leq C ||Du||_{\infty} \left( \int_{\mathbb{R}^d} \frac{dy}{(t_0 - t)^{k-\frac{1}{2}}} |\tilde{p}^{x'-1}(t, t_0, x', y)| (\theta_{l_0,t}(x') - y)_{k+1:n} \right) \\
+ \int_{\mathbb{R}^d} \frac{dy}{(t_0 - t)^{k-\frac{1}{2}}} \tilde{p}^{x'-1}(t, t_0, x', y) \\
\left( |(m^x_{l_0,t}(x') - y)_{k+1:n}| + |(m^x_{l_0,t}(x') - \theta_{l_0,t}(x'))_{k+1:n}| \right). \quad (3.45)\]

To deal with the last contribution in the r.h.s., we will need some auxiliary lemmas already used in [12] for Schauder estimates. Namely, analogously to Lemmas 1 and 3 therein, we have the following result.

**Lemma 16** There exists \( \vartheta = \vartheta((A)) \in (0, 1) \) s.t. for all \( j \in \llbracket 1, n \rrbracket, x \in \mathbb{R}^n, x' = (x_{i-1}, z, x_{i+n}) \in \mathbb{R}^{n+1} :\)

\[
\left| (m^x_{l_0,t}(x') - \theta_{l_0,t}(x'))_j \right| \leq c_0^\vartheta d_2^{j-1}(x, x') := c_0^\vartheta |(x - x')_i|^{\frac{2j-1}{2^{j-1}}}. \quad (3.46)\]

For the sake of completeness a proof is provided in Appendix A below. Recalling that \( t_0 = t + c_0 |(x - x')_i|^{2/(2i-1)} \) with \(|(x - x')_i| \leq 1\), we obtain

\[
\frac{||Du||_{\infty}}{(t_0 - t)^{k-\frac{1}{2}}} |(m^x_{l_0,t}(x') - \theta_{l_0,t}(x'))_{k+1:n}| \\
\leq \frac{||Du||_{\infty}}{c_0^{\vartheta}} |(x - x')_i|^{\frac{2k+1}{2^{k-1}}} c_0^\vartheta |(x - x')_i|^{\frac{2k+1}{2^{k-1}}} \\
\leq ||Du||_{\infty} c_0^\vartheta |(x - x')_i|^{\frac{2k+1}{2^{k-1}}} \leq ||Du||_{\infty} c_0^\vartheta |(x - x')_i|^{\alpha i}.\]
Plugging the above control in (3.45) and using as well Lemma 7 and Proposition 5, we obtain:

\[
|B^k_1(t_0, x')|_{(\xi', \xi')=(x', x)} \leq C \|D u\|_\infty \left\{ 2(t_0 - t) + c_0^{\theta - \frac{1}{2}} |(x - x')_i|^{\alpha_i} \right\}.
\]

(3.47)

Let us eventually control the term \(B^2_3(t_0, x')\) defined in (3.42) which we rewrite in the following way:

\[
\begin{align*}
B^2_3(t_0, x') & = \left\{ \int_{\mathbb{R}^d} dy D_{x_k} \tilde{p}(t, t_0, x', y) \right. \\
& \quad \left\{ [D_k u(t_0, y_{1:k-1}, (\theta_{t_0,t}(x'))_{k,n}) - D_k u(t_0, (\theta_{t_0,t}(x'))_k)] \right. \\
& \quad - \left. \int_{\mathbb{R}^d} dy D_{x_k} \tilde{p}(t, t_0, x', y) \left[ D_k u(t_0, y_{1:k-1}, (m^{\tilde{\xi}'}_{t_0,t}(x'))_{k+1:n}) \right. \\
& \quad - \left. D_k u(t_0, (m^{\tilde{\xi}'}_{t_0,t}(x'))_{k+1:n}) \right] (y - m^{\tilde{\xi}'}_{t_0,t}(x'))_k \right\} \\
& \quad + \left\{ \int_{\mathbb{R}^d} dy D_{x_k} \tilde{p}(t, t_0, x', y) \left[ D_k u(t_0, (\theta_{t_0,t}(x'))_{k+1:n}) \right. \\
& \quad - \left. D_k u(t_0, (\theta_{t_0,t}(x'))_{k+1:n}) \right] (y - \tilde{\theta}_{t_0,t}(x'))_k \right\}
\end{align*}
\]

where, thanks to Proposition 6 the last contribution is actually 0. For the first and second contributions in the above r.h.s. we have, thanks to the Hölder regularity of \(x_{1:k} \mapsto D_k u(\cdot, x_{1:k}, \cdot)\), Proposition 5 and Lemma 16:

\[
|B^2_3(t_0, x')|_{(\xi', \xi')=(x', x)} \\
\leq C \int_{\mathbb{R}^d} dy \left\{ \tilde{p}^{x'}(t, t_0, x', y) + \tilde{p}^{x}(t, t_0, x', y) \right. \\
\left. + \sum_{j=1}^{k-1} [(D_k u)_j(t_0, \cdot)]_{\alpha_j} (t_0 - t)^{\alpha_j \left( j - \frac{1}{2} \right)} \right\} \\
+ C \int_{\mathbb{R}^d} dy \tilde{p}^{x'}(t, t_0, x', y) \sum_{j=1}^{k} [(D_k u)_j(t_0, \cdot)]_{\alpha_j} (t_0 - t)^{\frac{1}{2}} \\
\leq C \left( \sum_{j=1}^{k} [(D_k u)_j(t_0, \cdot)]_{\alpha_j} \right) (t_0 - t)^{\frac{1}{2}}.
\]

(3.48)
Thus, plugging estimates (3.43), (3.44), (3.47) and (3.48) in (3.39) we deduce that

\[
|D_{x_k} \tilde{P}_{t_0,i}^{x'} u(t_0, x') - D_{x_k} \tilde{P}_{t_0,i}^{\tilde{x}'} u(t_0, \tilde{x}')|_{(x', \tilde{x}')=(x', x)} \\
\leq C \left[ \left\| Du \right\|_{\infty} \left\{ (t_0 - t) + c_0 \left( \frac{\delta}{k} - \left( k - \frac{1}{2} \right) \right) \right\} \left\| (x - x')_i \right\|_{\alpha_i} \right] \\
+ \left( \sum_{j=1}^{k} \left( (D_{k} u)_i (t_0, \cdot) \right)_{\alpha_j} \right) (t_0 - t)^{\frac{1}{2}}. 
\tag{3.49}
\]

From the definition of \( t_0 = t + \ell \min (|x - x'|)^{2/(2i - 1)} \) and \( \alpha_i = 1/(2i - 1) \), recalling that \( t_0 - t \) is small as well (i.e. \( t_0 - t \leq C (t_0 - t)^{1/2} \)), we obtain from (3.49):

\[
|D_{x_k} \tilde{P}_{t_0,i}^{x'} u(t_0, x') - D_{x_k} \tilde{P}_{t_0,i}^{\tilde{x}'} u(t_0, \tilde{x}')|_{(x', \tilde{x}')=(x', x)} \\
\leq C \left\{ \frac{1}{c_0} \left( \sum_{j=1}^{k} \left( (D_{k} u)_i (t_0, \cdot) \right)_{\alpha_j} \right) + \left( \frac{\delta}{c_0} - \left( k - \frac{1}{2} \right) \right) \frac{1}{c_0} \right\} \left\| Du \right\|_{\infty} \left\| (x - x')_i \right\|_{\alpha_i}. 
\]

Thus, from (3.21), there exists \( \tilde{\delta} := \tilde{\delta}((A)) \in (0, 1) \) such that

\[
|D_{x_k} u(t, x')|_{\partial S_{t}} \leq \left\| (x - x')_i \right\|_{\alpha_i} C \left( c_0^{-\tilde{\delta} - 1} \left\| Du \right\|_{\infty} + c_0^{\tilde{\delta}} \max_{2 \leq \ell \leq m \leq n} \sup_{s \in [0,T]} \left\| (D_{m} u)_s (\cdot, \cdot) \right\|_{\alpha_{\ell}} \right). 
\tag{3.50}
\]

**Conclusion: control of (3.18).**

Plugging (3.26), (3.38) and (3.50) into (3.18), (3.16), we eventually derive that for some positive \( \tilde{\delta} := \tilde{\delta}((A)) > 0 \):

\[
\left\| D_{x_k} u(t, x_{1:i-1}, \ldots, x_{i+1:n}) \right\|_{B^{\alpha_i}_{\infty, \infty}} \leq C \left( c_0^{-\tilde{\delta} - 1} \left\| Du \right\|_{\infty} + \left\| D_{i} u \right\|_{\infty} + \left( \tilde{\delta} + T^{\tilde{\delta}} \right) \right) \left( \sup_{s \in [0,T], 2 \leq \ell \leq m \leq n} \left\| (D_{m} u)_s (\cdot, \cdot) \right\|_{B^{\alpha_{\ell}}_{\infty, \infty}} \right). 
\]

The main point to close our circular argument then consists in taking the supremums w.r.t. \( x_{1:i-1}, x_{i+1:n}, i, k \) and \( t \in [0, T] \) on the l.h.s. and to tune the constant \( c_0 \) and the terminal time \( T \) in order to obtain \( C \left( c_0^{\tilde{\delta}} + T^{\tilde{\delta}} \right) \leq 1/2 \). We then derive that for all \( 2 \leq i \leq k \leq n \):

\[
\sup_{t \in [0,T]} \left\| (D_{k} u)_i (t, \cdot) \right\|_{B^{\alpha_i}_{\infty, \infty}} \leq C \left( \left\| Du \right\|_{\infty} + \left\| D_{i} u \right\|_{\infty} \right), 
\tag{3.51}
\]

which concludes the proof of Lemma 11. \( \square \)

### 3.3 Proof of of Lemma 15

We follow the proof of Lemma 10, concentrating on the case \( \ell \leq k - 1 \), the specific case \( \ell = k \) could be treated similarly considering the slightly different cancellation.
terms already discussed in Lemma 10. The quantity to estimate is now:

\[ D_\ell \left[ \Psi_{\ell, k, m}^{(s, y_{1: \ell - 1}, y_{\ell + 1: n})}(t, x, x) - \psi_{\ell, k, m}^{(s, y_{1: \ell - 1}, y_{\ell + 1: n})}(t, x', x) \right] \]  

(3.52)

Splitting the thermic part of the Besov norm as in (3.4), we obtain the same kind estimate for the non-singular in time part. Indeed, we point out that the difference (3.52) does not involve \( D_{x_1} \), therefore we gain a singularity of order 1/2 w.r.t. Inequality (3.5).

On the other hand, the difference of the derivatives of the frozen densities w.r.t. \( x_k \) can be handled with (3.27). Choosing \( \rho_\ell = 2\ell - 1 \) as in the proof of Lemma 10, and recalling that \( \alpha_i(i - 1/2) = 1/2 \) (see Lemma 10), it is plain to check that:

\[
\begin{align*}
\text{Upper} & \left[ D_\ell \left[ \Psi_{\ell, k, m}^{(s, y_{1: \ell - 1}, y_{\ell + 1: n})}(t, x, x) - \psi_{\ell, k, m}^{(s, y_{1: \ell - 1}, y_{\ell + 1: n})}(t, x', x) \right] \right] \\
& \leq \frac{\hat{C}((x - x')_i) t |\alpha_i|}{(s - t)^{1 - \gamma_\ell}} \hat{q}_\ell(x)(t, s, x, (y_{1: \ell - 1}, y_{\ell + 1: n})), \\
& \text{(3.53)}
\end{align*}
\]

where from the definition of \( \gamma_\ell \) in Lemma 10, \( 3/2 - \gamma_\ell < 1 \).

Turning to the singular in time contribution of the thermic part of the Besov norm of (3.52) we decompose with the notations of Lemma 10 (see e.g. (3.9), (3.10) which exhibit an additional spatial derivative):

\[
\begin{align*}
\text{Lower} & \left[ D_\ell \left[ \Psi_{\ell, k, m}^{(s, y_{1: \ell - 1}, y_{\ell + 1: n})}(t, x, x) - \psi_{\ell, k, m}^{(s, y_{1: \ell - 1}, y_{\ell + 1: n})}(t, x', x) \right] \right] \\
& = \int_0^{(s-t)^{\rho_\ell}} \frac{d w}{w^\frac{\rho_\ell}{2} + 1} \int_{\mathbb{R}^d} d\tilde{z} \left[ \mathcal{F}_{1, \ell, k, m}^{(s, y_{1: \ell - 1}, y_{\ell + 1: n})}(t, x, x') \right] (w, \tilde{z}), \\
& \text{(3.54)}
\end{align*}
\]

where \( \ell < k \leq m \) and:

\[
\begin{align*}
\mathcal{F}_{1, \ell, k, m}^{(s, y_{1: \ell - 1}, y_{\ell + 1: n})}(t, x, x')(w, \tilde{z})
& = \int_{\mathbb{R}^d} dy_\ell D_\ell \partial_w h_w(\tilde{z} - y_\ell) \left( D_{x_k} \tilde{p}_\xi(t, s, x, y) - D_{x_k} \tilde{p}_\xi(t, s, x', y) \right)_{\xi = x} \\
& \left\{ F_\ell(s, y_{1:k-1}, \theta_{s,t}^{k,m}(x)) - F_\ell(s, y_{1: \ell - 1}, \tilde{z}, y_{\ell + 1: k - 1}, \theta_{s,t}^{k,m}(x)) \right\}, (y - \theta_{s,t}(x))_m,
& \text{(3.54)}
\end{align*}
\]

with a slight abuse of notation when \( \ell = k - 1 \) and

\[
\begin{align*}
\mathcal{F}_{2, \ell, k, m}^{(s, y_{1: \ell - 1}, y_{\ell + 1: n})}(t, x, x')(w, \tilde{z})
& = \int_{\mathbb{R}^d} dy_\ell D_\ell \partial_w h_w(\tilde{z} - y_\ell) \left[ \left( D_{x_k} \tilde{p}_\xi(t, s, x, y) - D_{x_k} \tilde{p}_\xi(t, s, x, y_{1: \ell - 1}, \tilde{z}, y_{\ell + 1: n}) \right) \\
& \left( D_{x_k} \tilde{p}_\xi(t, s, x', y) - D_{x_k} \tilde{p}_\xi(t, s, x', y_{1: \ell - 1}, \tilde{z}, y_{\ell + 1: n}) \right) \right]_{\xi = x}
& \text{(3.54)}
\end{align*}
\]
\[
\begin{aligned}
&\left\{ F_\ell (s, y^{1:1}_\ell, z, y^{1:k}_\ell) - F_\ell (s, \theta_{s,t}) \right\} \\
&- D_{\ell-1} F_\ell (s, \theta_{s,t})(y - \theta_{s,t})_{\ell-1} (y - \theta_{s,t})_m. \\
\end{aligned}
\] (3.55)

Note now that, applying a Taylor formula around \( \tilde{z} \) for the above term into brackets, we then get from Lemma 14, equation (3.27) that:

\[
\left| \left( D_{x_i} \tilde{p}_\ell^k(t, s, x, y) - D_{x_i} \tilde{p}_\ell^k(t, s, x, y^{1:1}_\ell, \tilde{z}, y^{1:k}_\ell) \right) \right|
\]

\[
\leq \frac{\tilde{C}}{(s-t)^{\frac{1}{2}} + \frac{1}{2} (k-\frac{1}{2})} \int_0^1 d\lambda \left\| \hat{p}^k_{\ell-1}(t, s, x, y^{1:1}_\ell, \tilde{z} + \lambda (y^{1:k}_\ell - \tilde{z}), y^{1:k}_\ell) \right\| \left( \| x' - x \|^{1/2} | \tilde{z} - y^{1:k}_\ell | \right).
\]

With this control at hand, together with estimate (3.27), to handle the contributions

\[
\mathcal{T}_{1,\ell,k,m}(s,y^{1:1}_\ell,y^{1:k}_\ell,t,x',x) + \mathcal{T}_{2,\ell,k,m}(s,y^{1:1}_\ell,y^{1:k}_\ell,t,x',x) \]

we can mimic the proof of the estimation of the contributions

\[
\mathcal{T}_{1,\ell,i,l,k}(s,y^{1:1}_\ell,y^{1:k}_\ell,t,x',x) + \mathcal{T}_{2,\ell,i,l,k}(s,y^{1:1}_\ell,y^{1:k}_\ell,t,x',x)
\]

done in Lemma 10 to obtain

\[
\text{Lower} \left[ D_{\ell} \left[ \psi^{(s,y^{1:1}_\ell,y^{1:k}_\ell),t,x,x}_{\ell,k,m} \right] - \psi^{(s,y^{1:1}_\ell,y^{1:k}_\ell),t,x,x}_{\ell,k,m} \right] \]

\[
\leq \tilde{C} \hat{\gamma}_{\ell} (s-t)^{\rho_{\ell}} \left( \int_0^{(s-t)^{\rho_{\ell}}} \frac{d w}{w} \right)^{\frac{\gamma_{\ell}}{2} + 1} \\
\times \left\{ \frac{1}{w^{3/2} - (s-t)^{\rho_{\ell}}} + \frac{w^{-1 + \rho_{\ell}}}{(s-t)^{\ell-2}} + \frac{w^{-1}}{(s-t)^{\ell-2 + \frac{1}{2} + 1}} \right\} \\
\leq \tilde{C} \hat{\gamma}_{\ell} (s-t)^{\rho_{\ell}} (s-t)^{-\frac{3}{2} + \frac{\gamma_{\ell}}{2}} \| x' - x \|^{\alpha_{\ell}},
\]

where \( \gamma_{\ell} = 1/2 + \eta(\ell - 3/2) \), which together with (3.53) concludes the proof. \( \square \)

\textbf{Acknowledgements} We would like to thank the anonymous referees for their careful reading and comments which improved a lot the presentation of the manuscript. For the first author, this work has been partially supported by the ANR project ANR-15-IDEX-02. For the third author, the article was prepared within the framework of a subsidy granted to the HSE by the Government of the Russian Federation for the implementation of the Global Competitiveness Program.

\section*{A Sensitivity results for the mean: Proof of Lemma 16}

In order to prove Lemma 16, we first need to establish some controls on the sensitivity of the flows, see Lemma 17 below. Those results are obtained under the sole assumption
(A) and remain valid for the mollification procedure of the coefficients considered in (AM). We will then proceed to the final proof of Lemma 16 in Section A.2.

For our analysis, we now introduce the spatial homogeneous distance, which basically reflects the various scales of the system already seen e.g. in Proposition 5. Namely, for \((x, x') \in \mathbb{R}^{nd}\), we define:

\[
d(x, x') := \sum_{i=1}^{n} |(x - x')_i|^{\frac{1}{2}}. \quad (A.1)
\]

The distance is homogeneous in the sense that, for any \(\lambda > 0\),

\[
d(\lambda^{-1/2}T\lambda x, \lambda^{-1/2}T\lambda x') = \lambda^{1/2}d(x, x').
\]

**A.1 A first sensitivity result for the flow**

**Lemma 17** (Control of the flows). Under (A), there exists \(C := C((A), T)\) s.t. for all spatial points \((x, x') \in (\mathbb{R}^{nd})^2\), \(d(x, x') \leq 1\), \(0 \leq t < s \leq T \leq 1\) and \(i \in [1, n]\):

\[
|\theta_{s,t}(x) - \theta_{s,t}(x')|_i \leq C \left( (s - t)^{i - \frac{1}{2}} + d^{2i-1}(x - x') \right).
\]

The flow, \(\theta_{s,t}\), is, somehow, locally “almost” Lipschitz continuous in space w.r.t. the homogeneous distance \(d\), up to a time additive term. This time contribution is a consequence of the non-Lipschitz continuity of the drift \(F\). The analysis of the sensitivity of the flow was already done for \(F\) Lipschitz continuous in Proposition 4.1 of [48], and Appendix A.1 in [12] with different Hölder regularity of \(F\). Actually, as we consider a smoother drift than in [12], the following lemma can be seen as a by-product of Lemma 12 therein.

For the sake of completeness, we provide the corresponding, and more direct, analysis below.

**Proof** The analysis mainly relies on Grönwall type arguments coupled with suitable mollification techniques, because \(F\) is not Lipschitz continuous, and appropriate Young inequalities in order to make the intrinsic scales associated to the spatial variables appear.

Let \(\delta \in \mathbb{R}^n\) be the vector whose entries \(\delta_i > 0\) correspond to the mollification parameter of the drift \(F_i\) in the \(i^{th}\) variable for \(i \in [2, n]\). Namely, for all \(v \in [0, T], z \in \mathbb{R}^{nd}, i \in [2, n]\), we define

\[
F^\delta_i(v, z^{i-1:n}) := F_i(v, \cdot) * \rho_{\delta_i}(z) = \int_{\mathbb{R}^d} dw F_i(v, z_{i-1}, z_i - w, z_{i+1}, \ldots, z_n) \rho_{\delta_i}(w), \quad (A.2)
\]

with \(\rho_{\delta_i}(w) := (1/\delta_i^d) \rho (w/\delta_i)\) where \(\rho : \mathbb{R}^d \rightarrow \mathbb{R}_+\) is a usual mollifier, namely \(\rho\) has compact support and \(\int_{\mathbb{R}^d} \rho(z) dz = 1\). Eventually, we write \(F^\delta(v, z) := (F^\delta_1(v, z), F^\delta_2(v, z), \ldots, F^\delta_n(v, z))\). With a slight abuse of notation in the previous definitions, since the first component \(F^\delta_1\) is not mollified. The sublinearity of \(F^\delta_1\) is actually enough to obtain the desired control.
To be at the good current time scale for the contributions associated with the mollification, we pick \( \delta_i \) in order to have \( C := C((A), T) > 0 \) s.t. for all \( z \in \mathbb{R}^{nd}, u \in [t, s] \):

\[
(s - t)^{\frac{1}{2}} \mathbb{T}_{s-t}^{-1} \left( F(u, z) - F^\delta(u, z) \right) \leq C(s - t)^{-1}.
\]

(A.3)

By the previous definition of \( F^\delta \) in (A.2), identity (A.3) will be implied from (T)\( \beta \) if:

\[
\sum_{i=2}^{n} (s - t)^{\frac{1}{2} - i \delta_i^{\frac{2i-2}{2i-1}}} \leq C(s - t)^{-1}.
\]

(A.4)

Hence, we choose from now on, for each \( i \in \{2, n\} \):

\[
\delta_i = (s - t)^{(i-\frac{1}{2}) \delta_i^{\frac{2i-2}{2i-1}}}.
\]

(A.5)

Next, let us control the last components of the flow. By the definition of \( \theta_{s,t} \) in (2.4), we get:

\[
|\theta_{s,t}(x) - \theta_{s,t}(x')|_n
\leq |(x - x')|_n + \int_t^s dv \left( |F^\delta_n(v, \theta_{v,t}(x)) - F^\delta_n(v, \theta_{v,t}(x'))| + |F^\delta_n(v, \theta_{v,t}(x)) - F_n(v, \theta_{v,t}(x'))| \right)
\leq |(x - x')|_n + C \int_t^s dv \left( |(\theta_{v,t}(x) - \theta_{v,t}(x'))|_{n-1} \right)
+ \delta_n^{-1+\frac{2n-2}{2n-1}} \left( |(x - x')|_n + (s - t)^{\frac{2n-2}{2n-1} - \delta_n} \right) + (s - t)^{\frac{2n-2}{2n-1}},
\]

observing for the last inequality that since \( \beta_n > (2n - 2)/(2n - 1) \) and \( \delta_n \) is meant to be small, \( \delta_n^{\beta_n} \leq \delta_n^{(2n-2)/(2n-1)} \).

Hence by Grönwall’s lemma, we get:

\[
|\theta_{s,t}(x) - \theta_{s,t}(x')|_n
\leq C \exp \left( C(s - t) \delta_n^{-\frac{2n-2}{2n-1} + \frac{2n-2}{2n-1}} \right)
\times \left( |(x - x')|_n + (s - t)^{\frac{2n-2}{2n-1}} + \int_t^s dv \left( |(\theta_{v,t}(x) - \theta_{v,t}(x'))|_{n-1} \right) \right)
\leq C \exp \left( C(s - t)^{\frac{1}{2}} \right)
\left( |(x - x')|_n + (s - t)^{\frac{n-1}{2}} + \int_t^s dv \left( |(\theta_{v,t}(x) - \theta_{v,t}(x'))|_{n-1} \right) \right),
\]

(A.6)

using (A.5) for the last inequality. For the \((n - 1)^{th}\) component, the situation is quite different in the sense that we have to handle the non-Lipschitz continuity of \( F^\delta_{n-1} \) in
its \(n^{th}\) variable. Write:

\[
|\langle \theta_{s,t}(x) - \theta_{s,t}(x') \rangle_{n-1}| \\
\leq C \exp \left( C(s-t)^{3/2} \right) \left( (x-x')_{n-1} + (s-t)^{-\frac{3}{2}} \right) \\
+ \int_t^s dv \left| (\theta_{v,t}(x) - \theta_{v,t}(x'))_{n-2} \right| \\
+ |(x-x')_n| \beta_n + v - t \beta_n(n^{-\frac{1}{2}}) + \left( \int_t^v \left. dw \right| (\theta_{w,t}(x) - \theta_{w,t}(x'))_{n-1} \beta_n \right) \\
\leq C \exp \left( C(s-t)^{3/2} \right) \left( (x-x')_{n-1} + (s-t)^{-\frac{3}{2}} \right) \\
+ \int_t^s dv \left| (\theta_{v,t}(x) - \theta_{v,t}(x'))_{n-2} \right| \\
+ |(x-x')_n| \frac{2n-3}{2n-1} + \left( \int_t^v \left. dw \right| (\theta_{w,t}(x) - \theta_{w,t}(x'))_{n-1} \beta_n \right)
\]

(A.7)

from our choice of \(\delta_{n-1}\) in (A.5) for the second inequality. We also exploited for the last inequality that, since under \((T_\beta, \beta_n > (2n-2)/(2n-1), \beta_n (n - \frac{1}{2}) > n - 1 > n - 3/2\) and \(0 \leq t < s \leq T\) where \(T\) is small, then \((v-t)^{\beta_n(n-1/2)} \leq (v-t)^{n-3/2}\). Also, since \(d(x, x') \leq 1\), the same arguments yield \(|(x-x')_n| \beta_n \leq |(x-x')_n| \frac{2n-2}{2n-1} \leq |(x-x')_n| \frac{2n-3}{2n-1}\).

From (A.7), which still holds true replacing \(s\) by any \(\tilde{s} \in [t, s]\), we deduce that taking the supremum over \(\tilde{s} \in [t, s]\):

\[
\sup_{\tilde{s} \in [t, s]} |\langle \theta_{\tilde{s},t}(x) - \theta_{\tilde{s},t}(x') \rangle_{n-1}| \\
\leq C \exp \left( C(s-t)^{3/2} \right) \left( (x-x')_{n-1} + (s-t)^{-\frac{3}{2}} \right) \\
+ \int_t^s dv \left| (\theta_{v,t}(x) - \theta_{v,t}(x'))_{n-2} \right| \\
+ |(x-x')_n| \frac{2n-3}{2n-1} + \left( \int_t^v \left. dw \right| (\theta_{w,t}(x) - \theta_{w,t}(x'))_{n-1} \beta_n \right)
\]

Taking then the supremum in \(w \in [t, s]\) in the above integral, we obtain:

\[
\sup_{\tilde{s} \in [t, s]} |\langle \theta_{\tilde{s},t}(x) - \theta_{\tilde{s},t}(x') \rangle_{n-1}| \\
\leq C \exp \left( C(s-t)^{3/2} \right) \left( (x-x')_{n-1} + (s-t)^{-\frac{3}{2}} \right) \\
+ \int_t^s dv \left| (\theta_{v,t}(x) - \theta_{v,t}(x'))_{n-2} \right| \\
+ |(x-x')_n| \frac{2n-3}{2n-1} + \sup_{w \in [t, s]} \left| (\theta_{w,t}(x) - \theta_{w,t}(x'))_{n-1} \beta_n (s-t)^{\beta_n+1} \right|
\]

(A.8)
From Young’s inequality we now derive:

\[
\sup_{w \in [t,s]} \left| (\theta_{w,t}(x) - \theta_{w,t}(x')) \right|_{n-1}^\beta_n (s-t)^{\beta_n+1} \\
\leq \frac{1}{2C} \exp(C) \sup_{w \in [t,s]} \left| (\theta_{w,t}(x) - \theta_{w,t}(x')) \right|_{n-1} (s-t) + \bar{C}(s-t)^{1-\beta_n} \\
\leq \frac{1}{2C} \exp(C) \sup_{w \in [t,s]} \left| (\theta_{w,t}(x) - \theta_{w,t}(x')) \right|_{n-1} + \bar{C}(s-t)^{n-3/2},
\]

calling for the last inequality that \( s-t \leq 1 \), and since \( \beta_n > (2n-2)/(2n-1) \), \( 1-\beta_n < 1/(2n-1) \), we also have \( (s-t)^{1/(1-\beta_n)} < (s-t)^{2n-1}/(s-t)^{n-3/2} \). Plugging the above control into (A.8), we obtain up to a modification of \( C \):

\[
\frac{1}{2} \sup_{\tilde{w} \in [t,s]} \left| (\theta_{\tilde{w},t}(x) - \theta_{\tilde{w},t}(x')) \right|_{n-1} \\
\leq C \exp \left( C(s-t)^{1/2} \right) \\
\left( |(x-x')_{n-1}| + (s-t)^{n-3/2} + \int_t^s dv \left| (\theta_{v,t}(x) - \theta_{v,t}(x')) \right|_{n-2} + |(x-x')_{n}|^{2n-3/2n-1} \right).
\]

Plugging the above inequality into (A.6) we derive:

\[
|\theta_{s,t}(x) - \theta_{s,t}(x')|_n \\
\leq C \exp \left( C(s-t)^{1/2} \right) \left( |(x-x')_{n}| + (s-t)^{n-1/2} + |(x-x')_{n-1}|(s-t) \\
+ |(x-x')_{n}|^{2n-3/2n-1} + \int_t^s dv \int_t^v dw \left| (\theta_{w,t}(x) - \theta_{w,t}(x')) \right|_{n-2} \right) \\
\leq C \exp \left( C(s-t)^{1/2} \right) \left( |(x-x')_{n}| + (s-t)^{n-1/2} + |(x-x')_{n-1}|^{2n-3/2n-1} \\
+ \int_t^s dv \int_t^v dw \left| (\theta_{w,t}(x) - \theta_{w,t}(x')) \right|_{n-2} \right),
\]

using again the Young inequalities \(|(x-x')_{n}|^{(2n-3)/(2n-1)}(s-t) \leq C \left( |(x-x')_{n}| + (s-t)^{n-1/2} \right) \) and \(|(x-x')_{n-1}|(s-t) \leq C \left( |(x-x')_{n-1}|^{(2n-1)/(2n-3)} + (s-t)^{n-1/2} \right) \) for the last inequality. Iterating the procedure, we get:

\[
|\theta_{s,t}(x) - \theta_{s,t}(x')|_n \\
\leq C \left( (s-t)^{n-1/2} + \sum_{j=2}^n |(x-x')_{j}|^{2n-3/2n-1} \right) \\
+ \int_t^{v_n=s} dv_{n-1} \ldots \int_t^{v_2} dv_1 \left| (\theta_{v_1,t}(x') - \theta_{v_1,t}(x)) \right|_1.
\]

\( \odot \) Springer
Analogously, for $i \in \{2, n\}$, we obtain:

$$
|\langle \theta_{s,t}(\mathbf{x}) - \theta_{s,t}(\mathbf{x}') \rangle_i|
\leq C \left( (s-t)^{1-\frac{i}{2}} + \sum_{j=2}^{n} |(\mathbf{x} - \mathbf{x}')_j|^{2j-1} + \int_{t}^{s} dv_{i-1} \ldots \int_{t}^{v_{2}} dv_{1} |\langle \theta_{v_{1},t}(\mathbf{x}) - \theta_{v_{1},t}(\mathbf{x}) \rangle_1| \right). 
$$

(A.11)

We explicitly see from the above equation that each entry of the difference of the starting points appears at its intrinsic scale for the homogeneous distance $d$ introduced in (A.1).

**Remark 6** Observe that equations (A.10) and (A.11) are available for any fixed time $s \in [t, T]$. The first term, i.e. for $i = 1$ is controlled slightly differently. In other words, for any $\tilde{s} \in [t, s]$, write:

$$
|\langle \theta_{\tilde{s},t}(\mathbf{x}) - \theta_{\tilde{s},t}(\mathbf{x}') \rangle_1| \leq |(\mathbf{x} - \mathbf{x}')_1| + C \sum_{j=1}^{n} \int_{t}^{\tilde{s}} dv |\langle \theta_{v,t}(\mathbf{x}) - \theta_{v,t}(\mathbf{x}') \rangle_j|^{\beta_j},
$$

which in turn implies from (A.11), Remark 6 and convexity inequalities:

$$
\sup_{\tilde{s} \in [t, s]} |\langle \theta_{\tilde{s},t}(\mathbf{x}) - \theta_{\tilde{s},t}(\mathbf{x}') \rangle_1|
\leq |(\mathbf{x} - \mathbf{x}')_1| + C \left( (s-t) \sup_{v \in [t, s]} |\langle \theta_{v,t}(\mathbf{x}) - \theta_{v,t}(\mathbf{x}') \rangle_1|^{\beta_1} + \sum_{j=2}^{n} \int_{t}^{s} dv |\langle \theta_{v,t}(\mathbf{x}) - \theta_{v,t}(\mathbf{x}') \rangle_j|^{\beta_j} \right)
\leq |(\mathbf{x} - \mathbf{x}')_1| + \left( (s-t) \left[ \frac{1}{2} \sup_{v \in [t, s]} |\langle \theta_{\tilde{s},t}(\mathbf{x}) - \theta_{\tilde{s},t}(\mathbf{x}') \rangle_1| + C \right]
\right)
\right)\left( (s-t)^{j-\frac{1}{2}} + \sum_{k=2}^{n} |(\mathbf{x} - \mathbf{x}')_k|^{2k-1} + (s-t)^{j-1} \sup_{v \in [t, s]} |\langle \theta_{v,t}(\mathbf{x}) - \theta_{v,t}(\mathbf{x}') \rangle_1|^{\beta_j} \right).
$$

Recalling $\beta_j > (2j - 2)/(2j - 1)$ and $0 \leq t < s \leq T \leq 1$, $d(\mathbf{x}, \mathbf{x}') \leq 1$, we get:

$$
\frac{1}{2} \sup_{\tilde{s} \in [t, s]} |\langle \theta_{\tilde{s},t}(\mathbf{x}) - \theta_{\tilde{s},t}(\mathbf{x}') \rangle_1|
\leq |(\mathbf{x} - \mathbf{x}')_1| + C \left( (s-t) + \sum_{j=2}^{n} (s-t) \right)
\leq |(\mathbf{x} - \mathbf{x}')_1| + C \left( (s-t) + \sum_{j=2}^{n} (s-t) \right).
$$
\[
\left((s - t)^{\frac{1}{2}} + \sum_{k=2}^{n} |(x - x')_k|^{\frac{1}{2k-1}} + (s - t)^{(j-1)\beta_j} \sup_{v \in [t,s]} |(\theta_{v,t}(x) - \theta_{v,t}(x'))_1|^{\beta_j}\right).
\]  
(A.12)

Write now from the Young inequality:
\[
C(s - t)^{1+(j-1)\beta_j} \sup_{v \in [t,s]} |(\theta_{v,t}(x) - \theta_{v,t}(x'))_1|^{\beta_j}
\leq C(s - t) + \frac{1}{4} \sup_{v \in [t,s]} |(\theta_{v,t}(x) - \theta_{v,t}(x'))_1|.
\]

We eventually derive from (A.12) that:
\[
\sup_{v \in [t,s]} |(\theta_{v,t}(x) - \theta_{v,t}(x'))_1| \leq C \left((s - t)^{\frac{1}{2}} + d(x, x')\right),
\]
which gives the statement for \(i = 1\). Plugging now this inequality into (A.11), we get for each \(i \in \{2, n\}\):
\[
|(\theta_{s,t}(x) - \theta_{s,t}(x'))_i| 
\leq C \left((s - t)^{i-\frac{1}{2}} + d^{2i-1}(x, x') + (s - t)^i \sup_{v \in [t,s]} |(\theta_{v,t}(x) - \theta_{v,t}(x'))_1|\right)
\leq C \left((s - t)^{i-\frac{1}{2}} + d^{2i-1}(x, x') + (s - t)^i \left((s - t)^{\frac{1}{2}} + d(x, x')\right)\right)
\leq C \left((s - t)^{i-\frac{1}{2}} + d^{2i-1}(x, x')\right),
\]
using again for the last identity the Young inequality to derive that 
\((s - t)^{i-1}d(x, x') \leq C \left((s - t)^{i-1/2} + d^{2i-1}(x, x')\right)\). The proof is complete.

\[\square\]

**A.2 Sensitivity results for the mean: final proof of Lemma 16**

Again through the analysis, we assume w.l.o.g. that \(d(x, x') \leq 1\). We choose here to prove, for the sake of completeness, a slightly more general result than the one stated in the Lemma. Namely, we do not restrict to the case where \(x, x'\) only differ on their \(i^{th}\) component but consider arbitrary given points \(x, x' \in \mathbb{R}^{nd}\) to emphasize how the specific structure of the drift leads to the control of the linearization error in terms of the homogeneous distance.

The control is done with a distinction of two contributions to handle.
\[
\mathbf{m}^x_{s,t}(x') - \theta_{s,t}(x') = [\mathbf{m}^x_{s,t}(x') - \theta_{s,t}(x)] + [\theta_{s,t}(x) - \theta_{s,t}(x')]. \tag{A.13}
\]

By the proxy definition in (2.3), we deduce that the mean value of \(\hat{X}_v^{m,\xi}, \mathbf{m}^\xi_{u,v}\) is s.t.
\[
\mathbf{m}^x_{s,t}(x') - \theta_{s,t}(x) = x' - x + \int_t^s dv DF(v, \theta_{v,t}(x))[\mathbf{m}^x_{v,t}(x') - \theta_{v,t}(x)]. \tag{A.14}
\]
The sub-triangular structure of $DF$ yields that for each $i \in [2, n]$:

$$(m^x_{s,t}(x') - \theta_{s,t}(x))_i = x'_i - x_i + \int_t^s dv D_{i-1} F_i(v, \theta_{v,t}(x))[m^x_{v,t}(x')_i - \theta_{v,t}(x)_i].$$

Also, since $m^x_{v,t}(x')_1 = x'_1 + \int_t^s dv F_1(v, \theta_{v,t}(x))$, so we obtain that $[m^x_{v,t}(x')_1 - \theta_{v,t}(x)_1] = x'_1 - x_1$, we then obtain by iteration that:

$$(m^x_{s,t}(x') - \theta_{s,t}(x))_i = x'_i - x_i + \sum_{k=2}^i \left[ \int_t^{v_{i-1}} dv_{i-1} \ldots \int_t^{v_k} dv_{k-1} \prod_{j=k}^i D_{j-1} F_j(v_j, \theta_{v,t,\mathcal{I}}(x)) \right] [x'_{k-1} - x_{k-1}].$$

with the convention that for $i = 1$, $\sum_{k=2}^i = 0$. From the above control, equation (A.13) and the dynamics of the flow, and because the starting points are the same, the contributions involving differences of the spatial points $(x' - x)$ or flows only appear in iterated time integrals, we obtain:

$$| (m^x_{s,t}(x') - \theta_{s,t}(x'))_i | \leq \left| \sum_{k=2}^i \left[ \int_t^{v_{i-1}} dv_{i-1} \ldots \int_t^{v_k} dv_{k-1} \prod_{j=k}^i D_{j-1} F_j(v_j, \theta_{v,t,\mathcal{I}}(x)) \right] [x'_{k-1} - x_{k-1}] \right| + \int_t^s | F_i(v, \theta_{v,t}(x)) - F_i(v, \theta_{v,t}(x')) | dv$$

$$\leq C \left( \sum_{k=2}^{i-1} (s - t)^{i-k} |x_k - x'_k| + \int_t^s dv \left( \sum_{j=i}^n |(\theta_{v,t}(x) - \theta_{v,t}(x'))_j|^{\beta_j} + |(\theta_{v,t}(x) - \theta_{v,t}(x'))_{i-1} | \right) \right).$$

We derive from the previous Lemma 17 (control of the flows) recalling again that $\beta_j > (2j - 2)/(2j - 1)$ and $d(x, x') \leq 1$, $0 \leq t < s \leq T \leq 1$: 
\[ |(m_{s,t}(x') - \theta_{s,t}(x'))_i| \leq C \left( \sum_{k=2}^{i-1} (s-t)^{i-k} |x_k - x'_k| + (s-t)^{\frac{2i-2}{2}+1} + d^{2i-2}(x, x')(s-t) + ((s-t)^{i-1} + d^{2i-1}(x, x'))(s-t) \right). \]

In particular, for \( s = t_0 = t + c_0 d^2(x, x') \) with \( c_0 < 1 \), the previous equation yields:

\[ |(m_{t_0,t}^x(x') - \theta_{t_0,t}(x'))_i| \leq C \left( c_0 d^{2i-1}(x, x') + (c_0 + c_i) d^{2i-1}(x, x') + \left( c_0^2 + c_0 \right) d^{2i-1}(x, x') \right). \]

using again \( d(x, x') \leq 1 \) for the middle term. After summing and by convexity inequalities, we eventually deduce:

\[ d(m_{s,t}^x(x'), \theta_{s,t}(x')) \leq C c_0^{2i-1} d(x, x'). \]

This concludes the proof of Lemma 16. \( \square \)

**B Parametrix expansion with different freezing points**

In this section we show how the parametrix expansion (3.17) involving different freezing points can be derived. This can actually be done from the Duhamel formulation up to an additional discontinuity term. Restarting from (2.26) we can indeed rewrite from the Markov property that for given \( (t, x') \in [0, T] \times \mathbb{R}^{nd} \) and any \( r \in (t, T] \), \( \xi' \in \mathbb{R}^{nd} \):

\[ u(t, x') = \left[ \tilde{P}_{r, t}^{\xi'} u(r, \cdot) \right](x') + \int_t^r ds \left[ \tilde{P}_{s, t}^{\xi'} f(s, \cdot) \right](x') \]

\[ + \int_t^r ds \left[ \tilde{P}_{s, t}^{\xi'} (L_s - \tilde{L}_s^{\xi'}) u(s, \cdot) \right](x'). \] \( \text{(B.1)} \)

Differentiating the above expression in \( r \in (t, T] \) yields for any \( \xi' \in \mathbb{R}^{nd} \):

\[ 0 = \partial_r \left[ \tilde{P}_{r, t}^{\xi'} u(r, \cdot) \right](x') + \left[ \tilde{P}_{r, t}^{\xi'} f(r, \cdot) \right](x') + \left[ \tilde{P}_{r, t}^{\xi'} (L_r - \tilde{L}_r^{\xi'}) u(r, \cdot) \right](x'). \] \( \text{(B.2)} \)

Denoting by \( t_0 = t, T \] the time at which we change the freezing point and integrating (B.2) on \([t, t_0]\) for a first given \( \xi' \) and between \([t_0, T]\) with a possibly different \( \tilde{\xi}' \) yields:

\[ 0 = \left[ \tilde{P}_{t_0, t}^{\xi'} u(t_0, \cdot) \right](x') - u(t, x') + \int_t^{t_0} ds \left[ \tilde{P}_{s, t}^{\xi'} f(s, \cdot) \right](x'). \]
\[ + \int_t^{t_0} ds \left[ \tilde{P}_{s,t}^\xi \left( (L_s - \tilde{L}_s^\xi') u \right)(s, \cdot) \right](x') \]

\[ + \tilde{P}_{T,t}^\xi u(T, \cdot)(x') - \left[ \tilde{P}_{t_0,t}^\xi u(t_0, \cdot) \right](x') \]

\[ + \int_0^T ds \left[ \tilde{P}_{s,t}^\xi f(s, \cdot) \right](x') + \int_t^{T_0} ds \left[ \tilde{P}_{s,t}^\xi \left( (L_s - \tilde{L}_s^\xi') u \right)(s, \cdot) \right](x'). \]

Recalling that \( u(T, \cdot) = 0 \) (terminal condition), the above equation rewrites:

\[ u(t, x') = \int_t^T ds \left( \mathbb{1}_{s \leq t_0} \left[ \tilde{P}_{s,t}^\xi f(s, \cdot) \right](x') + \mathbb{1}_{s > t_0} \left[ \tilde{P}_{s,t}^\xi f(s, \cdot) \right](x') \right) \]

\[ + \tilde{P}_{t_0,t}^\xi u(t_0, \cdot)(x') - \left[ \tilde{P}_{t_0,t}^\xi u(t_0, \cdot) \right](x') \]

\[ + \int_t^T ds \left[ \mathbb{1}_{s \leq t_0} \left[ \tilde{P}_{s,t}^\xi \left( (L_s - \tilde{L}_s^\xi') u \right)(s, \cdot) \right](x') + \mathbb{1}_{s > t_0} \left[ \tilde{P}_{s,t}^\xi \left( (L_s - \tilde{L}_s^\xi') u \right)(s, \cdot) \right](x') \right). \]

We see that for \( \xi' \neq \tilde{\xi}' \) we have an additional discontinuity term deriving from the change of freezing point along the time variable. The above equation is precisely (3.17).

## C Auxiliary results concerning the multi-scale Gaussian densities, their derivatives and some related objects

In order to be self contained, we gather in this section the proof of some results related to the Gaussian dynamics in (2.3). Namely, we provide a complete proof of Proposition 4 and some auxiliary related results used throughout the previous proofs. We here freely use the notations of Sect. 2.1.

### C.1 About the objects appearing in the multi-scale density

#### C.1.1 Good scaling properties of the covariance matrix: proof of Proposition 4

We recall the correspondence between the notations of [22] and those of the current article.

- Notations and Assumptions from [22]. Consider the Gaussian process with dynamics

\[ dG_t = L_t G_t dt + B \Sigma_t dW_t \quad \text{(C.1)} \]
where \((\Sigma_t)_{t \in [0,T]}\) is a measurable deterministic \(\mathbb{R}^d \otimes \mathbb{R}^d\) valued family s.t. \(A_t := \Sigma_t \Sigma_t^*\) has uniformly non-degenerate spectrum, i.e. there exists \(\Lambda \geq 1\) s.t. for any \(t \in [0, T]\), \(\text{Spec}(A_t) \in [\Lambda^{-1}, \Lambda]\), and the measurable deterministic \(\mathbb{R}^{nd} \otimes \mathbb{R}^{nd}\) valued family \((L_t)_{t \in [0,T]}\) is such that for any \(t \in [0, T]\):

\[
L_t = \begin{pmatrix}
0_{d,d} & \cdots & \cdots & 0_{d,d} \\
\alpha_t^1 & 0_{d,d} & \cdots & 0_{d,d} \\
0_{d,d} & \alpha_t^2 & \cdots & 0_{d,d} \\
\vdots & \vdots & \ddots & \vdots \\
0_{d,d} & \cdots & 0_{d,d} & \alpha_t^{n-1} 0_{d,d}
\end{pmatrix},
\]

(C.2)

where the \((\alpha_t^i)_{i \in [1,n-1]}\) are \(\mathbb{R}^d \otimes \mathbb{R}^d\) valued.

This is special case of assumption \((A)_{\text{linear}}\) in [22]. Proposition 3.4 of that reference states that, whenever for all \(i \in [1, n-1]\) and \(t \in [0, T]\), \(\alpha_t^i\) belongs to \(E_i\) (closed convex subset of \(GL_d(\mathbb{R})\)), there exists a constant \(c \geq 1\) depending on \(E_i\), \(\Lambda\), \(\kappa\) s.t.

\[
\max_{i \in [1,n-1]} \sup_{t \in [0,T]} |\alpha_t^i| \leq \kappa, \quad n, \quad d
\]

such that the Gaussian process \((G_t)_{t \in [0,T]}\) introduced in (C.1) satisfies a good scaling property with constant \(c\) in the sense of Definition 3.2 of [22]. Precisely, denoting by \((R(s, t))_{0 \leq s, t \leq T}\) the resolvent matrix associated with \((L_t)_{t \geq 0}\), the covariance matrix

\[
K_t = \int_0^t ds R_{t,s} B A_s B^* R_{t,s}^*
\]

of the random variable \(G_t\) satisfies that for any \(y \in \mathbb{R}^{nd}\),

\[
c^{-1} t^{-1} |\mathbb{T}_t y|^2 \leq \langle K_t y, y \rangle \leq ct^{-1} |\mathbb{T}_t y|^2,
\]

or equivalently, for all \(t \in (0, T]\):

\[
c^{-1} t |\mathbb{T}_t^{-1} y|^2 \leq \langle K_t^{-1} y, y \rangle \leq ct |\mathbb{T}_t^{-1} y|^2.
\]

Derivation of the Proposition 4 from the previous results of [22]. From the dynamics of the process \((\check{X}_v^{(r, \xi)})_{v \in [t,T]}\) given in (2.3) and starting from \(x\) at time \(t\), see also the associated integrated expression in (2.5), it can be seen that the covariance matrix of the random variable \(\check{X}_v^{(r, \xi)}\) writes, with the notations of Sect. 2.1:

\[
\check{K}_{v,t}^{(r, \xi)} := \int_t^v du \check{R}^{(r, \xi)}(v, u) B a(u, \theta_{u,t}(\xi)) B^* \check{R}^{(r, \xi)}(v, u)^t,
\]

as given in the statement of Proposition 4. This covariance matrix also corresponds to the one of a Gaussian process with dynamics (C.1) setting for fixed \(0 \leq t <
\( s \leq T \) and for any \( r \in [0, s-t] \)

\[
\mathbf{L}_r = \begin{pmatrix}
D_x \mathbf{F}_2(t + r, \theta_{t+r, \tau}(\xi)) & \cdots & \cdots & \cdots & 0_{d,d} \\
0_{d,d} & \cdots & \cdots & \cdots & 0_{d,d} \\
0_{d,d} & D_x \mathbf{F}_3(t + r, \theta_{t+r, \tau}^{n=1/r}(\xi)) & 0_{d,d} & \cdots & 0_{d,d} \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0_{d,d} & \cdots & 0_{d,d} & D_x \mathbf{F}_n(t + r, \theta_{t+r, \tau}(\xi)) & 0_{d,d}
\end{pmatrix}.
\]

and

\[
\Sigma_r = \sigma(t + r, \theta_{t+r, \tau}(\xi)).
\]

Since the resolvent \( (\mathbf{R}(r, v))_{v \in [0, s-t]} \) associated with \( (\mathbf{L}_r)_{r \in [0, s-t]} \) writes \( \mathbf{R}(r, v) = \hat{\mathbf{R}}(t, \xi)(t + r, t + v) \), one readily derives that the covariance matrix \( \mathbf{K}_{s-t} \) of \( \mathbf{G}_{t-s} = \mathbf{R}(t - s, 0) \mathbf{x} + \int_0^{t-s} \mathbf{R}(t - s, r) \mathbf{B} \Sigma_r dW_r \) coincides with \( \hat{\mathbf{K}}_{s,t}^{(t, \xi)} \). Since \( \mathbf{K}_{s-t} = \hat{\mathbf{K}}_{s,t}^{(t, \xi)} \) satisfies a good scaling property, this proves Proposition 4.

C.1.2 Scaling Properties of the resolvent

We here aim at proving the following control. There exists \( \hat{C}_1 := \hat{C}_1((\mathbf{A}), T) \) s.t.

\[
\forall \xi \in \mathbb{R}^{nd}, \left| \hat{\mathbf{R}}_{(t, \xi)}^{(t, \xi), s,t} \right| \leq \hat{C}_1|\xi|.
\]

We will proceed following the arguments of Lemma 3.6 in [22] specifying how they can apply to the current setting. Let us restart from the previous linear Gaussian dynamics. Namely, for fixed \( t \in [0, T] \), let \( (\mathbf{G}_s)_{s \in [0, t]} \) be as in (C.1) and introduce the scaled process \( \hat{\mathbf{G}}_s^t = t^{1/2} \mathbb{T}_{-1}^{t} \mathbf{G}_s \), \( s \in [0, 1] \). Intuitively, from the previously established good scaling property of \( \mathbf{G} \), all the components of the process \( \hat{\mathbf{G}}_s^t \) actually evolve at a macro scale, i.e. its covariance matrix is of order one at time \( s = 1 \).

It is then easily checked that \( \mathbf{G}_s^t \) satisfies (C.1), i.e. \( d\hat{\mathbf{G}}_s^t = \hat{\mathbf{L}}_s^t d\hat{\mathbf{G}}_s + B \hat{\Sigma}_s^t d\hat{\dot{W}}_s^t \) with:

\[
\hat{\mathbf{L}}_s^t = t \mathbb{T}_{s}^{-1} \mathbf{L}_{s,t} \mathbb{T}_t, \quad \hat{\Sigma}_s^t = \mathbf{S}_{s,t}, \quad \hat{\dot{W}}_s^t = t^{-1/2} W_{s,t}, \quad s \in [0, 1],
\]

and importantly that from the specific subdiagonal structure considered in (C.3) \( \hat{\mathbf{L}}_s^t = \mathbf{L}_{s,t} \). In any case, we obtain \( |\hat{\mathbf{L}}_s^t| \leq (1 + T^n) \kappa \) as soon as \( \sup_{s \in [0,1]} |\mathbf{L}_{s,t}| \leq \kappa \).

It is then clear that denoting by \( \hat{\mathbf{R}}^t \) the resolvent associated with \( (\hat{\mathbf{L}}_s^t)_{s \in [0,1]} \) it holds that there exists \( \hat{C}_1 := \hat{C}_1((\mathbf{A}), T) \) s.t. for all \( s_0, s_1 \in [0, 1], |\hat{\mathbf{R}}^t(s_1, s_0)| \leq \hat{C}_1 \). On the other hand, direct computations also yield that

\[
\hat{\mathbf{R}}^t(s_1, s_0) = \mathbb{T}_{s}^{-1} \mathbf{R}(s_1 t, s_0 t) \mathbb{T}_t \iff \mathbf{R}(s_1 t, s_0 t) = \mathbb{T}_t \hat{\mathbf{R}}^t(s_1, s_0) \mathbb{T}_{s}^{-1} \implies [\mathbf{R}(t, 0)]^* = \mathbb{T}_t^{-1} [\hat{\mathbf{R}}^t(1, 0)]^* \mathbb{T}_t.
\]
where $R$ stands for the resolvent associated with $L$. The final bound (C.4) on the rescaled resolvent associated with the frozen process $((s - t)^{1/2} \tilde{X}^\varepsilon_{t}(t,x))_{t \in [s,t]}$ is eventually derived from the same previous correspondence exhibited to prove the good scaling property of Proposition 4 in the previous paragraph.

C.2 Proof of Lemma 14

This Section is dedicated to the proof of the Lemma 14 about the Hölder controls for the derivatives of the frozen densities in the diagonal-diagonal regime, i.e. when $x, x' \in \mathbb{R}^{nd}$ only differ in their $i^{th}$, $d$-dimensional component and $s, t \in S^c_i$, i.e. $c_0(|x - x|) = (2^{1/2})^{d-1} \leq s - t \leq T - s$.

Write:

$$D_{x_k}D_{y_1}^\beta \tilde{p}^\xi(t, s, x, y) - D_{x_k}D_{y_1}^\beta \tilde{p}^\xi(t, s, x', y)$$

$$= - \int_0^1 d\lambda D_{x_k} \left(D_{x_k}D_{y_1}^\beta \tilde{p}^\xi(t, s, x + \lambda(x' - x), y)\right) \cdot (x' - x)_i. \quad (C.5)$$

From (2.15) in Proposition 5 we thus derive:

$$|D_{x_k}D_{y_1}^\beta \tilde{p}^\xi(t, s, x, y) - D_{x_k}D_{y_1}^\beta \tilde{p}^\xi(t, s, x', y)|$$

$$\leq \frac{C(|x' - x|)}{(s - t)^{i - \frac{1}{2} + \frac{k - 1}{2} + \frac{d}{2}}} \int_0^1 d\lambda \tilde{p}^\xi_{c-1}(t, s, x + \lambda(x' - x), y). \quad (C.6)$$

Now, from the definition of $\tilde{p}_{c-1}$ in (2.15), recalling as well from (2.7) that $x \mapsto \tilde{m}^\xi_{s,t}(x) := m^\xi_{s,t}(x)$ is affine, we get:

$$|\tilde{p}^\xi_{c-1}(t, s, x + \lambda(x' - x), y)|$$

$$\leq \frac{C}{(s - t)^{n^{2d}/2}} \exp\left(-c(s - t)|\mathbb{T}^{-1}_{s-t}(\tilde{m}^\xi_{s,t}(x + \lambda(x' - x)) - y)|^2\right)$$

$$\leq \frac{C}{(s - t)^{n^{2d}/2}} \exp\left(c(s - t)|\mathbb{T}^{-1}_{s-t}(\tilde{R}^{(t,\xi)}(s, t)(x - x'))|^2\right)$$

$$\exp\left(-\frac{c}{2}(s - t)|\mathbb{T}^{-1}_{s-t}(\tilde{m}^\xi_{s,t}(x) - y)|^2\right).$$

Using the rescaling arguments of the proof of Proposition 5 on the resolvent (see equation (2.18)), we then get $(s - t)^{1/2}|\mathbb{T}^{-1}_{s-t}(\tilde{R}^{(t,\xi)}(s, t)(x - x'))| \leq C(s - t)^{1/2}|\mathbb{T}^{-1}_{s-t}(x - x')| = C(s - t)^{-1/2}|(x' - x)| \leq C$, from the very definition of $S^c_i$. Hence,

$$|\tilde{p}^\xi_{c-1}(t, s, x + \lambda(x' - x), y)| \leq \frac{C}{(s - t)^{n^{2d}/2}} \exp\left(-\frac{c}{2}(s - t)|\mathbb{T}^{-1}_{s-t}(\tilde{m}^\xi_{s,t}(x) - y)|^2\right). \quad (C.7)$$
Recalling that for given \( \mathbf{x}, \mathbf{x}' \) in the global diagonal regime and \( s \in S^c_i, |\mathbf{x}_i - \mathbf{x}'_i| \leq c_0^{-\frac{i-\frac{1}{2}}{2}} (s-t)^{-\frac{1}{2}} \), we derive

\[
\frac{|\mathbf{x}_i - \mathbf{x}'_i|}{(s-t)^{i-\frac{1}{2}}} \leq \left( \frac{|\mathbf{x}_i - \mathbf{x}'_i|}{(s-t)^{i-\frac{1}{2}}} \right)^{\alpha_i} \left( \frac{|\mathbf{x}_i - \mathbf{x}'_i|}{(s-t)^{i-\frac{1}{2}}} \right)^{1-\alpha_i} \leq \frac{|\mathbf{x}_i - \mathbf{x}'_i|^{\alpha_i}}{(s-t)^{i-\frac{1}{2}}} c_0^{-\frac{i-\frac{1}{2}}{2} (1-\alpha_i)} ,
\]

which from (C.6) and (C.7) precisely gives (3.27) recalling that \( \alpha_i = 1/(2i-1) \) (see the definition in the statement of Lemma 10).

### D Proof of Lemma 12 through reverse Taylor formula

**Proof of Lemma 12** We assume here, for the sake of simplicity and without loss of generality, that \( d = 1 \) (scalar case). When \( d > 1 \), the proof below can be reproduced componentwise. Fix \( l \in [2,n] \) and \( k \in [l+1,n] \). For \( y \in \mathbb{R}^{nd} \), \( \mathbf{y}' = (y_1, \cdots, y_{k-1}, y'_k, y_{k+1}, \cdots, y_n) \), write:

\[
D_l u(s, \mathbf{y}') - D_l u(s, \mathbf{y}) = \int_0^1 d \mu \left( D_l u(s, \mathbf{y}') - D_l u(s, \mathbf{y}'_1, \cdots, \mathbf{y}'_l, \mathbf{y}_{l+1:n}) \right)
\]

\[
+ \{ D_l u(s, \mathbf{y}_1, \cdots, \mathbf{y}_l, \mathbf{y}_{l+1:n}) - D_l u(s, \mathbf{y}) \}
\]

\[
+ \{ D_l u(s, \mathbf{y}_1, \cdots, \mathbf{y}_l, \mathbf{y}_{l+1:n}) - D_l u(s, \mathbf{y}_1, \cdots, \mathbf{y}_l, \mathbf{y}_{l+1:n}) \}
\]

\[
= \sum_{\ell=1}^3 \Delta^l \ell(s, \mathbf{y}', \mathbf{y}) ,
\]

(D.8)

where \( d = |\mathbf{y}' - \mathbf{y}| = |y'_k - y_k| \). The first two terms can be dealt directly.

\[
|\Delta^l 1(s, \mathbf{y}', \mathbf{y})| + |\Delta^l 2(s, \mathbf{y}', \mathbf{y})| \leq 2[(D_l u)_l(s, \cdot)]_{\alpha l} d^{\delta_l} \alpha l .
\]

(D.9)

For \( \Delta^l 3(s, \mathbf{y}', \mathbf{y}) \), we use an explicit reverse Taylor expansion which yields:

\[
|\Delta^l 3(s, \mathbf{y}', \mathbf{y})| = d^{-\delta_l} \left| u(s, \mathbf{y}_1, \cdots, \mathbf{y}_l, \mathbf{y}_{l+1:n}) - u(s, \mathbf{y}')
\]

\[
\sum_{\ell=l}^d \sum_{\alpha l} \sum_{\delta l} (D_l u)_l(s, \cdot)|_{\alpha l} d^{\delta l} \alpha l .
\]

Since we now recall
from Lemma 11 that \( \alpha_l = 1/(2l - 1) \), we get

\[
\delta_l \alpha_l = \frac{1}{2^{l-1} + 1} \cdot \frac{1}{2l - 1} = \frac{1}{2l} =: \zeta_l.
\]

We then write from (D.8) and the definition of \( \delta_l \) that:

\[
|D_l u(s, y') - D_l u(s, y)| \leq 2(\|Du\|_{\infty} + [(D_l u)_l(s, \cdot)]_{\alpha_l})|y' - y_k|^{\zeta_l},
\]

which gives the result.

\[\square\]

**References**

1. Athreya, S., Butkovsky, O., Mytnik, L.: Strong existence and uniqueness for stable stochastic differential equations with distributional drift. Ann. Probab. **48**(1), 178–210 (2020)
2. Adams, D.R., Hedberg, L.I.: Function Spaces and Potential Theory. Springer, Cham (1996)
3. Bass, R.F.: Diffusions and Elliptic Operators. Springer, Cham (1997)
4. Bass, R., Chen, Z.Q.: Brownian motion with singular drift. Ann. Probab. **31–2**, 791–817 (2003)
5. Bramanti, M., Cupini, G., Lanconelli, E., Priola, E.: Global \( L^p \) estimates for degenerate Ornstein-Uhlenbeck operators. Math. Z. **266**(4), 789–816 (2010)
6. Bramanti, M., Cupini, G., Lanconelli, E., Priola, E.: Global \( L^p \) estimates for degenerate Ornstein-Uhlenbeck operators with variable coefficients. Math. Nachr. **286**(11–12), 1087–1101 (2013)
7. Beck, L., Flandoli, F., Gubinelli, M., Maurelli, M.: Stochastic ODEs and stochastic linear PDEs with critical drift: regularity, duality and uniqueness. Electr. J. Prob. **24**, 136 (2019)
8. Bramanti, M., Zhu, M.: \( L^p \) and Schauder estimates for nonvariational operators structured on Hörmander vector fields with drift. Anal. PDE **6**(8), 1793–1855 (2013)
9. Cannizzaro, G., Chouk, K.: Multidimensional sdes with singular drift and universal construction of the polymer measure with white noise potential. Ann. Probab. **46**(3), 1710–1763 (2018)
10. 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), 259–286 (2017)
11. Chaudru de Raynal, P.-E.: Weak regularization by stochastic drift: result and counter example. Discrete Cont. Dyn. Syst. (Ser. A) **38–3**, 1269–1291 (2018)
12. Chaudru de Raynal, P.-E., Honoré, I., Menozzi, S.: Sharp Schauder estimates for some degenerate Kolmogorov equations. Ann. Scuola Normale Superiore di Pisa, Scienze **22–3**, 989–1089 (2021)
13. Chaudru de Raynal, P.-E., Menozzi, S.: On multidimensional stable-driven stochastic differential equations with Besov drift. arXiv:1907.12263, (2019)
14. Chaudru de Raynal, P.-E., Menozzi, S.: Regularization effects of a noise propagating through a chain of differential equations: an almost sharp result. Trans. Am. Math. Soc. **375–1**, 1–45 (2022). https://doi.org/10.1090/tran/7947
15. Chaudru de Raynal, P.-E., Menozzi, S., Priola, E.: Schauder estimates for drifted fractional operators in the supercritical case. J. Funct. Anal. **278**(8), 108425 (2020)
16. Chaudru de Raynal, P.-E., Menozzi, S., Priola, E.: Weak well-posedness of multidimensional stable driven SDEs in the critical case. Stoch. Dyn. **9**(0), 2040004 (2020)
17. Catellier, R., Gubinelli, M.: Averaging along irregular curves and regularisation of ODEs. Stoch. Proc. Appl. **126–8**, 2323–2366 (2016)
18. Chen, Z.Q., Zhang, X., Zhao, G.: Well-posedness of supercritical SDE driven by Lévy processes with irregular drifts. Trans. Am. Math. Soc. **374–11**, 7621–7655 (2021)
19. Davie, A.M.: Uniqueness of solutions of stochastic differential equations. Int. Math. Res. Not. IMRN **2007**(24), rnm124, 26 pp (2007)
20. Delarue, F., Diehl, R.: Rough paths and 1d SDE with a time dependent distributional drift: application to polymers. Probab. Theory Relat. Fields, pp. 1–63, (2015)
21. Delarue, F., Flandoli, F.: The transition point in the zero noise limit for a 1d Peano example. Discrete Cont. Dyn. Syst. **34**(10), 4071–4083 (2014)
22. Delarue, F., Menozzi, S.: Density estimates for a random noise propagating through a chain of differential equations. J. Funct. Anal. 259–6, 1577–1630 (2010)
23. Di Perna, R., Lions, P.L.: Ordinary differential equations, transport theory and Sobolev spaces. Inv. Math. 98, 511–547 (1989)
24. Fedrizzi, E., Flandoli, F.: Pathwise uniqueness and continuous dependence for SDEs with nonregular drift. Stochastics 83–3, 241–257 (2011)
25. Fedrizzi, E., Flandoli, F., Priola, E., Vovelle, J.: Regularity of stochastic kinetic equations. Electr. J. Prob. 22, 48 (2017)
26. Flandoli, F., Gubinelli, M., Priola, E.: Well-posedness of the transport equation by stochastic perturbation. Inventiones Mathematicae 180(1), 1–53 (2010)
27. Flandoli, F., Issoglio, E., Russo, F.: Multidimensional stochastic differential equations with distributional drift. Trans. Am. Math. Soc. 369(10.1090/tran/6729), 1665–1688 (2017)
28. Flandoli, F.: Random Perturbation of PDEs and Fluid Dynamic Models. Lecture Notes in Mathematics, Springer, Heidelberg (2010)
29. Flandoli, F.: Regularizing properties of brownian paths and a result of Davie. Stoch. Dyn. 11(02n03), 323–331 (2011)
30. Friedman, A.: Partial Differential Equations of Parabolic Type. Prentice-Hall, London (1964)
31. Gradinaru, M., Offret, Y.: Existence and asymptotic behaviour of some time-inhomogeneous diffusions. Annales Institut. H. Poincaré, Probabilités et Statistiques, 49(1), 182–207 (2013)
32. Gikhman, I.I., Skorokhod, A.V.: Introduction to the theory of random processes. Saunders Math Books (1969)
33. Huang, L., Menozzi, S.: A parametrix approach for some degenerate stable driven SDEs. Annales Institut. H. Poincaré 52(4), 1925–1975 (2016)
34. Huang, L., Menozzi, S., Priola, E.: Lp estimates for degenerate non-local kolmogorov operators. J. de Mathématiques Pures et Appliquées 121, 162–215 (2019)
35. Hörmander, L.: Hypoelliptic second order differential equations. Acta Mathematica 119, 147–171 (1967)
36. Harang, F.A., Perkowski, N.: $C^\infty$ regularization of ODEs perturbed by noise, (2020)
37. Hao, Z., Wu, M., Zhang, X.: Schauder estimates for nonlocal kinetic equations and applications. J. de Mathématiques Pures et Appliquées 140, 139–184 (2020)
38. Kolmogorov, A.N.: Zufällige Bewegungen (zur Theorie der Brownschen Bewegung). Ann. Math. 2–35, 116–117 (1934)
39. Krylov, N.V., Priola, E.: Elliptic and parabolic second-order PDEs with growing coefficients. Comm. Partial Diff. Eq. 35(1), 1–22 (2010)
40. Krylov, N., Röckner, M.: Strong solutions of stochastic equations with singular time dependent drift. Prob. Theory Rel. Fields 131, 154–196 (2005)
41. Krylov, N.V.: Lectures on elliptic and parabolic equations in Hölder spaces. Graduate Studies in Mathematics 12. AMS, (1996)
42. Krylov, N.V.: The heat equation in $L^q((0, t), L^p)$-spaces with weights. SIAM J. Math. Anal. 32(5), 1117–1141 (2001)
43. Lorenzi, L.: Estimates of the derivatives for a class of parabolic degenerate operators with unbounded coefficients in $\mathbb{R}^N$. Annali della Scuola Normale Superiore di Pisa. Classe di Scienze. Serie V, 2, 01 (2005)
44. Lemarié-Rieusset, P.-G.: Recent Developments in the Navier-Stokes Problem. CRC Press, London (2002)
45. Ling, C., Zhao, G.: Nonlocal elliptic equation in hölder space and the martingale problem. J. Diff. Eq. 314, 653–699 (2022)
46. Marino, L.: Schauder estimates for degenerate stable Kolmogorov equations. Bull. des Sci. Mathématiques 162, 102885 (2020)
47. Menozzi, S.: Parametrix techniques and martingale problems for some degenerate Kolmogorov equations. Electr. Commun. Prob. 17, 234–250 (2011)
48. Menozzi, S.: Martingale problems for some degenerate Kolmogorov equations. Stoc. Proc. Appl. 128–3, 756–802 (2018)
49. Mikulevicius, R., Pragarauskas, H.: On the Cauchy problem for integro-differential operators in Hölder classes and the uniqueness of the martingale problem. Potential Anal. 40(4), 539–563 (2014)
50. Menozzi, S., Pesce, A., Zhang, X.: Density and gradient estimates for non degenerate Brownian SDEs with unbounded measurable drift. J. Diff. Eq. 272, 330–369 (2021)
51. McKean, H.P., Singer, I.M.: Curvature and the eigenvalues of the Laplacian. J. Diff. Geometry 1, 43–69 (1967)
52. Nualart, D., Ouknine, Y.: Regularization of differential equations by fractional noise. Stoch. Process. Appl. 102(1), 103–116 (2002)
53. Pigato, P.: Density estimates and short-time asymptotics for a hypoelliptic diffusion process. Stoch. Process. Appl. 145, 117–142 (2022)
54. Priola, E.: Pathwise uniqueness for singular SDEs driven by stable processes. Osaka J. Math. 49(2), 421–447 (2012)
55. Priola, E.: On weak uniqueness for some degenerate SDEs by global $L^p$ estimates. Potential Anal. 42(1), 247–281 (2015)
56. Priola, E.: Davie’s type uniqueness for a class of sdes with jumps. Ann. Inst. H. Poincaré Probab. Statist. 54(2), 694–725 (2018)
57. Priola, E.: On Davie’s uniqueness for some degenerate SDEs, (2019)
58. Shaposhnikov, A.V.: Some remarks on Davie’s uniqueness theorem. Proc. Edinburgh Math. Soc. 59(4), 1019–1035 (2016)
59. Stroock, D.W.: Partial differential equations for probabilists. Cambridge Studies in Advanced Mathematics, vol. 112. Cambridge University Press, Cambridge (2008)
60. Stroock, D.W., Varadhan, S.R.S.: Multidimensional Diffusion Processes. Springer-Verlag, Berlin Heidelberg New York (1979)
61. Shaposhnikov, A., Wresch, L.: Pathwise vs. path-by-path uniqueness, (2020)
62. Triebel, H.: Theory of Function Spaces, II. Birkhauser, Basel (1983)
63. Talay, D., Tubaro, L.: Expansion of the global error for numerical schemes solving stochastic differential equations. Stoch. Anal. Appl. 8-4, 94–120 (1990)
64. Veretennikov, A.Y.: Strong solutions and explicit formulas for solutions of stochastic integral equations. Matematicheski Sbornik. Novaya Seriya, 111(153)(3):434–452, 480 (1980)
65. Veretennikov, A.Y.: Stochastic equations with diffusion that degenerates with respect to part of the variables. Izvestiya Akademii Nauk SSSR. Seriya Matematicheskaya 47(1), 189–196 (1983)
66. Wang, F.Y., Zhang, X.: Degenerate SDE with Holder-Dini drift and non-Lipschitz noise coefficient. SIAM J. Math. Anal. 48(3), 2189–2226 (2016)
67. Zhang, X.: Well-posedness and large deviation for degenerate SDEs with Sobolev coefficients. Revista Matematica Iberoamericana, 29(1), (2010)
68. Zhang, X.: Stochastic Hamiltonian flows with singular coefficients. Sci. China Math. 61(8), 1353–1384 (2018)
69. Zvonkin, A.: A transformation of the phase space of a diffusion process that will remove the drift. Mat. Sb. (N.S.), 93(135):129–149, 152, (1974)
70. Zhang, X., Zhao, G.: Heat kernel and ergodicity of SDEs with distributional drifts. arXiv:1710.10537, (2017)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.