Strong regularization by Brownian noise propagating through a weak Hörmander structure

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

Strong regularization by Brownian noise propagating through a weak Hörmander structure

Paul-Éric Chaudru de Raynal[*], **Igor Honoré**[†] **and Stéphane Menozzi**[‡]

September 25, 2020

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.

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:

$$\begin{aligned} &d\mathbf{X}_t^1 = \mathbf{F}_1(t,\mathbf{X}_t^1,\dots,\mathbf{X}_t^n)dt + \sigma(t,\mathbf{X}_t^1,\dots,\mathbf{X}_t^n)dW_t,\\ &d\mathbf{X}_t^2 = \mathbf{F}_2(t,\mathbf{X}_t^1,\dots,\mathbf{X}_t^n)dt,\\ &d\mathbf{X}_t^3 = \mathbf{F}_3(t,\mathbf{X}_t^2,\dots,\mathbf{X}_t^n)dt, \qquad\qquad t\geq 0,\\ &\vdots\\ &d\mathbf{X}_t^n = \mathbf{F}_n(t,\mathbf{X}_t^{n-1},\mathbf{X}_t^n)dt, \end{aligned} \tag{1.1}$$

where $(W_t)_{t\geq 0}$ stands for a d-dimensional Brownian motion on some filtered probability space $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\geq 0},\mathbb{P})$ and for all $i\in[\![1,n]\!]$[1], $t\geq 0$ the component $\mathbf{X}_t^i$ is $\mathbb{R}^d$-valued as well (i.e. $\mathbf{X}_t\in\mathbb{R}^{nd}$). We suppose that the $(\mathbf{F}_i)_{i\in[\![2,n]\!]}$ satisfy a kind of weak Hörmander condition, i.e. the matrices $\left(D_{\mathbf{x}_{i-1}}\mathbf{F}_i(t,\cdot)\right)_{i\in[\![2,n]\!]}$ have full rank. However, the coefficients $(\mathbf{F}_i)_{i\in[\![2,n]\!]}$ can be rather *rough* in their other entries, namely, Hölder continuous. We assume as well that the diffusion coefficient σ 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.* [DPL89]) 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 [Fla11a] and the references therein for an overview of the topic), has been widely studied since the pioneering works of Zvonkin [Zvo74] and Veretennikov [Ver80] 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

[*]Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LAMA, 73000 Chambéry, France. pe.deraynal@univ-smb.fr

[†]Laboratoire de Modélisation Mathématique d'Evry (LaMME), Université d'Evry Val d'Essonne, 23 Boulevard de France 91037 Evry. igor.honore@univ-evry.fr

[‡]Laboratoire de Modélisation Mathématique d'Evry (LaMME), Université d'Evry Val d'Essonne, 23 Boulevard de France 91037 Evry, France and Laboratory of Stochastic Analysis, HSE, Shabolovka 31, Moscow, Russian Federation. stephane.menozzi@univ-evry.fr

[1]We will use throughout the paper the notation $[\![\cdot,\cdot]\!]$ for integer intervals.

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 [KR05] ($L_q - L_p$ drift), Zhang [Zha10] ($L_q - L_p$ drift and weakly Lipschitz diffusion matrix), Fedrizzi and Flandoli [FF11] ($L_q - L_p$ and Hölder drift). We also mention Flandoli *et al.* [FGP10] and Beck *et al.* [BFGM19] 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 [Ver83], where the author extended the result in [Ver80] to some specific case of the considered chain (1.1) for $n = 2$. In comparison with our setting, the author does not impose any non-degeneracy condition on $D_{\mathbf{x}_1}\mathbf{F}_2(t,\cdot)$. The to price pay is anyhow that all coefficients (i.e. with the notations of (1.1) $\mathbf{F}_1, \mathbf{F}_2, \sigma$) 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 $\big(D_{\mathbf{x}_{i-1}}\mathbf{F}_i(t,\cdot)\big)$, $i \in [\![2,n]\!]$ precisely allows the noise to propagate through the chain passing from the i^{th} to the $(i+1)^{\text{th}}$ 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 [Hör67]), 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 [CdR17], Wang and Zhang [WZ16], Fedrizzi *et al.* [FFPV17], Zhang [Zha18]. 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 $\mathbf{F}_i$ is β_j-Hölder continuous in the j^{th} variable for some $\beta_j \in \Big((2j-2)/(2j-1), 1\Big]$ when $i \le j$, so that we recover the critical index mentioned above when $j = 2$. We refer to Section 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. $\mathbf{x}$ or $\mathbf{y}$, any element of $\mathbb{R}^{nd}$, writing as well $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ where for $i \in [\![1,n]\!]$, $\mathbf{x}_i \in \mathbb{R}^d$. For practical purpose we will be led in our analysis to consider *subcomponents* of a vector $\mathbf{x} \in \mathbb{R}^{nd}$. Namely, for any $1 \le i \le j \le n$ and $\mathbf{x} \in \mathbb{R}^{nd}$, we introduce the notation $\mathbf{x}_{i:j} := (\mathbf{x}_i, \dots, \mathbf{x}_j)$. Accordingly, we write the drift as the mapping

$$(s,\mathbf{x}) \in \mathbb{R}^+ \times \mathbb{R}^{nd} \mapsto \mathbf{F}(s,\mathbf{x}) \quad = \quad \begin{pmatrix} \mathbf{F}_1(s,\mathbf{x}) \\ \vdots \\ \mathbf{F}_n(s,\mathbf{x}) \end{pmatrix} := \begin{pmatrix} \mathbf{F}_1(s,\mathbf{x}) \\ \mathbf{F}_2(s,\mathbf{x}) \\ \mathbf{F}_3(s,\mathbf{x}_{2:n}) \\ \vdots \\ \mathbf{F}_n(s,\mathbf{x}_{n-1:n}) \end{pmatrix},$$

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

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

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

For a smooth function $\mathbf{\Psi} : [0,T] \times \mathbb{R}^{nd} \to \mathbb{R}^{nd}$, where $T > 0$ is a fixed given time, writing for $(t, \mathbf{x}) \in [0,T] \times \mathbb{R}^{nd}$, $\mathbf{\Psi}(t,\mathbf{x}) = \big(\mathbf{\Psi}_1(t,\mathbf{x}), \dots, \mathbf{\Psi}_n(t,\mathbf{x})\big)$ where for each $i \in [\![1,n]\!]$, $\mathbf{\Psi}_i$ is $\mathbb{R}^d$ valued, we denote by

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

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

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

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

$$|\sigma(t,\mathbf{x}) - \sigma(t,\mathbf{x}')| \leq \kappa|\mathbf{x}-\mathbf{x}'|.$$

(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 $t \geq 0$, $(\mathbf{x},\zeta) \in \mathbb{R}^{nd} \times \mathbb{R}^d$:

$$\Lambda^{-1}|\zeta|^2 \leq \langle a(t,\mathbf{x})\zeta,\zeta\rangle \leq \Lambda|\zeta|^2.$$

($\mathbf{T}_\beta$) For all $j \in [\![1,n]\!]$, the functions $(\mathbf{F}_i)_{i\in[\![1,j]\!]}$ are uniformly β_j-Hölder continuous in the j^{th} spatial variable with $\beta_j \in \big((2j-2)/(2j-1), 1\big]$, uniformly w.r.t. the other spatial variables of $\mathbf{F}_i$ and in time. In particular, there exists a finite constant $C_\beta > 0$ s.t. with the notations of (1.2),

$$\max_{i\leq j\in[\![1,n]\!]^2} \sup_{s\in[0,T]} [(\mathbf{F}_i)_j(s,\cdot)]_{\beta_j} \leq C_\beta.$$

($\mathbf{H}_\eta$) For all $i \in [\![2,n]\!]$, there exists a closed convex subset $\mathcal{E}_{i-1} \subset GL_d(\mathbb{R})$ (set of invertible $d\times d$ matrices) s.t., for all $t \geq 0$ and $(\mathbf{x}_{i-1},\dots,\mathbf{x}_n) \in \mathbb{R}^{(n-i+2)d}$, $D_{\mathbf{x}_{i-1}}\mathbf{F}_i(t,\mathbf{x}_{i-1},\dots,\mathbf{x}_n) \in \mathcal{E}_{i-1}$. For example, $\mathcal{E}_{i-1}$ may be a closed ball included in $GL_d(\mathbb{R})$ the latter being an open set. Moreover, $D_{\mathbf{x}_{i-1}}\mathbf{F}_i$ is η -Hölder continuous w.r.t. $\mathbf{x}_{i-1}$ uniformly in $\mathbf{x}_{i:n}$ and time. We also assume without loss of generality that $\eta \in \Big(0, \inf_{j\in[\![2,n]\!]}\big\{\beta_j - (2j-2)/(2j-1)\big\}\Big)$, i.e. η is meant to be *small*.

From now on, we will say that assumption $(\mathbf{A})$ is in force provided that **(ML)**, **(UE)**, **($\mathbf{T}_\beta$)** and **($\mathbf{H}_\eta$)** 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 [KR05], 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 subsection 1.5 for further details.*

1.3 Related discussions and perspectives

One may wonder if the thresholds in **($\mathbf{T}_\beta$)** 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 [Fla11a]. Such phenomenon has been considered for ODEs perturbed by many noise classes (in particular α-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 [SW20] 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 $(\mathbf{T}_\beta)$, we introduce, for a given scalar α-stable[2], $\alpha \in (0,2]$ or H-fractional Brownian, $H \in (0,1)$ noise $\mathcal{W}$ of self similarity index $\gamma > 0$, the following simplified version of (1.1)

$$d\mathbf{X}_t = \mathbf{A}\mathbf{X}_t dt + Bd\mathcal{W}_t + \mathbf{P}(\mathbf{X}_t)dt,\ \mathbf{A} = \begin{pmatrix} 0 & \cdots & \cdots & \cdots & 0 \\ 1 & 0 & \cdots & 0 & \vdots \\ 0 & 1 & 0 & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{X}_0 = \mathbf{0}, \tag{1.4}$$

where $d = 1$, $\mathbf{P} = (\mathbf{P}_1, \dots, \mathbf{P}_n)^*$ and for each $i \in [\![1, n]\!]$, $\mathbf{P}_i(\mathbf{x}) = \sum_{j=i}^n P_i^j(\mathbf{x})$ with $P_i^j := c_{i,j}\mathrm{sgn}(\mathbf{x}_j)\big(|\mathbf{x}_j|^{\beta_i^j} \wedge 1\big)$ with $c_{i,j} \in \mathbb{R}$ and $B = (1, 0, \dots, 0)^* = (1, \mathbf{0}_{1,n-1})^*$ where "$*$" stands for the transpose. The dynamics (1.4) can be viewed as a perturbation of $d\bar{\mathbf{X}}_t = \mathbf{A}\bar{\mathbf{X}}_t dt + Bd\mathcal{W}_t$, $\bar{\mathbf{X}}_0 = 0$, corresponding to the noise and its iterated integrals, i.e. $\bar{\mathbf{X}}_t = (\mathcal{W}_t, \int_0^t \mathcal{W}_s ds, \dots, \int_0^t dt_1 \int_0^{t_1} dt_2 \dots \int_0^{t_{n-1}} dt_n \mathcal{W}_{t_n})^*$, by a Peano type drift $\mathbf{P}$. With the notations of (1.1), the full drift of (1.4) writes $\mathbf{F}(t, \mathbf{x}) = \mathbf{F}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{P}(\mathbf{x})$.

Note that the above truncation in $\mathbf{P}$ 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 $(\mathbf{T}_\beta)$ and $(\mathbf{H}_\eta)$. We emphasize that, when $n = 1$, setting $\beta_1^1 =: \beta$, $c_{1,1} = 1$, we obtain the following dynamics

$$d\mathbf{X}_t = \mathrm{sgn}(\mathbf{X}_t)\big(|\mathbf{X}_t|^\beta \wedge 1\big)dt + d\mathcal{W}_t, \quad \mathbf{X}_0 = 0, \tag{1.5}$$

known as the (localized version of the) stochastic Peano example and when $n = 2$, $\mathbf{P}_1 \equiv 0, c_{2,2} = 1$, we get that

$$d\mathbf{X}_t^1 = d\mathcal{W}_t, \quad d\mathbf{X}_t^2 = \mathrm{sgn}(\mathbf{X}_t^2)\big(|\mathbf{X}_t^2|^{\beta_2^2} \wedge 1\big)dt + d\left\{\int_0^t ds\mathcal{W}_s\right\}, \quad \mathbf{X}_0^1 = \mathbf{X}_0^2 = 0, \tag{1.6}$$

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. [SV79]). 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 α-stable process. The *formal* associated PDE writes

$$\partial_t u(t, \mathbf{x}) + \langle \mathrm{sgn}(\mathbf{x})(|\mathbf{x}|^\beta \wedge 1), \mathbf{D}u(t, \mathbf{x})\rangle + \frac{1}{2}\Delta^{\frac{\alpha}{2}} u(t, \mathbf{x}) = 0,$$

where $\Delta^{\alpha/2}$ stands for the usual fractional Laplacian. Introduce for $\lambda > 0$ the corresponding rescaled function $u_\lambda(t, \mathbf{x}) := u(\lambda t, \lambda^{1/\alpha}\mathbf{x})$ (scaling reflecting the parabolic scale for t and $\mathbf{x}$). We get that u_λ satisfies the equation:

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

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

- if $\beta > 1 - \alpha$, the scaled drift coefficient goes to zero with λ 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 λ goes to zero.

[2]For simplicity reasons, we restrain our considerations to rotationally invariant stable processes with generator $1/2$ the usual fractional Laplacian.

In other words, a sufficient condition on the exponent β 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 [DF14] and the counter-examples to weak uniqueness in [CdR18, CdRM17]. From these pathwise analysis, one can formally generalize the above heuristic and get that, for the system (1.5) driven by a general γ self-similar noise $\mathcal{W}$, the condition

$$\beta > 1 - \frac{1}{\gamma}, \tag{1.7}$$

is needed for the weak well-posedness to hold. When $\gamma = 1/\alpha$, α in $(0, 1]$ we refer to e.g. [CdRMP20a, CdRMP20b] for results in that direction and to [DD15, CC18, FIR17, ZZ17, CdRM19, LZ19] for results related to the case where α in $(1, 2]$ so that the above rule allows β to take negative values, the drift being then a distribution.

Thresholds for weak well-posedness and thresholds in $(\mathbf{T}_\beta)$. 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

$$\partial_t u(t, \mathbf{x}_1, \mathbf{x}_2) + \langle \mathbf{x}_1, D_{\mathbf{x}_2} u(t, \mathbf{x}_1, \mathbf{x}_2)\rangle + \langle \mathrm{sgn}(\mathbf{x}_2)(|\mathbf{x}_2|^{\beta_2^2} \wedge 1), D_{\mathbf{x}_2} u(t, \mathbf{x}_1, \mathbf{x}_2)\rangle + \frac{1}{2}\Delta_{\mathbf{x}_1} u(t, \mathbf{x}_1, \mathbf{x}_2) = 0.$$

We introduce the corresponding rescaled function $u_\lambda(t, \mathbf{x}) := u(\lambda t, \lambda^{1/2}\mathbf{x}_1, \lambda^{3/2}\mathbf{x}_2)$ (scaling reflecting the usual parabolic scale for t and $\mathbf{x}_1$ and the fast regime associated with the integral of the Brownian motion) which solves the equation

$$\partial_t u_\lambda(t, \mathbf{x}_1, \mathbf{x}_2) + \langle \mathbf{x}_1, D_{\mathbf{x}_2} u_\lambda(t, \mathbf{x}_1, \mathbf{x}_2)\rangle + \lambda^{-\frac{1}{2}}\langle \mathrm{sgn}(\mathbf{x}_2)(|\lambda^{\frac{3}{2}}\mathbf{x}_2|^{\beta_2^2} \wedge 1), D_{\mathbf{x}_2} u_\lambda(t, \mathbf{x}_1, \mathbf{x}_2)\rangle + \frac{1}{2}\Delta_{\mathbf{x}_1} u_\lambda(t, \mathbf{x}_1, \mathbf{x}_2) = 0,$$

so that, again, the terms $\partial_t u_\lambda$, $\Delta_{\mathbf{x}_1} u_\lambda(t, \mathbf{x}_1, \mathbf{x}_2)$ and $\langle \mathbf{x}_1, D_{\mathbf{x}_2} u_\lambda(t, \mathbf{x}_1, \mathbf{x}_2)\rangle$ associated with the principal part of the partial differential operator in the above PDE are *comparable*. On the other hand, when $\mathbf{x}_2$ belongs to compact subsets of $\mathbb{R}$:

- if $\beta_2^2 > 1/3$, the scaled drift term $\lambda^{-1/2}\langle \mathrm{sgn}(\mathbf{x}_2)(|\lambda^{3/2}\mathbf{x}_2|^{\beta_2^2} \wedge 1), D_{\mathbf{x}_2} u_\lambda(t, \mathbf{x}_1, \mathbf{x}_2)\rangle$ goes to zero with λ, 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 λ goes to zero.

Observe that $1/3 = 1 - 1/(3/2)$ and that $3/2$ is precisely the self similarity index of $\int_0^\cdot W_s ds$, so that we recover the rule (1.7). This threshold has also been shown to be (almost) sharp through counter-examples in [CdR18]. In comparison, the threshold introduced in $(\mathbf{T}_\beta)$ gives $\beta_2^2 = \beta_2 > 2/3 = 1 - 1/[2 \times (3/2)] > 1/3$ 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. $\sum_{i=2}^n \sum_{j=i}^n \lambda^{-i+3/2}\langle \mathrm{sgn}(\mathbf{x}_j)(|\lambda^{j-1/2}\mathbf{x}_j|^{\beta_i^j} \wedge 1), D_{\mathbf{x}_i}\rangle$, explodes when $\beta_i^j < (2i-3)/(2j-1)$, $i \le j$ in $[\![1, n]\!]^2$. These thresholds have also been shown to be (almost) sharp in [CdRM17] through counter-examples. On the other hand, along the diagonal of the system (i.e. for the indexes β_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 $(j-1)^{\text{th}}$ iterated in time integral of the Brownian motion. Still in comparison, we assumed in $(\mathbf{T}_\beta)$ that each $\beta_j^j = \beta_j$ is (strictly) greater than $(2j-2)/(2j-1) = 1 - 1/[2 \times (j-1/2)]$. 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. $\beta_i^j = \beta_j$, $i \in [\![1, n]\!]$). We are hence not able to prove strong well-posedness for (1.5) as soon as $\beta \le 1 - 1/(2\gamma)$, $\gamma = n + 1/2$, $n \in \mathbb{N}^*$. 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 [Zvo74] and Veretennikov [Ver80] 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 [KR05] for $L_q - L_p$ singular drifts (see also [Zha10] and [FF11]). When $\mathcal{W}$ is a pure jump process ($\alpha \in [1, 2)$), Priola showed in [Pri12] that strong well-posedness holds for any $\beta > 1 - \alpha/2$ for bounded Hölder drifts and Chen *et al.* obtained in [CZZ17] 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 [NO02] 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 γ 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

$$\mathbf{X}_t \quad = \quad \mathbf{u}(t, \mathbf{X}_t) + \mathbf{X}_0 - \mathbf{u}(0, \mathbf{X}_0) - M_{0,t}(\alpha, \mathbf{u}, \mathbf{X}) + \mathcal{W}_t, \tag{1.8}$$

with

$$M_{0,t}(\alpha, \mathbf{u}, \mathbf{X}) := \begin{cases} \displaystyle\int_0^t \mathbf{D}\mathbf{u}(s, \mathbf{X}_s) \cdot d\mathcal{W}_s, & \text{if } \alpha = 2; \\ \displaystyle\int_0^t \int_{\mathbb{R}\setminus\{0\}} \tilde{N}(dr, dz)\{\mathbf{u}(s, \mathbf{X}_{s^-} + z) - \mathbf{u}(s, \mathbf{X}_{s^-})\}, & \text{if } \alpha < 2, \end{cases} \tag{1.9}$$

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

$$\partial_t \mathbf{u}(t, \mathbf{x}) + \langle \mathrm{sgn}(\mathbf{x})(|\mathbf{x}|^\beta \wedge 1), \mathbf{D}\mathbf{u}(t, \mathbf{x})\rangle + \frac{1}{2}\Delta^{\frac{\alpha}{2}} \mathbf{u}(t, \mathbf{x}) = \mathrm{sgn}(\mathbf{x})(|\mathbf{x}|^\beta \wedge 1), \quad \mathbf{u}(T, \cdot) = 0. \tag{1.10}$$

Equation (1.8) then follows from the Itô formula provided $\mathbf{u}$ 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 $\mathbf{u}$.

For this new equation to be well-posed in any dimension[3], it is commonly assumed that the integrand in the stochastic integral $M(\alpha, \mathbf{u}, \mathbf{X})$ 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 case $\alpha < 2$, it follows from the interpolation type Lemma 4.1 in [Pri12] 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.* [Kry96], 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 (see (1.16) below as well). From the corresponding parabolic bootstrap (Schauder estimates for (1.10), see e.g. [Fri08, KP10] for $\alpha = 2$ or [CdRMP20a, MP14] in the pure jump case), we could expect at best that $\|u(t, \cdot)\|_{C^{\alpha+\beta}} \leq \|\mathrm{sgn}(\cdot)(|\cdot|^\beta \wedge 1)\|_{C^\beta}$. Hence, we obtain that β must satisfies $\alpha + \beta > 1 + \alpha/2 \Leftrightarrow \beta > 1 - \alpha/2$. In other words, denoting again by γ the self similarity index of the noise, we get that strong well-posedness holds for

$$\beta > 1 - \frac{1}{2\gamma}. \tag{1.11}$$

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 [Kry01] in the $L_q - L_p$ framework of Krylov and Röckner).

[3]Note indeed that the scalar case induces very specific features, see *e.g.* [BC03, GO13, ABM20, CdRM19].

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

$$\mathbf{X}_t = \mathbf{u}(t, \mathbf{X}_t) + \mathbf{x}_0 - \mathbf{u}(0, \mathbf{x}_0) - \int_0^t \mathbf{D}\mathbf{u}(s, \mathbf{X}_s) B dW_s + W_t, \tag{1.12}$$

where $\mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_n)^*$ and each $\mathbf{u}_i$ solves

$$\partial_t \mathbf{u}_i(t, \mathbf{x}) + \langle \underbrace{\mathbf{A}\mathbf{x} + \mathbf{P}(\mathbf{x})}_{=\mathbf{F}(\mathbf{x})}, \mathbf{D}\mathbf{u}_i(t, \mathbf{x}) \rangle + \frac{1}{2} \Delta_{\mathbf{x}_1} \mathbf{u}_i(t, \mathbf{x}) = \mathbf{F}_i(\mathbf{x}). \tag{1.13}$$

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 $B = (1, \mathbf{0}_{1,n-1})^*$), the gradient $D_1\mathbf{u}$ to be Lipschitz in all the spatial directions. We thus ask each component of the function $\mathbf{u}$ 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 β_j the regularity index of any component $\mathbf{F}_i$, i in $[\![1, n]\!]$ w.r.t. the j^{th} variable, j in $[\![i, n]\!]$.

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 $(\mathbf{F}_i)_{i \in [\![1,n]\!]}$ coming from both the Zvonkin transform and the weak Hörmander setting. Especially, we cannot expect the solutions $(\mathbf{u}_i)_{i \in [\![1,n]\!]}$ to be bounded. The associated parabolic bootstrap is thus more tricky than in (1.10). In the companion work [CdRHM20], assuming that the source term in the right-hand side of (1.13) is bounded and β_j-Hölder w.r.t. the j^{th}-variable, we derived Schauder estimates and proved that for any i in $[\![1, n]\!]$, any j in $[\![1, n]\!]$, for fixed $(\mathbf{x}_{1:j-1}, \mathbf{x}_{j+1:n})$ in $\mathbb{R}^{n-1}$ and t in $[0, T]$, the map

$$(\mathbf{u}_i)_j(t, \cdot) : \mathbb{R} \ni z_j \mapsto \mathbf{u}_i(t, \mathbf{x}_{1:j-1}, z_j, \mathbf{x}_{j+1:n}) \in \mathbb{R} \tag{1.14}$$

belongs to $C^{2/(2j-1)+\beta_j}$ for any β_j in $\big(0, 1/[2j-1]\big)$, uniformly w.r.t. t and $(\mathbf{x}_{1:j-1}, \mathbf{x}_{j+1:n})$. This means that the smoothing effect of the hypoelliptic operator $\partial_t + \Delta_{\mathbf{x}_1} + \langle \mathbf{A}\mathbf{x}, \mathbf{D} \rangle$ yields a regularity gain of order $2/[2j-1]$ in the j^{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 $(\mathbf{u}_i)_j(t, \cdot)$ belongs to $C^{2/(2j-1)+\beta_j}$, $(D_1\mathbf{u}_i)_j$ 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^{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 $(\mathbf{u}_i)_{i \in [\![1,n]\!]}$ 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 β_j in $\big((2j-2)/(2j-1), 1\big)$, the map $(D_1\mathbf{u}_i)_j(t, \cdot)$ belongs uniformly to $C^{1/(2j-1)+\beta_j-\varepsilon}$ for any $0 < \varepsilon << 1$ (see Section 1.5.1 for details). To obtain the Lipschitz control in all the spatial directions of the gradient $D_1\mathbf{u}$, we therefore need $1/(2j-1)+\beta_j > 1 \Leftrightarrow \beta_j > (2j-2)/(2j-1)$, which is precisely the thresholds assumed in $(\mathbf{T}_\beta)$.

Thus, one may conclude that the thresholds in $(\mathbf{T}_\beta)$ 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.* [CdR17, WZ16, FFPV17]. 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 [Dav07] and has then been formalized through the works of Flandoli [Fla11a, Fla11b]. The main particularity of that approach relies on the fact that the system (1.5) is considered "ω by ω", 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 [CG16] 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 [CG16], differs significantly from the PDE

based trick of the Zvonkin method. Furthermore, it seems that, *modulo* an additional work, the Zvonkin transform allows to recover, in the Markovian framework, path by path results. We refer *e.g.* to Shaposhnikov [Sha16], who revisited the result by Davie, and to Priola [Pri18, Pri19], who extended the Davie result to the α-stable, α 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 [CG16] and the range corresponding formally to the case $H = 0$ has been investigated recently in [HP20], one may wonder if the range $\gamma = 1/\alpha + n$, n in $\mathbb{N}$ and α 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 natural *proxy process* corresponds to $\bar{\mathbf{X}}$ (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 appropriate 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 characterization of Besov spaces (see *e.g.* Chapter 2.6.4 in [Tri83]) 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 "$\mathcal{W} = Z^\alpha$", α in $(0,2)$, up to a modification of the associated *proxy* and heat kernels, see *e.g.* [HM16, HMP19, Mar20].

Let us try to understand what are the main steps by restricting ourselves to the (already discussed) prototype system (1.4). In this case, the Zvonkin transform associated with (1.4) writes

$$\mathbf{X}_t = \mathbf{u}(t,\mathbf{X}_t) + \mathbf{X}_0 - \mathbf{u}(0,\mathbf{X}_0) - \int_0^t \int_{\mathbb{R}\setminus\{0\}} \tilde{N}_s(ds,dz)(\mathbf{u}(s,\mathbf{X}_{s^-} + Bz) - \mathbf{u}(s,\mathbf{X}_{s^-})) + Z_t^\alpha, \tag{1.15}$$

where $\mathbf{u} = (\mathbf{u}_1,\dots,\mathbf{u}_n)^*$ and each $\mathbf{u}_i$ solves the Integro-Partial Differential Equation (IPDE in short) associated with (1.1) i.e. the Laplacian therein is replaced by a fractional Laplacian $\Delta_{\mathbf{x}_1}^{\alpha/2}$. As already discussed, we need the integrand in the stochastic integral in (1.15) to be Lipschitz w.r.t. the spatial variable. Yet again, a sufficient condition is given by a modification of the interpolation type Lemma 4.1 in [Pri12]: we have for $\mathbf{x}$, $\mathbf{y}$ in $\mathbb{R}^n$ and z in $\mathbb{R}$, $|z| \le 1$:

$$|\mathbf{u}(t,\mathbf{x}+Bz) - \mathbf{u}(t,\mathbf{x}) - \mathbf{u}(t,\mathbf{y}+Bz) + \mathbf{u}(t,\mathbf{y})| \le C|z|^\eta [(\mathbf{Du})_1]_\eta |\mathbf{x}-\mathbf{y}|, \tag{1.16}$$

where $[(\mathbf{Du})_1]_\eta$ denotes the η-Hölder modulus w.r.t. the variable $\mathbf{x}_1$ of the full gradient $\mathbf{Du}$. We therefore have to control this η-Hölder modulus.

In [Mar20], Marino proved Schauder estimates for (more general versions of) the IPDE associated with (1.4). He established therein that for any j in $[\![1,n]\!]$, for fixed $(\mathbf{x}_{1:j-1},\mathbf{x}_{j+1:n})$ in $\mathbb{R}^{n-1}$ and t in $[0,T]$, the map $(\mathbf{u}_i)_j(t,\cdot)$ belongs to $C^{\alpha/[1+\alpha(j-1)]+\beta_j}$ uniformly in t and $(\mathbf{x}_{1:j-1},\mathbf{x}_{j+1:n})$, provided that for each j in $[\![1,n]\!]$: $\alpha > \beta_j$ (this condition comes from integrability purposes), $\alpha + \beta_j > 1$ and β_j in $\big(0, 1/(1+\alpha(j-1))\big)$. This means that the smoothing effect of the hypoelliptic operator $\partial_t + (1/2)\Delta_{\mathbf{x}_1}^{\alpha/2} + \langle \mathbf{Ax}, \mathbf{D}\rangle$ yields a regularity gain of order $\alpha/[1+\alpha(j-1)]$ in the j^{th} variable. We thus understand (at least formally), that for each j in $[\![1,n]\!]$, $D_j\big(\mathbf{u}(t,\cdot)\big)_j$ belongs to $C^{\alpha/[1+\alpha(j-1)]+\beta_j-1}$. Note that, due to homogeneity, controlling the the η-Hölder modulus of $D_j\mathbf{u}(t,\cdot)$ w.r.t. the non-degenerate variable amounts to control the $\eta/[1+\alpha(j-1)]$-Hölder modulus of $\big(D_j\mathbf{u}(t,\cdot)\big)_j$ i.e. a η-Hölder regularity w.r.t. the non-degenerate variable corresponds to a $\eta/[1+\alpha(j-1)]$-Hölder regularity w.r.t. the j^{th} variable. According to the associated parabolic bootstrap, we thus need the

following condition to be fulfilled:

$$\frac{\alpha}{1+\alpha(j-1)}+\beta_j-1\geq\frac{\eta}{1+\alpha(j-1)}>\frac{\alpha/2}{1+\alpha(j-1)}\Leftrightarrow\beta_j>\frac{1+\alpha(j-3/2)}{1+\alpha(j-1)},$$

recalling that, for integrability purposes coming from the Lévy measure associated with the pure jump process, η must again be strictly greater than $\alpha/2$. These are precisely the thresholds obtained in the recent work [HWZ20] in the kinetic case (i.e. when $n=2$). Following our method, we should be able to extend in some sense the Schauder control in order to obtain that for any β_j in $\big((1+\alpha(j-3/2))/(1+\alpha(j-1)),1\big)$ the map $(\mathbf{u}_i)_j(t,\cdot)$ belongs uniformly to $C^{\alpha/[1+\alpha(j-1)]+\beta_j-\varepsilon}$ for any $0<\varepsilon<<1$, so that the result would follow.

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 [CdRM17] 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 [Zvo74] which heavily relies on the connection between SDEs and PDEs. We rewrite (1.1) in the shortened form

$$d\mathbf{X}_t=\mathbf{F}(t,\mathbf{X}_t)dt+B\sigma(t,\mathbf{X}_t)dW_t,$$

where we recall that $\mathbf{F}=(\mathbf{F}_1,\dots,\mathbf{F}_n)$ is an $\mathbb{R}^{nd}$-valued function and B is the embedding matrix from $\mathbb{R}^d$ into $\mathbb{R}^{nd}$, i.e. $B=(\mathbf{I}_{d,d},\mathbf{0}_{d,d},\dots,\mathbf{0}_{d,d})^*=(\mathbf{I}_{d,d},\mathbf{0}_{d,(n-1)d})^*$, where $\mathbf{I}_{d,d},\mathbf{0}_{d,d}$ 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 $(t,\mathbf{x})\in[0,T]\times\mathbb{R}^{nd}$ let

$$L_t\varphi(\mathbf{x})=\langle\mathbf{F}(t,\mathbf{x}),\mathbf{D}\varphi(t,\mathbf{x})\rangle+\frac{1}{2}\mathrm{Tr}\Big(a(t,\mathbf{x})D^2_{\mathbf{x}_1}\varphi(\mathbf{x})\Big),\tag{1.17}$$

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:

$$\begin{cases}(\partial_t\mathbf{u}_i+L_t\mathbf{u}_i)(t,\mathbf{x})=\mathbf{F}_i(t,\mathbf{x}),\ (t,\mathbf{x})\in[0,T)\times\mathbb{R}^{nd},\\ \mathbf{u}_i(T,\mathbf{x})=\mathbf{0}_d,\qquad i\in[\![1,n]\!].\end{cases}\tag{1.18}$$

Remark 2. *Note that above we adopted the following convention for notational convenience: for each i in $[\![3,n]\!]$, all t in $[0,T]$, $\mathbf{x}$ in $\mathbb{R}^{nd}$, $\mathbf{F}_i(t,\mathbf{x}):=\mathbf{F}_i(t,\mathbf{x}_{i-1:n})$ i.e. the independence of each map $\mathbf{F}_i(t,\cdot)$ with respect to the first $i-2$ components of the vector $\mathbf{x}=(\mathbf{x}_1,\dots,\mathbf{x}_n)\in\mathbb{R}^{nd}$ is implicitly assumed.*

Denote by $\mathbf{U}=(\mathbf{u}_1,\dots,\mathbf{u}_n)$ its global solution. Let now $(\mathbf{F}^m)_{m\geq0},(a^m)_{m\geq0}$ 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 $\mathbf{F}^m$, and converging in supremum norm to $(\mathbf{F},a)$ (such sequences are easily obtained from [CdR17]). Then, for each m, the regularized systems of PDEs associated with (1.18) write:

$$\begin{cases}(\partial_t\mathbf{u}_i^m+L_t^m\mathbf{u}_i^m)(t,\mathbf{x})=\mathbf{F}_i^m(t,\mathbf{x}),\ (t,\mathbf{x})\in[0,T)\times\mathbb{R}^{nd},\\ \mathbf{u}_i^m(T,\mathbf{x})=\mathbf{0}_d,\qquad i\in[\![1,n]\!],\end{cases}\tag{1.19}$$

where L_t^m is obtained from (1.17) replacing $\mathbf{F}$ by $\mathbf{F}^m$ and a by a^m.

The above system (1.19) is well-posed and admits a unique smooth solution $\mathbf{U}^m=(\mathbf{u}_1^m,\dots,\mathbf{u}_n^m)$. 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 [GS69] or Talay and Tubaro [TT90]). Hence, applying Itô's formula, one easily deduces that

$$\int_0^t ds\mathbf{F}(s,\mathbf{X}_s)=-\mathbf{U}^m(0,\mathbf{x})+\mathbf{U}^m(t,\mathbf{X}_t)-\int_0^t\mathbf{D}\mathbf{U}^m(s,\mathbf{X}_s)B\sigma(s,\mathbf{X}_s)dW_s+\int_0^t\mathcal{R}_s^m(\mathbf{X}_s)ds,\tag{1.20}$$

where

$$\begin{aligned}\mathcal{R}_s^m(\mathbf{X}_s):=&[\mathbf{F}(s,\mathbf{X}_s)-\mathbf{F}^m(s,\mathbf{X}_s)]-(L_s-L_s^m)\mathbf{U}^m(s,\mathbf{X}_s)\\ =&[\mathbf{F}(s,\mathbf{X}_s)-\mathbf{F}^m(s,\mathbf{X}_s)]-(\mathbf{F}-\mathbf{F}^m)(s,\mathbf{X}_s)\cdot\mathbf{D}\mathbf{U}^m(s,\mathbf{X}_s)\\ &-\frac{1}{2}\mathrm{Tr}\big((a-a^m)(s,\mathbf{X}_s)D^2_{\mathbf{x}_1}\mathbf{U}^m(s,\mathbf{X}_s)\big).\end{aligned}\tag{R_m}$$

This representation (1.20) 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 $\mathbf{U}^m$, $m \geq 0$ will be smoother than the source term $\mathbf{F}$ so that the right-hand side of (1.20) 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 (1.19) 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 $C_T := C_T((\mathbf{A})) > 0$ satisfying $C_T \to 0$ when $T \to 0$ such that for every $m \geq 0$, the solution $\mathbf{U}^m$ satisfies, with the notation of* (1.3)*:*

$$\|\mathbf{D}\mathbf{U}^m\|_\infty + \|\mathbf{D}(D_1\mathbf{U}^m)\|_\infty \leq C_T. \tag{1.21}$$

Remark 3 (On well-posedness of the initial PDE (1.18))**.** *We also point out that, from the uniformity in m in the previous theorem, we could also derive some regularizing properties for the system* (1.18) *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 12). These controls precisely allow to derive, through the Arzelà-Ascoli theorem, a well-posedness result for equation* (1.18) *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.

Proof of Theorem 1. Let now $\mathbf{X}$ and $\mathbf{X}'$ be two solutions of (1.1). Using the representation (1.20) to express the difference of the *bad* drifts in terms of the function $\mathbf{U}^m$ and its derivative up to a remainder, we write:

$$\begin{aligned}
&\mathbf{X}_t - \mathbf{X}'_t \\
= \;& \mathbf{U}^m(t,\mathbf{X}_t) - \mathbf{U}^m(t,\mathbf{X}'_t) - \int_0^t \left[\mathbf{D}\mathbf{U}^m(s,\mathbf{X}_s)B\sigma(s,\mathbf{X}_s) - \mathbf{D}\mathbf{U}^m(s,\mathbf{X}'_s)B\sigma(s,\mathbf{X}'_s)\right] dW_s \\
&+\Big[\mathcal{R}^m_t(\mathbf{X}) - \mathcal{R}^m_t(\mathbf{X}')\Big] + \int_0^t B\left[\sigma(s,\mathbf{X}_s) - \sigma(s,\mathbf{X}'_s)\right] dW_s.
\end{aligned}$$

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:

$$\begin{aligned}
&\mathbb{E}\left[\sup_{t\leq T}|\mathbf{X}_t - \mathbf{X}'_t|^2\right] \\
\leq \;& 5\Big(\mathbb{E}\left[\sup_{t\leq T}|\mathbf{U}^m(t,\mathbf{X}_t) - \mathbf{U}^m(t,\mathbf{X}'_t)|^2\right] + \mathbb{E}\left[\int_0^T ds\,|[\mathbf{D}\mathbf{U}^m B](s,\mathbf{X}_s) - [\mathbf{D}\mathbf{U}^m B](s,\mathbf{X}'_s)|^2\,\|\sigma\|^2_\infty\right] \\
&+\mathbb{E}\left[\int_0^T ds(\|\mathbf{D}\mathbf{U}^m B\|_\infty + \mathbf{1})\,|[\sigma(s,\mathbf{X}_s) - \sigma(s,\mathbf{X}'_s)]|^2\right] + 2T\|\mathcal{R}^m_\cdot(\cdot)\|^2_\infty\Big).
\end{aligned}$$

Note that thanks to the particular structure of B one has $\mathbf{D}\mathbf{U}^m B = (D_1\mathbf{U}^m, \mathbf{0}_{d,d}, \ldots, \mathbf{0}_{d,d})^*$. 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

$$\mathbb{E}\left[\sup_{t\leq T}|\mathbf{X}_t - \mathbf{X}'_t|^2\right] \leq \bar{C}_T\mathbb{E}\left[\sup_{t\leq T}|\mathbf{X}_t - \mathbf{X}'_t|^2\right] + 10T\|\mathcal{R}^m_\cdot(\cdot)\|^2_\infty. \tag{1.22}$$

Letting $m \to +\infty$, since $\|a - a^m\|_\infty$ and $\|\mathbf{F} - \mathbf{F}^m\|_\infty$ tend to 0, it readily follows from (R_m) and the bound (1.21) in Theorem 2 that $\|\mathcal{R}^m_\cdot(\cdot)\|_\infty \underset{m}{\to} 0$. 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. [SV79]. □

[4]By "small enough" we mean that there exists a time $\mathcal{T} > 0$ depending only on known parameters in **(A)** s.t. for any $T \leq \mathcal{T}$ the statement of the theorem holds.

1.5 Regularization properties of the underlying PDE (1.18): strategy of the proof and primer

As mentioned above, the regularization properties of the PDE (1.19) given by estimate (1.21) 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 Hölder coefficients. In that setting, the estimates of Theorem 2 are well known, see *e.g.* the book of Friedman [Fri64] or Bass [Bas97]. For unbounded drift and source terms, those estimates have been recently established in [KP10]. In our case, the story is rather different since the diffusion matrix Ba 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 $(D_{\mathbf{x}_{i-1}}\mathbf{F}_i)_{i\in[\![2,n]\!]}$ allows the noise to propagate in the indicated directions thanks to the drift. It can be viewed as a weak type of Hörmander 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 Hörmander [Hör67] or, for a probabilistic viewpoint, the monograph of Stroock [Str08]). The tricky point in our *weak Hörmander setting* is that the pointwise gradient estimates (1.21) 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 [DM10] for density estimates, Menozzi [Men11], [Men18] and Priola [Pri15] for the martingale problem and also Bramanti *et al.* [BCLP10], [BCLP13] for related L^p estimates and Bramanti and Zhu [BZ11] for the VMO framework). We can mention the work of Lorenzi [Lor05] 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 $\mathbf{F}^m, a^m$, but only on known parameters appearing in **(A)**. Again, this is what would also allow to transfer those bounds to equation (1.18) from a suitable compactness argument, extending well known results for non-degenerate diffusions with Hölder 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 [MS67] or the book of Friedman [Fri64]. 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 (1.19) 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 $\tilde{\mathbf{X}}^m$ of $\mathbf{X}^m$ 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 $(\boldsymbol{\theta}^m_{s,t})_{s\in[t,T]}$ that solves the deterministic counterpart of (1.1) with mollified coefficients (*i.e.* with $\sigma^m \equiv \mathbf{0}_{d,d}$) namely, $\dot{\boldsymbol{\theta}}^m_{s,t} = \mathbf{F}^m(s, \boldsymbol{\theta}^m_{s,t})$. This *proxy* process $\tilde{\mathbf{X}}^m$ may be seen as a (non-linear) generalization of the so-called Kolmogorov example [Kol34] and we refer the reader to the work of Delarue and Menozzi [DM10] and Menozzi [Men11] 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 [CdRM17], 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 (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 $\mathbf{F}_i$ 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 [CdR17]. 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 [CdR17] 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 [CdR17] equilibrated *at hand* through elementary cancellation arguments, whereas the current approach takes advantage of the

[5] Pay attention that this is not the case for L whose coefficients do not have the required smoothness in $(\mathbf{T}_\beta)$ to compute the corresponding Lie brackets.

[6] Meaning that the freezing curve $\boldsymbol{\theta}$ solves the corresponding ODE associated with (1.1) in a forward form.

full-force duality results between Besov spaces (see Sections 1.5.2 and 2.2 below). This forward perturbative approach has also been successfully used in [CdRHM20] to establish some weaker regularization properties of the PDE (1.18) 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 $(\tilde{\mathbf{X}}_t)_{t\geq 0}$ with dynamics:

$$d\tilde{\mathbf{X}}_t = \mathbf{A}_t\tilde{\mathbf{X}}_t dt + BdW_t, \tag{1.23}$$

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

$$\mathbf{A}_t = \begin{pmatrix} \mathbf{0}_{d,d} & \cdots & \cdots & \cdots & \mathbf{0}_{d,d} \\ \mathbf{a}_{2,1}(t) & \mathbf{0}_{d,d} & \cdots & \cdots & \mathbf{0}_{d,d} \\ \mathbf{0}_{d,d} & \mathbf{a}_{3,2}(t) & \mathbf{0}_{d,d} & \cdots & \mathbf{0}_{d,d} \\ \vdots & \mathbf{0}_{d,d} & \ddots & \ddots & \vdots \\ \mathbf{0}_{d,d} & \cdots & \mathbf{0}_{d,d} & \mathbf{a}_{n,n-1}(t) & \mathbf{0}_{d,d} \end{pmatrix}. \tag{1.24}$$

The entries $(\mathbf{a}_{i,i-1}(t))_{i\in[\![2,n]\!]}$ are uniformly in time non-degenerate elements of $\mathbb{R}^d\otimes\mathbb{R}^d$ (which expresses the weak Hörmander condition). The corresponding generator $\tilde{L}_t$ writes for any $\varphi\in C_0^2(\mathbb{R}^{nd},\mathbb{R})$:

$$\tilde{L}_t\varphi(\mathbf{x}) = \langle \mathbf{A}_t\mathbf{x}, \mathbf{D}\varphi(\mathbf{x})\rangle + \frac{1}{2}\Delta_{\mathbf{x}_1}\varphi(\mathbf{x}).$$

In such a case, each component $\tilde{\mathbf{u}}_i$, $i\in[\![1,n]\!]$ of the solution $\tilde{\mathbf{U}}$ of the corresponding system of PDEs

$$\begin{cases} (\partial_t + \tilde{L}_t)\tilde{\mathbf{U}}(t,\mathbf{x}) = \mathbf{F}(t,\mathbf{x}),\ (t,\mathbf{x})\in[0,T)\times\mathbb{R}^{nd}, \\ \tilde{\mathbf{U}}(T,\mathbf{x}) = \mathbf{0}_{nd}, \end{cases} \tag{1.25}$$

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

$$\tilde{\mathbf{u}}_i(t,\mathbf{x}) = -\int_t^T ds\int_{\mathbb{R}^{nd}} d\mathbf{y}\mathbf{F}_i(s,\mathbf{y})\tilde{p}(t,s,\mathbf{x},\mathbf{y}). \tag{1.26}$$

Above, $\tilde{p}$ stands for the transition density of the Gaussian process $(\tilde{\mathbf{X}}_v)_{v\geq 0}$ with dynamics (1.23). Using the resolvent associated with $(\mathbf{A}_v)_{v\in[t,s]}$, i.e. $\partial_s\tilde{\mathbf{R}}_{s,t} = \mathbf{A}_s\tilde{\mathbf{R}}_{s,t}$, $\tilde{\mathbf{R}}_{t,t} = \mathbf{I}_{nd,nd}$, the above equation can be explicitly integrated. Precisely, for a fixed starting point $\mathbf{x}$ at time t:

$$\tilde{\mathbf{X}}_v = \tilde{\mathbf{R}}_{v,t}\mathbf{x} + \int_t^v \tilde{\mathbf{R}}_{v,u}BdW_u. \tag{1.27}$$

Hence, at time $s>t$ the covariance matrix of the random variable $\tilde{\mathbf{X}}_s$ writes as $\tilde{\mathbf{K}}_{s,t} := \int_t^s du\tilde{\mathbf{R}}_{s,u}BB^*\tilde{\mathbf{R}}_{s,u}^*$. From (1.27), the density at time $v=s$ and at the spatial point $\mathbf{y}$ therefore writes:

$$\tilde{p}(t,s,\mathbf{x},\mathbf{y}) = \frac{1}{(2\pi)^{\frac{nd}{2}}\det(\tilde{\mathbf{K}}_{s,t})^{\frac{1}{2}}}\exp\left(-\frac{1}{2}\Big\langle(\tilde{\mathbf{K}}_{s,t})^{-1}\big(\tilde{\mathbf{R}}_{s,t}\mathbf{x}-\mathbf{y}\big),\tilde{\mathbf{R}}_{s,t}\mathbf{x}-\mathbf{y}\Big\rangle\right). \tag{1.28}$$

Note that the resolvent also appears in (1.27) and in the density. Since the drift in (1.23) is unbounded, the term $\tilde{\mathbf{R}}_{s,t}\mathbf{x}$ actually corresponds to the transport of the initial condition $\mathbf{x}$ through the associated deterministic and linear differential system. It is well known, see e.g. [DM10] and Section 2.1 below, that the covariance $\tilde{\mathbf{K}}_{s,t}$ enjoys what we will call a *good scaling property*. Precisely, for a given $T>0$ there exists $C := C\big((\mathbf{A}_v)_{v\in[0,T]},T\big)\geq 1$ s.t. for any $\boldsymbol{\xi}\in\mathbb{R}^{nd}$,

$$C^{-1}(s-t)^{-1}|\mathbb{T}_{s-t}\boldsymbol{\xi}|^2 \leq \langle\tilde{\mathbf{K}}_{s,t}\boldsymbol{\xi},\boldsymbol{\xi}\rangle \leq C(s-t)^{-1}|\mathbb{T}_{s-t}\boldsymbol{\xi}|^2, \tag{1.29}$$

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

$$\mathbb{T}_u = \begin{pmatrix} u\mathbf{I}_{d,d} & \mathbf{0}_{d,d} & \cdots & \mathbf{0}_{d,d} \\ \mathbf{0}_{d,d} & u^2\mathbf{I}_{d,d} & \mathbf{0}_{d,d} & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0}_{d,d} & \cdots & \mathbf{0}_{d,d} & u^n\mathbf{I}_{d,d} \end{pmatrix}. \tag{1.30}$$

Importantly, the *good scaling property* stated in (1.29) indicates that, for a given initial time t and for each $i\in[\![1,n]\!]$, each $\mathbb{R}^d$-valued component $\tilde{\mathbf{X}}^i_s$ 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 $\tilde{p}$ also satisfies the bound

$$\tilde{p}(t,s,\mathbf{x},\mathbf{y}) \le \frac{C}{(s-t)^{\frac{n^2d}{2}}}\exp\left(-C^{-1}(s-t)|\mathbb{T}^{-1}_{s-t}(\tilde{\mathbf{R}}_{s,t}\mathbf{x}-\mathbf{y})|^2\right) =: \bar{C}\hat{p}_{C^{-1}}(t,s,\mathbf{x},\mathbf{y}),$$

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

$$\begin{aligned} |D_{\mathbf{x}_l}D^r_{\mathbf{x}_1}\tilde{p}(t,s,\mathbf{x},\mathbf{y})| &\le \frac{C}{(s-t)^{\frac{n^2d}{2}+(l-\frac12)+\frac{r}{2}}}\exp\left(-C^{-1}(s-t)|\mathbb{T}^{-1}_{s-t}(\tilde{\mathbf{R}}_{s,t}\mathbf{x}-\mathbf{y})|^2\right) \\ &\le \frac{\bar{C}}{(s-t)^{(l-\frac12)+\frac{r}{2}}}\hat{p}_{C^{-1}}(t,s,\mathbf{x},\mathbf{y}), \end{aligned} \tag{1.31}$$

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 (1.21) of Theorem 2 for the current system (1.25), it follows from the specific structure of the matrix B that we have to estimate for any $l\in[\![1,n]\!]$ the quantities $D_{\mathbf{x}_l}D^r_{\mathbf{x}_1}\tilde{\mathbf{u}}_i(t,\mathbf{x})$, $r\in\{0,1\}$. From (1.31), we thus have

$$|D_{\mathbf{x}_l}D^r_{\mathbf{x}_1}\tilde{\mathbf{u}}_i(t,\mathbf{x})| \le \bar{C}\int_t^T ds\int_{\mathbb{R}^{nd}} d\mathbf{y}|\mathbf{F}_i(s,\mathbf{y})|(s-t)^{-(l-\frac12)-\frac{r}{2}}\hat{p}_{C^{-1}}(t,s,\mathbf{x},\mathbf{y}). \tag{1.32}$$

We now face two problems: first the $\mathbf{F}_i$ 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 $\mathbf{F}_i$ exploiting precisely the fact that, once integrated through the variables $\mathbf{y}_l$ to $\mathbf{y}_n$, the transition density $\tilde{p}$ does not depend on the variable $\mathbf{x}_l$ anymore. This is due to the structure of $\mathbf{A}$ in (1.24), which in particular yields that the resolvent $(\tilde{\mathbf{R}}_{s,t})_{0\le t\le s\le T}$ is lower triangular. Precisely, denoting for conciseness by $\boldsymbol{\theta}_{s,t}(\mathbf{x})=\tilde{\mathbf{R}}_{s,t}\mathbf{x}$ (which is coherent with the notation below when handling non-linear flows), we write:

$$\int_{\mathbb{R}^{(l-1)d}} d\mathbf{y}_{1:l-1}\mathbf{F}_i(t,\mathbf{y}_1,\dots,\mathbf{y}_{l-1},\boldsymbol{\theta}^l_{s,t}(\mathbf{x}),\dots,\boldsymbol{\theta}^n_{s,t}(\mathbf{x}))D_{\mathbf{x}_1}D_{\mathbf{x}_l}\int_{\mathbb{R}^{(n-(l-1))d}} d\mathbf{y}_{l:n}\tilde{p}(t,s,\mathbf{x},\mathbf{y})=0.$$

This is what will be called a *cancellation (or centering) argument* in the following. When using this property, we obtain that

$$\begin{aligned} &|D_{\mathbf{x}_1}D_{\mathbf{x}_l}\tilde{\mathbf{u}}_i(t,\mathbf{x})| \\ &= \left|\int_t^T ds\int_{\mathbb{R}^{nd}} d\mathbf{y}\left(\mathbf{F}_i(s,\mathbf{y})-\mathbf{F}_i(t,\mathbf{y}_1,\dots,\mathbf{y}_{l-1},\boldsymbol{\theta}^l_{s,t}(\mathbf{x}),\dots,\boldsymbol{\theta}^n_{s,t}(\mathbf{x}))\right)D_{\mathbf{x}_1}D_{\mathbf{x}_l}\tilde{p}(t,s,\mathbf{x},\mathbf{y})\right|. \end{aligned}$$

We thus obtain from (1.31):

$$\begin{aligned} &|D_{\mathbf{x}_1}D_{\mathbf{x}_l}\tilde{\mathbf{u}}_i(t,\mathbf{x})| \\ \le\ & \bar{C}\int_t^T ds\int_{\mathbb{R}^{nd}} d\mathbf{y}\left|\mathbf{F}_i(s,\mathbf{y})-\mathbf{F}_i(t,\mathbf{y}_1,\dots,\mathbf{y}_{l-1},\boldsymbol{\theta}^l_{s,t}(\mathbf{x}),\dots,\boldsymbol{\theta}^n_{s,t}(\mathbf{x}))\right|(s-t)^{-(l-\frac12)-\frac12}\hat{p}_{C^{-1}}(t,s,\mathbf{x},\mathbf{y}). \end{aligned}$$

Then, using the regularity assumed of $\mathbf{F}_i$, which satisfies $(\mathbf{T}_\beta)$, we get that for some constants $\bar{C}, C$ (which possibly change from line to line)

$$
\begin{aligned}
&|D_{\mathbf{x}_1} D_{\mathbf{x}_l} \tilde{\mathbf{u}}_i(t,\mathbf{x})| \\
\leq &\bar{C} \int_t^T ds \int_{\mathbb{R}^{nd}} d\mathbf{y} \sum_{j=l}^n |\mathbf{y}_j - \boldsymbol{\theta}^j_{s,t}(\mathbf{x})|^{\beta_j} (s-t)^{-(l-\frac{1}{2})-\frac{1}{2}} \hat{p}_{C^{-1}}(t,s,\mathbf{x},\mathbf{y}) \\
\leq &\bar{C} \int_t^T ds \int_{\mathbb{R}^{nd}} d\mathbf{y} \sum_{j=l}^n \left| \frac{\mathbf{y}_j - \boldsymbol{\theta}^j_{s,t}(\mathbf{x})}{(s-t)^{j-\frac{1}{2}}} \right|^{\beta_j} (s-t)^{\beta_j(j-\frac{1}{2})} (s-t)^{-(l-\frac{1}{2})-\frac{1}{2}} \hat{p}_{C^{-1}}(t,s,\mathbf{x},\mathbf{y}) \qquad \text{(Abs)} \\
\leq &\bar{C} \int_t^T ds \int_{\mathbb{R}^{nd}} d\mathbf{y} \sum_{j=l}^n (s-t)^{-(l-\frac{1}{2})-\frac{1}{2}+\beta_j(j-\frac{1}{2})} \hat{p}_{C^{-1}}(t,s,\mathbf{x},\mathbf{y}),
\end{aligned}
$$

thanks to Remark 4 below for the derivation of the last inequality. Note that the above term is integrable only if for each $j \in [\![l,n]\!]$, $-(l-1/2)-1/2+\beta_j(j-1/2) > -1 \iff \beta_j > \big((2l-2)/(2j-1)\big)$. This condition actually holds if for any $i \in [\![1,n]\!]$, $\beta_i > \big((2i-2)/(2i-1)\big)$ which is exactly the infimum assumed in $(\mathbf{T}_\beta)$. 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.

Remark 4 (On the absorption of spatial differences by the *proxy* density)**.** *From a technical viewpoint, we also insist here that in the above inequalities in* (Abs)*, we explicitly made the contribution* $(s-t)^{-(j-1/2)}|\mathbf{y}_j - \boldsymbol{\theta}^j_{s,t}(\mathbf{x})|$ *appear since it precisely corresponds to the absolute value of the* j^{th} *entry of the vector* $(t-s)^{1/2}\mathbb{T}^{-1}_{t-s}(\boldsymbol{\theta}_{s,t}(\mathbf{x})-\mathbf{y})$ *which also precisely appears with the current notation* $\boldsymbol{\theta}_{s,t}(\mathbf{x}) = \tilde{\mathbf{R}}_{s,t}\mathbf{x}$*, in its square norm, in the off-diagonal exponential bound of* (1.31)*. This is why for the last inequality this contribution disappeared, i.e. it can be absorbed by the exponential (again up to a modification of the concentration constant* C *which is modified at most a finite number of times).*

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 $\tilde{\mathbf{X}}^m$ with generator $\tilde{L}^m$. Observe that equation (1.19) can be equivalently rewritten as:

$$
\begin{cases}
(\partial_t \mathbf{u}^m_i + \tilde{L}^m_t \mathbf{u}^m_i)(t,\mathbf{x}) = \mathbf{F}^m_i(t,\mathbf{x}) - (L^m_t - \tilde{L}^m_t)\mathbf{u}^m_i(t,\mathbf{x}), \ (t,\mathbf{x}) \in [0,T)\times\mathbb{R}^{nd}, \\
\mathbf{u}^m_i(T,\mathbf{x}) = \mathbf{0}_d, \qquad i \in [\![1,n]\!].
\end{cases}
$$

Since $\mathbf{u}^m_i(t,\mathbf{x})$ is smooth (see the arguments following (1.19)), so is the term $(L^m_t - \tilde{L}^m_t)\mathbf{u}^m_i(t,\mathbf{x})$ 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 $\mathbf{u}^m_i$, $i \in [\![1,n]\!]$ of our solution $\mathbf{U}^m$ of the systems (1.19) it holds that for any $(t,\mathbf{x}) \in [0,T]\times\mathbb{R}^{nd}$

$$
\mathbf{u}^m_i(t,\mathbf{x}) = \int_t^T ds \int_{\mathbb{R}^{nd}} d\mathbf{y} \Big\{ -\mathbf{F}^m_i(s,\mathbf{y}) + (L^m_s - \tilde{L}^m_s)\mathbf{u}^m_i(s,\mathbf{y}) \Big\} \tilde{p}^m(t,s,\mathbf{x},\mathbf{y}). \tag{1.33}
$$

This corresponds to the so-called first order parametrix expansion of the initial equation (1.19) around the *proxy* generator $\tilde{L}^m$. Above, the additional term in the right-hand side is, in comparison with (1.26), 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 (1.21) 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 $\mathbf{u}^m_i(t,\mathbf{x})$, we will obtain the required estimate provided the multiplicative constants associated with the terms $\|\mathbf{D}\mathbf{u}_i{}^m\|_\infty$ and $\|D_1\mathbf{D}\mathbf{u}_i{}^m\|_\infty$, that will appear in the corresponding upper-bound for the above right-hand side, are small enough (see also Section 2 of [CdR17] and Section 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 $\int_t^T ds \int_{\mathbb{R}^{nd}} d\mathbf{y}(L_s^m - \tilde{L}_s^m)\mathbf{u}_i^m(s,\mathbf{y})\tilde{p}^m(t,s,\mathbf{x},\mathbf{y})$ 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 $D_\ell D_j \mathbf{u}_i^m$, $\ell, j \in [\![1,n]\!]$ 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 [CdR17], the only thing we could hope is that the gradient in the degenerate directions viewed as a function of the degenerate variables, i.e. $\mathbf{D}_{2:n}\mathbf{u}_i^m(t,\mathbf{x}_1,\cdot) := \big(D_2\mathbf{u}_i^m(t,\mathbf{x}_1,\cdot),\ldots,D_n\mathbf{u}_i^m(t,\mathbf{x}_1,\cdot)\big)^*$ for any $(t,\mathbf{x}_1) \in [0,T]\times\mathbb{R}^d$, 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 $\mathbf{D}_{2:n}\mathbf{u}_i^m(t,\mathbf{x}_1,\cdot)$, 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 Section 2 for details and to Proposition 3.6 in the book of Lemarié-Rieusset [LR02] 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 $\mathbf{D}u_i^m$. 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 $\mathbf{F}_1$ belongs to a suitable $L_q - L_p$ space w.r.t. time and the non-degenerate variable $\mathbf{x}_1$. 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 $\mathbf{U}^m$, namely $D_{\mathbf{x}_1}D_{\mathbf{x}_1}\mathbf{u}_1^m$ (which is a part of the main estimate (1.21) in Theorem 2). The point is to establish for this quantity an $L_q - L_p$ control. This cannot be derived from the previously described approach and requires harmonic analysis techniques (see also [Kry01]). The main problem to establish the estimate is mainly due to the source term, which is actually $\mathbf{F}_1$. To prove it, the main idea consists in exploiting the results of Menozzi [Men18] (where such an estimate is proved under the assumption that the drift is Lipschitz) through the tools developed in [CdRM17] (*backward* parametrix approach for drift $\mathbf{F}$ 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.[Ver80] and [FFPV17]). 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.

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 $\mathbf{F}$, a are infinitely differentiable functions with bounded derivatives of all orders for a and greater than 1 for the coefficient $\mathbf{F}$.

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 (1.17) 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 $(\mathbf{F}_i)_{i\in[\![1,n]\!]}$ in the dynamics (1.1)). Namely, we concentrate on

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

where $(L_t)_{t\geq 0}$ is defined in (1.17) 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 $\gamma := \gamma((\mathbf{A})) > 0$ and $C := C((\mathbf{A})) > 0$ s.t.*

$$\|\mathbf{D}u\|_\infty + \|\mathbf{D}(D_1 u)\|_\infty \leq CT^\gamma, \tag{2.2}$$

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

The proof of Theorem 3 is performed in Section 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 Section 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 $\mathbf{U} = (\mathbf{u}_1,\dots,\mathbf{u}_n)$ where for $i \in [\![1,n]\!]$, $\mathbf{u}_i$ solves (1.18), we derive that $\mathbf{u}_i^j$, $j \in [\![1,d]\!]$ solves (2.1) with $f(t,\mathbf{x}) = -\mathbf{F}_i^j(t,\mathbf{x})$. Theorem 2 thus follows in whole generality applying Theorem 3 to each component of the solution of the systems (1.18). □

2.1 Gaussian *proxy* and associated controls

2.1.1 Linearization of the dynamics

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

$$\begin{aligned} d\tilde{\mathbf{X}}_v^{(\tau,\boldsymbol{\xi})} &= [\mathbf{F}(v,\boldsymbol{\theta}_{v,\tau}(\boldsymbol{\xi})) + D\mathbf{F}(v,\boldsymbol{\theta}_{v,\tau}(\boldsymbol{\xi}))(\tilde{\mathbf{X}}_v^{(\tau,\boldsymbol{\xi})} - \boldsymbol{\theta}_{v,\tau}(\boldsymbol{\xi}))]dv + B\sigma(v,\boldsymbol{\theta}_{v,\tau}(\boldsymbol{\xi}))dW_v,\ \forall v\in[t,s], \\ \tilde{\mathbf{X}}_t^{(\tau,\boldsymbol{\xi})} &= \mathbf{x}, \end{aligned} \tag{2.3}$$

where

$$\dot{\boldsymbol{\theta}}_{v,\tau}(\boldsymbol{\xi}) = \mathbf{F}(v,\boldsymbol{\theta}_{v,\tau}(\boldsymbol{\xi})),\ v\in[0,T],\ \boldsymbol{\theta}_{\tau,\tau}(\boldsymbol{\xi}) = \boldsymbol{\xi}, \tag{2.4}$$

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

$$D\mathbf{F}(v,\mathbf{z}) = \begin{pmatrix} \mathbf{0}_{d,d} & \cdots & \cdots & \cdots & \mathbf{0}_{d,d} \\ D_{\mathbf{z}_1}\mathbf{F}_2(v,\mathbf{z}) & \mathbf{0}_{d,d} & \cdots & \cdots & \mathbf{0}_{d,d} \\ \mathbf{0}_{d,d} & D_{\mathbf{z}_2}\mathbf{F}_3(v,\mathbf{z}_{2:n}) & \mathbf{0}_{d,d} & \mathbf{0}_{d,d} & \vdots \\ \vdots & \mathbf{0}_{d,d} & \ddots & \ddots & \vdots \\ \mathbf{0}_{d,d} & \cdots & \mathbf{0}_{d,d} & D_{\mathbf{z}_{n-1}}\mathbf{F}_n(v,\mathbf{z}_{n-1},\mathbf{z}_n) & \mathbf{0}_{d,d} \end{pmatrix}.$$

In the following, we will often refer to the Gaussian process $(\tilde{\mathbf{X}}_v^{(\tau,\boldsymbol{\xi})})_{v\geq t}$ 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 $v \in [t,s]$:

$$\begin{aligned}
&\tilde{\mathbf{X}}_v^{(\tau,\boldsymbol{\xi})} \\
&= \tilde{\mathbf{R}}^{(\tau,\boldsymbol{\xi})}(v,t)\mathbf{x} + \int_t^v du \tilde{\mathbf{R}}^{(\tau,\boldsymbol{\xi})}(v,u)\Big(\mathbf{F}(u,\boldsymbol{\theta}_{u,\tau}(\boldsymbol{\xi})) - D\mathbf{F}(u,\boldsymbol{\theta}_{u,\tau}(\boldsymbol{\xi}))\boldsymbol{\theta}_{u,\tau}(\boldsymbol{\xi})\Big) + \int_t^v \tilde{\mathbf{R}}^{(\tau,\boldsymbol{\xi})}(v,u)B\sigma(u,\boldsymbol{\theta}_{u,\tau}(\boldsymbol{\xi}))dW_u \\
&=: \mathbf{m}_{v,t}^{(\tau,\boldsymbol{\xi})}(\mathbf{x}) + \int_t^v \tilde{\mathbf{R}}^{(\tau,\boldsymbol{\xi})}(v,u)B\sigma(u,\boldsymbol{\theta}_{u,\tau}(\boldsymbol{\xi}))dW_u, \qquad (2.5)
\end{aligned}$$

where $(\tilde{\mathbf{R}}^{(\tau,\boldsymbol{\xi})}(v,u))_{t\le u,v\le s}$ stands for the resolvent associated with the partial gradients $(D\mathbf{F}(v,\boldsymbol{\theta}_{v,\tau}(\boldsymbol{\xi})))_{v\in[t,s]}$ which satisfies for $v \in [t,s]$:

$$\partial_v \tilde{\mathbf{R}}^{(\tau,\boldsymbol{\xi})}(v,t) = D\mathbf{F}(v,\boldsymbol{\theta}_{v,\tau}(\boldsymbol{\xi}))\tilde{\mathbf{R}}^{(\tau,\boldsymbol{\xi})}(v,t),\ \tilde{\mathbf{R}}^{(\tau,\boldsymbol{\xi})}(t,t) = \mathbf{I}_{nd\times nd}. \qquad (2.6)$$

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

Also, for $v \in [t,s]$, we recall that $\mathbf{m}_{v,t}^{(\tau,\boldsymbol{\xi})}(\mathbf{x})$ stands for the mean of $\tilde{\mathbf{X}}_v^{(\tau,\boldsymbol{\xi})}$ and corresponds as well to the solution of (2.3) when $\sigma = 0$ and the starting point is $\mathbf{x}$. Importantly, we point out that $\mathbf{x} \in \mathbb{R}^{nd} \mapsto \mathbf{m}_{v,t}^{(\tau,\boldsymbol{\xi})}(\mathbf{x})$ is **affine** w.r.t. the starting point $\mathbf{x}$. Precisely, for $\mathbf{x},\mathbf{x}' \in \mathbb{R}^{nd}$:

$$\mathbf{m}_{v,t}^{(\tau,\boldsymbol{\xi})}(\mathbf{x}+\mathbf{x}') = \tilde{\mathbf{R}}^{(\tau,\boldsymbol{\xi})}(v,t)\mathbf{x}' + \mathbf{m}_{v,t}^{(\tau,\boldsymbol{\xi})}(\mathbf{x}). \qquad (2.7)$$

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

$$d\mathbf{m}_{v,t}^{(\tau,\boldsymbol{\xi})}(\mathbf{x}) = [\mathbf{F}(v,\boldsymbol{\theta}_{v,\tau}(\boldsymbol{\xi})) + D\mathbf{F}(v,\boldsymbol{\theta}_{v,\tau}(\boldsymbol{\xi}))(\mathbf{m}_{v,t}^{(\tau,\boldsymbol{\xi})}(\mathbf{x}) - \boldsymbol{\theta}_{v,\tau}(\boldsymbol{\xi}))]dv,\quad \mathbf{m}_{t,t}^{(\tau,\boldsymbol{\xi})}(\mathbf{x}) = \mathbf{x},$$

it holds from (2.4) that for $(\tau,\boldsymbol{\xi}) = (t,\mathbf{x})$:

$$\begin{aligned}
d\big(\mathbf{m}_{v,t}^{(t,\mathbf{x})}(\mathbf{x}) - \boldsymbol{\theta}_{v,t}(\mathbf{x})\big) &= [\mathbf{F}(v,\boldsymbol{\theta}_{v,t}(\mathbf{x})) + D\mathbf{F}(v,\boldsymbol{\theta}_{v,t}(\mathbf{x}))(\mathbf{m}_{v,t}^{(t,\mathbf{x})}(\mathbf{x}) - \boldsymbol{\theta}_{v,t}(\mathbf{x}))]dv - \mathbf{F}(v,\boldsymbol{\theta}_{v,t}(\mathbf{x}))dv \\
&= [D\mathbf{F}(v,\boldsymbol{\theta}_{v,t}(\mathbf{x}))(\mathbf{m}_{v,t}^{(t,\mathbf{x})}(\mathbf{x}) - \boldsymbol{\theta}_{v,t}(\mathbf{x}))]dv, \\
\mathbf{m}_{t,t}^{(t,\mathbf{x})}(\mathbf{x}) - \boldsymbol{\theta}_{t,t}(\mathbf{x}) &= 0.
\end{aligned}$$

Therefore, the Grönwall lemma yields:

$$\mathbf{m}_{v,t}^{(\tau,\boldsymbol{\xi})}(\mathbf{x})|_{(\tau,\boldsymbol{\xi})=(t,\mathbf{x})} = \mathbf{m}_{v,t}^{(t,\mathbf{x})}(\mathbf{x}) = \boldsymbol{\theta}_{v,t}(\mathbf{x}). \qquad (\mathrm{LF})$$

Namely, when freezing the parameters at the initial time t and the starting spatial point $\mathbf{x}$, the linearized flow $\mathbf{m}_{v,t}^{(t,\mathbf{x})}(\mathbf{x})$ and the non-linear one $\boldsymbol{\theta}_{v,t}(\mathbf{x})$ coincide.

From the non-degeneracy of σ and the Hörmander like condition, the Gaussian process defined by (2.5) admits a density $\tilde{p}^{(\tau,\boldsymbol{\xi})}(t,s,\mathbf{x},\cdot)$ which is suitably controlled (see Proposition 5 below and for instance [DM10], [CdRM17]).

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 $\tilde{\mathbf{X}}_v^{(\tau,\boldsymbol{\xi})}$ in* (2.5) *writes:*

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

Uniformly in $(\tau,\boldsymbol{\xi}) \in [0,T]\times\mathbb{R}^{nd}$ and $s \in [0,T]$, it satisfies a good scaling property in the sense of Definition 3.2 in [DM10] (see also Proposition 3.4 of that reference). That is, for any fixed $T > 0$, there exists $C_{2.8} := C_{2.8}((\mathbf{A}),T) \ge 1$ s.t. for all $0 \le t < v \le s \le T$, for any $(\tau,\boldsymbol{\xi}) \in [0,T]\times\mathbb{R}^{nd}$:

$$\forall \boldsymbol{\zeta} \in \mathbb{R}^{nd},\ C_{2.8}^{-1}(v-t)^{-1}|\mathbb{T}_{v-t}\boldsymbol{\zeta}|^2 \le \langle \tilde{\mathbf{K}}_{v,t}^{(\tau,\boldsymbol{\xi})}\boldsymbol{\zeta},\boldsymbol{\zeta}\rangle \le C_{2.8}(v-t)^{-1}|\mathbb{T}_{v-t}\boldsymbol{\zeta}|^2, \qquad (2.8)$$

where we again use the notation introduced in (1.30) *for the scaling matrix $\mathbb{T}_{v-t}$.*

Under **(A)**, Proposition 4 readily follows from Proposition 3.4 in [DM10] (see also the scaling Lemma 3.6 therein). A complete proof is provided in Appendix C.1.1 below.

Remark 5 (On some important consequences of the good scaling property)**.** *We state here some rather direct yet important controls that follow from Proposition 4.*

- *Setting for any* $s \in (t,T]$,

$$\widehat{\tilde{\mathbf{K}}_1^{(\tau,\boldsymbol{\xi}),s,t}} := (s-t)\mathbb{T}_{s-t}^{-1}\tilde{\mathbf{K}}_{s,t}^{(\tau,\boldsymbol{\xi})}\mathbb{T}_{s-t}^{-1}, \tag{S}$$

it follows from (2.8) *that for any* $\boldsymbol{\zeta} \in \mathbb{R}^{nd}$*:*

$$C_{2.8}^{-1}|\boldsymbol{\zeta}|^2 \le \langle \widehat{\tilde{\mathbf{K}}_1^{(\tau,\boldsymbol{\xi}),s,t}}\boldsymbol{\zeta},\boldsymbol{\zeta}\rangle \le C_{2.8}|\boldsymbol{\zeta}|^2 \Longrightarrow \exists \bar{C} := \bar{C}((\mathbf{A}),T) \ge 1,\ \bar{C}^{-1} \le \det(\widehat{\tilde{\mathbf{K}}_1^{(\tau,\boldsymbol{\xi}),s,t}}) \le \bar{C}. \tag{S_{ND}}$$

- *Equations* (S) *and* (S_{ND}) *then readily give that:*

$$\bar{C}^{-1}(s-t)^{n^2d} \le \det(\tilde{\mathbf{K}}_{s,t}^{(\tau,\boldsymbol{\xi})}) \le \bar{C}(s-t)^{n^2d}, \tag{D}$$

as well as, for any $\boldsymbol{\zeta} \in \mathbb{R}^{nd}$*,*

$$C_{2.8}^{-1}|\boldsymbol{\zeta}|^2 \le \langle (\widehat{\tilde{\mathbf{K}}_1^{(\tau,\boldsymbol{\xi}),s,t}})^{-1}\boldsymbol{\zeta},\boldsymbol{\zeta}\rangle \le C_{2.8}|\boldsymbol{\zeta}|^2 \Rightarrow C_{2.8}^{-1}(s-t)|\mathbb{T}_{s-t}^{-1}\boldsymbol{\zeta}|^2 \le \langle (\tilde{\mathbf{K}}_{s,t}^{(\tau,\boldsymbol{\xi})})^{-1}\boldsymbol{\zeta},\boldsymbol{\zeta}\rangle \le C_{2.8}(s-t)|\mathbb{T}_{s-t}^{-1}\boldsymbol{\zeta}|^2. \tag{$\mathrm{S}_{ND}^{\mathrm{Inv}}$}$$

This last bound will be of particular relevance to control the tails of the Gaussian density of $\tilde{\mathbf{X}}_s^{(\tau,\boldsymbol{\xi})}$*.*

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* $\tilde{\mathbf{X}}_s^{(\tau,\boldsymbol{\xi})}$ *in* (2.5) *admits a Gaussian density* $\tilde{p}^{(\tau,\boldsymbol{\xi})}(t,s,\mathbf{x},\cdot)$ *which writes for any* $\mathbf{y} \in \mathbb{R}^{nd}$*:*

$$\tilde{p}^{(\tau,\boldsymbol{\xi})}(t,s,\mathbf{x},\mathbf{y}) := \frac{1}{(2\pi)^{\frac{nd}{2}}\det(\tilde{\mathbf{K}}_{s,t}^{(\tau,\boldsymbol{\xi})})^{\frac{1}{2}}}\exp\left(-\frac{1}{2}\left\langle (\tilde{\mathbf{K}}_{s,t}^{(\tau,\boldsymbol{\xi})})^{-1}(\mathbf{m}_{s,t}^{(\tau,\boldsymbol{\xi})}(\mathbf{x})-\mathbf{y}), \mathbf{m}_{s,t}^{(\tau,\boldsymbol{\xi})}(\mathbf{x})-\mathbf{y}\right\rangle\right), \tag{2.9}$$

with $\tilde{\mathbf{K}}_{s,t}^{(\tau,\boldsymbol{\xi})}$ *as in Proposition 4. Also, there exist* $C := C((\mathbf{A}),T) > 0$ *and* $\bar{C} := \bar{C}(C)$ *s.t. for all* $l \in [\![1,n]\!]$*,* $q,\ r \in \{0,1\}$*, we have:*

$$\begin{aligned} |D_{\mathbf{x}_l}^q D_{\mathbf{x}_1}^r \tilde{p}^{(\tau,\boldsymbol{\xi})}(t,s,\mathbf{x},\mathbf{y})| &\le \frac{C}{(s-t)^{\frac{n^2d}{2}+(l-\frac{1}{2})q+\frac{r}{2}}}\exp\left(-C^{-1}(s-t)\big|\mathbb{T}_{s-t}^{-1}\big(\mathbf{m}_{s,t}^{(\tau,\boldsymbol{\xi})}(\mathbf{x})-\mathbf{y}\big)\big|^2\right) \\ &=: \frac{\bar{C}}{(s-t)^{(l-\frac{1}{2})q+\frac{r}{2}}}\hat{p}_{C^{-1}}^{(\tau,\boldsymbol{\xi})}(t,s,\mathbf{x},\mathbf{y}). \end{aligned} \tag{2.10}$$

Proof. Expression (2.9) readily follows from (2.5). The control (2.10) is then a direct consequence of Proposition 4 for $q = r = 0$ (upper bound of the density). Differentiating w.r.t. $\mathbf{x}$ recalling from (2.7) that $\mathbf{x} \mapsto \mathbf{m}_{s,t}^{(\tau,\boldsymbol{\xi})}(\mathbf{x})$ is affine yields:

$$D_{\mathbf{x}_j}\tilde{p}^{(\tau,\boldsymbol{\xi})}(t,s,\mathbf{x},\mathbf{y}) = -\Big[\big[\tilde{\mathbf{R}}^{(\tau,\boldsymbol{\xi})}(s,t)\big]^*(\tilde{\mathbf{K}}_{s,t}^{(\tau,\boldsymbol{\xi})})^{-1}(\mathbf{m}_{s,t}^{(\tau,\boldsymbol{\xi})}(\mathbf{x})-\mathbf{y})\Big]_j\tilde{p}^{(\tau,\boldsymbol{\xi})}(t,s,\mathbf{x},\mathbf{y}), \tag{2.11}$$

where we denoted here for a vector $\mathbf{z} = (\mathbf{z}_1,\dots,\mathbf{z}_n) \in \mathbb{R}^{nd}$, $[\mathbf{z}]_j = \mathbf{z}_j$ for readibility. The point is now to use scaling arguments. We can first rewrite with the notations of (S):

$$\big[\tilde{\mathbf{R}}^{(\tau,\boldsymbol{\xi})}(s,t)\big]^*(\tilde{\mathbf{K}}_{s,t}^{(\tau,\boldsymbol{\xi})})^{-1} = (s-t)\big[\tilde{\mathbf{R}}^{(\tau,\boldsymbol{\xi})}(s,t)\big]^*\mathbb{T}_{s-t}^{-1}(\widehat{\tilde{\mathbf{K}}_1^{(\tau,\boldsymbol{\xi}),s,t}})^{-1}\mathbb{T}_{s-t}^{-1}. \tag{2.12}$$

Put it differently, $\widehat{\tilde{\mathbf{K}}_1^{(\tau,\boldsymbol{\xi}),s,t}}$ can also be seen as the covariance matrix at time 1 of the rescaled process $\big((s-t)^{\frac{1}{2}}\mathbb{T}_{s-t}^{-1}\tilde{\mathbf{X}}_{t+v(s-t)}^{t,\mathbf{x},(\tau,\boldsymbol{\xi})}\big)_{v\in[0,1]}$. From the good-scaling property of Proposition 4, it was already observed in (S_{ND}) that $\widehat{\tilde{\mathbf{K}}_1^{(\tau,\boldsymbol{\xi}),s,t}}$ is a non-degenerate bounded matrix. From (2.6), a similar scaling argument yields that the resolvent $\mathbf{R}^{(\tau,\xi)}(s,t)$ can also be written as:

$$[\tilde{\mathbf{R}}^{(\tau,\boldsymbol{\xi})}(s,t)]^* = \mathbb{T}_{s-t}^{-1}\Big[\widehat{\tilde{\mathbf{R}}}^{(\tau,\boldsymbol{\xi}),s,t}(1,0)\Big]^*\mathbb{T}_{s-t}, \tag{2.13}$$

where again $\widehat{\tilde{\mathbf{R}}}^{(\tau,\boldsymbol{\xi}),s,t}(1,0)$ is the resolvent at time 1 of the rescaled system associated with (2.6), i.e. for any $v \in [0,1]$, $\mathbb{T}_{s-t}[\tilde{\mathbf{R}}^{(\tau,\boldsymbol{\xi})}(t+v(s-t),t)]^*\mathbb{T}_{s-t}^{-1} = \big[\widehat{\tilde{\mathbf{R}}}^{(\tau,\boldsymbol{\xi}),s,t}(v,0)\big]^*$. From the analysis performed in Lemma 5.1

in [HM16] (see also the proof of Propositions 3.3 and 3.4 in [DM10]) one derives that there exists $\hat{C}_1$ s.t. for any $\boldsymbol{\zeta} \in \mathbb{R}^{nd}$, $|\big[\widehat{\widehat{\mathbf{R}}}^{(\tau,\boldsymbol{\xi}),s,t}(1,0)\big]^* \boldsymbol{\zeta}| \leq \hat{C}_1 |\boldsymbol{\zeta}|$. We refer to Appendix C for a complete proof of this last bound (see equation (C.3)). Equations (2.11), (2.12) and (2.13) therefore yield:

$$\begin{aligned}
& |D_{\mathbf{x}_j} \tilde{p}^{(\tau,\boldsymbol{\xi})}(t,s,\mathbf{x},\mathbf{y})| \\
\leq\ & (s-t)^{-j+\frac{1}{2}} \left| \left(\big[\widehat{\widehat{\mathbf{R}}}^{(\tau,\boldsymbol{\xi}),s,t}(1,0)\big]^* \big(\widehat{\tilde{\mathbf{K}}_1^{s,t}}\big)^{-1} \big((s-t)^{\frac{1}{2}} \mathbb{T}_{s-t}^{-1}(\mathbf{m}_{s,t}^{(\tau,\boldsymbol{\xi})}(\mathbf{x}) - \mathbf{y})\big) \right)_j \right| \tilde{p}^{(\tau,\boldsymbol{\xi})}(t,s,\mathbf{x},\mathbf{y}) \\
\leq\ & C(s-t)^{-j+\frac{1}{2}} (s-t)^{\frac{1}{2}} |\mathbb{T}_{s-t}^{-1}(\mathbf{m}_{s,t}^{(\tau,\boldsymbol{\xi})}(\mathbf{x}) - \mathbf{y})| \tilde{p}^{(\tau,\boldsymbol{\xi})}(t,s,\mathbf{x},\mathbf{y}).
\end{aligned}$$

From the explicit expression (2.9), Proposition 4 and the above equation, we eventually derive:

$$\begin{aligned}
& |D_{\mathbf{x}_j} \tilde{p}^{(\tau,\boldsymbol{\xi})}(t,s,\mathbf{x},\mathbf{y})| \\
\leq\ & \frac{C}{(s-t)^{j-\frac{1}{2}}} \Big((s-t)^{\frac{1}{2}} |\mathbb{T}_{s-t}^{-1}(\mathbf{m}_{s,t}^{(\tau,\boldsymbol{\xi})}(\mathbf{x}) - \mathbf{y})| \Big) \frac{1}{(s-t)^{\frac{n^2 d}{2}}} \exp\Big(-C^{-1}(s-t) |\mathbb{T}_{s-t}^{-1}(\mathbf{m}_{s,t}^{(\tau,\boldsymbol{\xi})}(\mathbf{x}) - \mathbf{y})|^2 \Big) \\
\leq\ & \frac{\bar{C}}{(s-t)^{j-\frac{1}{2}}} \hat{p}_{C^{-1}}^{(\tau,\boldsymbol{\xi})}(t,s,\mathbf{x},\mathbf{y}),
\end{aligned}$$

which gives the statement for one partial derivative. The controls on the higher order derivatives are obtained similarly (see e.g. the proof of Lemma 5.5 of [DM10] for the bounds on $D_{\mathbf{x}_1}^2 \tilde{p}^{(\tau,\boldsymbol{\xi})}(t,s,\mathbf{x},\mathbf{y})$). □

Remark 6. *When the freezing couple $(\tau,\boldsymbol{\xi})$ corresponds to the couple $(t,\mathbf{x})$, where t is the starting time and $\mathbf{x}$ the starting position of $\tilde{\mathbf{X}}^{(\tau,\boldsymbol{\xi})}$, we importantly derive from* (LF) *that* (2.10) *can be specified as follows:*

$$|D_{\mathbf{x}_l}^q D_{\mathbf{x}_1}^r \tilde{p}^{(\tau,\boldsymbol{\xi})}(t,s,\mathbf{x},\mathbf{y})|_{(\tau,\boldsymbol{\xi})=(t,\mathbf{x})} \leq \frac{C}{(s-t)^{\frac{n^2 d}{2}+(l-\frac{1}{2})q+\frac{r}{2}}} \exp\Big(-C^{-1}(s-t) \big| \mathbb{T}_{s-t}^{-1}\big(\boldsymbol{\theta}_{s,t}(\mathbf{x}) - \mathbf{y}\big) \big|^2 \Big). \quad (\mathrm{G}_{NL})$$

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

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

$$\int_{\mathbb{R}^{nd}} d\mathbf{y} D_{\mathbf{x}_k} \tilde{p}^{(\tau,\boldsymbol{\xi})}(t,s,\mathbf{x},\mathbf{y}) \Big\langle z, (\mathbf{y} - \mathbf{m}_{s,t}^{(\tau,\boldsymbol{\xi})}(\mathbf{x}))_k \Big\rangle = z. \quad (2.14)$$

Proof. From Proposition 5, we have $\int_{\mathbb{R}^{nd}} \tilde{p}^{(\tau,\boldsymbol{\xi})}(t,s,\mathbf{x},\mathbf{y})(\mathbf{y} - \mathbf{m}_{s,t}^{(\tau,\boldsymbol{\xi})}(\mathbf{x}))_k d\mathbf{y} = \mathbf{0}_d$. Differentiating this expression w.r.t. $\mathbf{x}_k$ and using the Leibniz formula (recalling as well the identity (2.7) which yields $D_{\mathbf{x}_k}[\mathbf{m}_{s,t}^{(\tau,\boldsymbol{\xi})}(\mathbf{x})]_k = (\tilde{\mathbf{R}}^{(\tau,\boldsymbol{\xi})}(s,t))_{k,k} = \mathbf{I}_{d,d}$) gives (2.14). □

2.1.2 Density and associated inhomogeneous semi-group: regularization properties

For fixed $(t,\mathbf{x}) \in [0,T] \times \mathbb{R}^{nd}$, we give in this section some important properties concerning the regularization effects in time of the density $\tilde{p}^{(\tau,\boldsymbol{\xi})}(t,s,\mathbf{x},\cdot)$, $s \in (t,T]$ and the associated semi-group. From now on, for notational simplicity, we will write with a slight abuse of notation $\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y}) := \tilde{p}^{(t,\boldsymbol{\xi})}(t,s,\mathbf{x},\mathbf{y})$, 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 Section 2.2.

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

$$\int_{\mathbb{R}^{nd}} d\mathbf{y} \Big| D_{\mathbf{x}_l} D_{\mathbf{x}_1}^{\ell} \tilde{p}^{\boldsymbol{\xi}}(s,t,\mathbf{x},\mathbf{y}) \Big|_{\boldsymbol{\xi}=\mathbf{x}} \big| \big(\mathbf{y} - \boldsymbol{\theta}_{s,t}(\mathbf{x})\big)_j \big|^{\gamma_1} \big| \big(\mathbf{y} - \boldsymbol{\theta}_{s,t}(\mathbf{x})\big)_k \big|^{\gamma_2} \leq C(s-t)^{-\frac{\ell}{2} - (l-\frac{1}{2}) + \gamma_1(j-\frac{1}{2}) + \gamma_2(k-\frac{1}{2})}. \quad (2.15)$$

Proof. Equation (2.15) actually follows from (G_{NL}) observing that:

$$\begin{aligned}
&\int_{\mathbb{R}^{nd}} d\mathbf{y}\Big|D_{\mathbf{x}_l} D^{\ell}_{\mathbf{x}_1}\tilde{p}^{\boldsymbol{\xi}}(s,t,\mathbf{x},\mathbf{y})\Big|_{\boldsymbol{\xi}=\mathbf{x}}\big|\big(\mathbf{y}-\boldsymbol{\theta}_{s,t}(\mathbf{x})\big)_j\big|^{\gamma_1}\big|\big(\mathbf{y}-\boldsymbol{\theta}_{s,t}(\mathbf{x})\big)_k\big|^{\gamma_2}\\
\leq\ &\int_{\mathbb{R}^{nd}} d\mathbf{y}\frac{C}{(s-t)^{\frac{n^2d}{2}+(l-\frac{1}{2})+\frac{\ell}{2}}}\exp\Big(-C^{-1}(s-t)\big|\mathbb{T}^{-1}_{s-t}\big(\boldsymbol{\theta}_{s,t}(\mathbf{x})-\mathbf{y}\big)\big|^2\Big)\\
&\times\Big(\frac{|(\boldsymbol{\theta}_{s,t}(\mathbf{x})-\mathbf{y})_j|}{(s-t)^{j-\frac{1}{2}}}\Big)^{\gamma_1}(s-t)^{\gamma_1(j-\frac{1}{2})}\times\Big(\frac{|(\boldsymbol{\theta}_{s,t}(\mathbf{x})-\mathbf{y})_k|}{(s-t)^{k-\frac{1}{2}}}\Big)^{\gamma_2}(s-t)^{\gamma_2(k-\frac{1}{2})},
\end{aligned}\tag{2.16}$$

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:

$$\Big(\frac{|(\boldsymbol{\theta}_{s,t}(\mathbf{x})-\mathbf{y})_j|}{(s-t)^{j-\frac{1}{2}}}\Big)^{\gamma_1}\Big(\frac{|(\boldsymbol{\theta}_{s,t}(\mathbf{x})-\mathbf{y})_k|}{(s-t)^{k-\frac{1}{2}}}\Big)^{\gamma_2}\leq|(s-t)^{\frac{1}{2}}(\mathbb{T}^{-1}_{s-t}(\boldsymbol{\theta}_{s,t}(\mathbf{x})-\mathbf{y}))|^{\gamma_1+\gamma_2},$$

we eventually derive from (2.16) that up to a modification of C, which can be chosen uniformly w.r.t. $\gamma_1,\gamma_2\in(0,1]$:

$$\begin{aligned}
&\int_{\mathbb{R}^{nd}} d\mathbf{y}\Big|D_{\mathbf{x}_l} D^{\ell}_{\mathbf{x}_1}\tilde{p}^{\boldsymbol{\xi}}(s,t,\mathbf{x},\mathbf{y})\Big|_{\boldsymbol{\xi}=\mathbf{x}}\big|\big(\mathbf{y}-\boldsymbol{\theta}_{s,t}(\mathbf{x})\big)_j\big|^{\gamma_1}\Big|_{\boldsymbol{\xi}=\mathbf{x}}\big|\big(\mathbf{y}-\boldsymbol{\theta}_{s,t}(\mathbf{x})\big)_k\big|^{\gamma_2}\\
\leq\ &\int_{\mathbb{R}^{nd}} d\mathbf{y}\frac{C}{(s-t)^{\frac{n^2d}{2}+(l-\frac{1}{2})+\frac{\ell}{2}}}\exp\Big(-C^{-1}(s-t)\big|\mathbb{T}^{-1}_{s-t}\big(\boldsymbol{\theta}_{s,t}(\mathbf{x})-\mathbf{y}\big)\big|^2\Big)(s-t)^{\gamma_1(j-\frac{1}{2})+\gamma_2(k-\frac{1}{2})}\\
\leq\ &C(s-t)^{-\frac{\ell}{2}-(l-\frac{1}{2})+\gamma_1(j-\frac{1}{2})+\gamma_2(k-\frac{1}{2})},
\end{aligned}$$

which precisely gives (2.15). □

Semi-group. Fix $t\in[0,T]$, $\boldsymbol{\xi}\in\mathbb{R}^{nd}$. 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), $\mathbf{x}\in\mathbb{R}^{nd}$:

$$\tilde{P}^{\boldsymbol{\xi}}_{s,t}g(\mathbf{x}):=\int_{\mathbb{R}^{nd}} d\mathbf{y}\tilde{p}^{(t,\boldsymbol{\xi})}(t,s,\mathbf{x},\mathbf{y})g(\mathbf{y}).\tag{2.17}$$

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 ϑ-Hölder continuous functions where $\vartheta:=(\vartheta_1,\dots,\vartheta_n)\in(0,1]^n$ is a multi-index and for $i\in[\![1,n]\!]$, ϑ_i stands for the Hölder regularity of f in the variable $\mathbf{x}_i$.*

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

$$D_{\mathbf{x}_l}\tilde{P}^{\boldsymbol{\xi}}_{s,t}\Big(f(\cdot_{1,k-1},\mathbf{a}_{k:n})\Big)(\mathbf{x})=0.\tag{2.18}$$

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

$$\begin{aligned}
|D_{\mathbf{x}_1}D_{\mathbf{x}_l}\tilde{P}^{\boldsymbol{\xi}}_{s,t}f(\mathbf{x})| &\leq C\sum_{j=l}^{n}[f_j(s,\cdot)]_{\vartheta_j}(s-t)^{-l+\vartheta_j(j-\frac{1}{2})},\\
|D_{\mathbf{x}_l}\tilde{P}^{\boldsymbol{\xi}}_{s,t}f(\mathbf{x})| &\leq C\sum_{j=l}^{n}[f_j(s,\cdot)]_{\vartheta_j}(s-t)^{-(l-\frac{1}{2})+\vartheta_j(j-\frac{1}{2})},
\end{aligned}\tag{2.19}$$

with the notations introduced in (1.2)*.*

Proof. Centering arguments like (2.18) will be a crucial tool in the analysis below. To justify such an identity, write:

$$\begin{aligned}
\tilde{P}^{\boldsymbol{\xi}}_{s,t}\Big(f(\cdot_{1:k-1},\mathbf{a}_{k:n})\Big)(\mathbf{x}) &= \int_{\mathbb{R}^{nd}} d\mathbf{y}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y})f(\mathbf{y}_{1,k-1},\mathbf{a}_{k:n})\\
&= \int_{\mathbb{R}^{nd}} d\mathbf{y}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y}+\mathbf{m}^{(t,\boldsymbol{\xi})}_{s,t}(\mathbf{x}))f\big(\mathbf{y}_{1,k-1}+(\mathbf{m}^{(t,\boldsymbol{\xi})}_{s,t}(\mathbf{x}))_{1:k-1},\mathbf{a}_{k:n}\big).
\end{aligned}$$

Note now from (2.9) that $\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y}+\mathbf{m}_{s,t}^{(t,\boldsymbol{\xi})}(\mathbf{x}))$ does not depend on $\mathbf{x}$. Observe as well that $(\mathbf{m}_{s,t}^{(t,\boldsymbol{\xi})}(\mathbf{x}))_{1:k-1}$ does not depend on $\mathbf{x}_l$. Thus, the right-hand side of the above equality does not depend on $\mathbf{x}_l$, which implies (2.18).

Let us now prove (2.19). The idea is to use first a centering argument w.r.t. the variables l to n. Namely, from (2.18) and Proposition 5, it holds that for any $\mathbf{a} \in \mathbb{R}^{nd}$,

$$\begin{aligned}\left|D_{\mathbf{x}_1} D_{\mathbf{x}_l}\left[\tilde{P}_{s,t}^{\boldsymbol{\xi}} f(s,\cdot)\right](\mathbf{x})\right| &= \left|D_{\mathbf{x}_1} D_{\mathbf{x}_l}\left[\tilde{P}_{s,t}^{\boldsymbol{\xi}}\Big(f(s,\cdot) - f(s,\cdot_{1:l-1},\mathbf{a}_{l:n})\Big)\right](\mathbf{x})\right| \\ &\le \bar{C}(s-t)^{-(l-1/2)-1/2}\int_{\mathbb{R}^{nd}} d\mathbf{y}\hat{p}_{C^{-1}}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y})\Big|f(s,\mathbf{y}) - f(s,\mathbf{y}_{1:l-1},\mathbf{a}_{l:n})\Big| \\ &\le \bar{C}(s-t)^{-(l-1/2)-1/2}\sum_{j=l}^{n}[f_j(s,\cdot))]_{\vartheta_j}\int_{\mathbb{R}^{nd}} d\mathbf{y}\hat{p}_{C^{-1}}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y})\Big|\mathbf{y}_j - \mathbf{a}_j\Big|^{\vartheta_j}.\end{aligned}$$

Taking now $\mathbf{a} = \mathbf{m}_{s,t}^{(t,\boldsymbol{\xi})}(\mathbf{x})$, we derive from (2.10) that:

$$\begin{aligned}\left|D_{\mathbf{x}_1} D_{\mathbf{x}_l}\left[\tilde{P}_{s,t}^{\boldsymbol{\xi}} f(s,\cdot)\right](\mathbf{x})\right| &\le \bar{C}(s-t)^{-(l-1/2)-1/2}\sum_{j=l}^{n}[f_j(s,\cdot))]_{\vartheta_j}\int_{\mathbb{R}^{nd}} d\mathbf{y}\hat{p}_{C^{-1}}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y})\Big|(\mathbf{y}-\mathbf{m}_{s,t}^{(t,\boldsymbol{\xi})}(\mathbf{x}))_j\Big|^{\vartheta_j} \\ &\le C\sum_{j=l}^{n}[f_j(s,\cdot)]_{\vartheta_j}(s-t)^{-l+\vartheta_j(j-\frac{1}{2})},\end{aligned}$$

which gives the first bound in (2.19). The control for $|D_{\mathbf{x}_l}\tilde{P}_{s,t}^{\boldsymbol{\xi}} f(\mathbf{x})|$ is derived similarly. □

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 $(\mathbf{T}_\beta)$ and $(\mathbf{H}_\eta)$.

Lemma 9. *From the smoothness assumption on the drift coefficient in* $(\mathbf{T}_\beta)$*, there exists* $C := C((\mathbf{A}))$ *s.t. for all* $\ell \in [\![2,n]\!]$*,* $k \in [\![\ell, n+1]\!]$*, and for any* $(s,\mathbf{x},\mathbf{a}) \in [0,T]\times(\mathbb{R}^{nd})^2$*:*

$$\begin{aligned}&\left|\Big(\mathbf{F}_\ell(s,\mathbf{y}_{\ell-1:k-1},\mathbf{a}_{k:n}) - \mathbf{F}_\ell(s,\mathbf{a}_{\ell-1:n}) - D_{\ell-1}\mathbf{F}_\ell(s,\mathbf{a}_{\ell-1:n})\big(\mathbf{y}-\mathbf{a}\big)_{\ell-1}\Big)\right| \\ &\le C\Big\{\sum_{j=\ell}^{k-1}\big\{[(\mathbf{F}_\ell)_j(s,\cdot)]_{\beta_j}|(\mathbf{y}-\mathbf{a})_j|^{\beta_j}\big\} + [(D_{\ell-1}\mathbf{F}_\ell)_{\ell-1}(s,\cdot))]_\eta|(\mathbf{y}-\mathbf{a})_{\ell-1}|^{1+\eta}\Big\},\end{aligned}$$

with the convention that for $k = n+1$*,* $\mathbf{F}_\ell(s,\mathbf{y}_{\ell-1:k-1},\mathbf{a}_{k:n}) = \mathbf{F}_\ell(s,\mathbf{y}_{\ell-1:n})$*.*

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 Section 1.5 (see the connection between equations (1.19) and (1.33)). The *proxy* used here is the Gaussian process introduced in Section 2.1 for a suitable freezing parameter $\boldsymbol{\xi}$ to be specified later on and whose generator is given by, for all $\varphi \in C_0^2(\mathbb{R}^{nd},\mathbb{R})$ and $(t,\mathbf{x}) \in [0,T]\times\mathbb{R}^{nd}$,

$$\begin{aligned}\tilde{L}_t^{\boldsymbol{\xi}}\varphi(\mathbf{x}) &:= \langle \mathbf{F}(t,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})) + D\mathbf{F}(t,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))(\mathbf{x}-\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})), D_{\mathbf{x}}\varphi(\mathbf{x})\rangle \\ &\quad + \frac{1}{2}\mathrm{tr}(\sigma\sigma^*(t,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))D_{\mathbf{x}_1}^2\varphi(\mathbf{x})).\end{aligned}$$

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

$$\begin{aligned}&u(t,\mathbf{x}) \\ &= \int_t^T ds\left[\tilde{P}_{s,t}^{\boldsymbol{\xi}} f(s,\cdot)\right](\mathbf{x}) + \int_t^T ds\left[\tilde{P}_{s,t}^{\boldsymbol{\xi}}\Big((L_s - \tilde{L}_s^{\boldsymbol{\xi}})u\Big)(s,\cdot)\right](\mathbf{x}) \\ &= \int_t^T ds\left[\tilde{P}_{s,t}^{\boldsymbol{\xi}} f(s,\cdot)\right](\mathbf{x}) + \int_t^T ds\left[\tilde{P}_{s,t}^{\boldsymbol{\xi}}\Big(\big\langle \mathbf{F}_1(s,\cdot) - \mathbf{F}_1(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})), D_1 u(s,\cdot)\big\rangle\Big)\right](\mathbf{x}) \\ &\quad + \int_t^T ds\left[\tilde{P}_{s,t}^{\boldsymbol{\xi}}\left(\frac{1}{2}\mathrm{Tr}\left[a(s,\cdot) - a(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))D_1^2 u(s,\cdot)\right]\right)\right](\mathbf{x}) \\ &\quad + \int_t^T ds\left[\tilde{P}_{s,t}^{\boldsymbol{\xi}}\left(\sum_{i=2}^{n}\big\langle(\mathbf{F}_i(s,\cdot) - \mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})) - D_{i-1}\mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))(\cdot - \boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))_{i-1}), D_i u(s,\cdot)\big\rangle\right)\right](\mathbf{x}),\end{aligned} \tag{2.20}$$

for any $\boldsymbol{\xi}$ in $\mathbb{R}^{nd}$, according to the notations introduced in Remark 2 for the entries $(\mathbf{F}_i)_{i\in[\![1,n]\!]}$ of $\mathbf{F}$.

To establish (2.2) we need to differentiate the above expression w.r.t. $(\mathbf{x}_l)_{l\in[\![1,n]\!]}$ and then w.r.t. $\mathbf{x}_1$. Differentiating first this expression w.r.t. $\mathbf{x}_l, l\in[\![1,n]\!]$ we obtain:

$$
\begin{aligned}
& D_{\mathbf{x}_l}u(t,\mathbf{x})\\
= \;& \int_t^T dsD_{\mathbf{x}_l}\left[\tilde{P}^{\boldsymbol{\xi}}_{s,t}f(s,\cdot)\right](\mathbf{x})+\int_t^T dsD_{\mathbf{x}_l}\left[\tilde{P}^{\boldsymbol{\xi}}_{s,t}\Big((L_s-\tilde{L}^{\boldsymbol{\xi}}_s)u\Big)(s,\cdot)\right](\mathbf{x})\\
= \;& \left\{\int_t^T dsD_{\mathbf{x}_l}\left[\tilde{P}^{\boldsymbol{\xi}}_{s,t}f(s,\cdot)\right](\mathbf{x})+\int_t^T dsD_{\mathbf{x}_l}\left[\tilde{P}^{\boldsymbol{\xi}}_{s,t}\left(\langle(\mathbf{F}_1(s,\cdot)-\mathbf{F}_1(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))),D_1u(s,\cdot)\rangle\right)\right](\mathbf{x})\right.\\
& \left.+\int_t^T dsD_{\mathbf{x}_l}\left[\tilde{P}^{\boldsymbol{\xi}}_{s,t}\left(\frac{1}{2}\mathrm{Tr}\left[\Big(a(s,\cdot)-a(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))\Big)D_1^2u(s,\cdot)\right]\right)\right](\mathbf{x})\right\}\\
& +\left\{\int_t^T dsD_{\mathbf{x}_l}\left[\tilde{P}^{\boldsymbol{\xi}}_{s,t}\Big(\sum_{i=2}^n\big\langle(\mathbf{F}_i(s,\cdot)-\mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))-D_{i-1}\mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))(\cdot-\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))_{i-1}),D_iu(s,\cdot)\big\rangle\Big)\right](\mathbf{x})\right\}\\
=: \;& \left\{\int_t^T ds\sum_{k=1}^3 H^{\boldsymbol{\xi}}_{l,k}(s,\mathbf{x})\right\}+\left\{\int_t^T dsI^{\boldsymbol{\xi}}_l(s,\mathbf{x})\right\}=:\int_t^T dsH^{\boldsymbol{\xi}}_l(s,\mathbf{x})+\int_t^T dsI^{\boldsymbol{\xi}}_l(s,\mathbf{x}). \qquad (2.21)
\end{aligned}
$$

The term $H^{\boldsymbol{\xi}}_l(s,\mathbf{x})$ gathers the sensitivities w.r.t. $\mathbf{x}_l$ of the source term and the non-degenerate part of the difference of the operators, whereas $I^{\boldsymbol{\xi}}_l(s,\mathbf{x})$ precisely gathers the sensitivities w.r.t. $\mathbf{x}_l$ of the degenerate part of the difference of the operators. We will now start from the representation (2.21) which we will again differentiate w.r.t. the non-degenerate variable $\mathbf{x}_1$ 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 Section 1.5, when differentiating the kernel associated with the frozen semi-group defined by (2.17) we generate an *a priori* not 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 $D_{\mathbf{x}_l}\big(D_{\mathbf{x}_1}u(t,\mathbf{x})\big)$ 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 (2.21). We then only concentrate on this term and omit the proof of the statement concerning the boundedness of the gradient $D_{\mathbf{x}_l}u(t,\mathbf{x})$ 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 $D_{\mathbf{x}_1}H^{\boldsymbol{\xi}}_l(s,\mathbf{x})$) and then the degenerate part (*i.e.* the estimate for $D_{\mathbf{x}_1}I^{\boldsymbol{\xi}}_l(s,\mathbf{x})$) which is a bit more involved.

Source term and non-degenerate part of the operator: estimates for $(D_{\mathbf{x}_1}H^{\boldsymbol{\xi}}_l(s,\mathbf{x}))_{\boldsymbol{\xi}=\mathbf{x}}$. We first focus on the source term and the derivatives w.r.t. the non-degenerate variable $\mathbf{x}_1$, three first terms in the r.h.s. denoted by $H^{\boldsymbol{\xi}}_{l,1}(s,\mathbf{x}), H^{\boldsymbol{\xi}}_{l,2}(s,\mathbf{x}), H^{\boldsymbol{\xi}}_{l,3}(s,\mathbf{x})$ in (2.21), taking $\boldsymbol{\xi}=\mathbf{x}$ after the additional differentiation.

For each $l\in[\![1,n]\!]$, one readily derives from (2.19) in Lemma 8 that for the source term[7]:

$$
\Big|D_{\mathbf{x}_1}H^{\boldsymbol{\xi}}_{l,1}(s,\mathbf{x})\Big|_{\boldsymbol{\xi}=\mathbf{x}}=\Big|D_{\mathbf{x}_1}D_{\mathbf{x}_l}\left[\tilde{P}^{\boldsymbol{\xi}}_{s,t}f(s,\cdot)\right](\mathbf{x})\Big|\Big|_{\boldsymbol{\xi}=\mathbf{x}}\le C\sum_{j=l}^n[f_j(s,\cdot)]_{\beta_j}(s-t)^{-l+\beta_j(j-\frac{1}{2})}. \qquad (2.22)
$$

Those terms are integrable in time as soon as

$$
\beta_j(j-\frac{1}{2})-l>-1,\quad j\in[\![l,n]\!]\Leftarrow\beta_j\in\left(\frac{2j-2}{2j-1},1\right]. \qquad (2.23)
$$

Similarly, from (2.20), (2.21), for the drift associated with the non-degenerate part, we first rewrite from

[7] Observe that for this contribution, from (2.19) the bound would hold for any freezing parameter $\boldsymbol{\xi}$. We choose here to take $\boldsymbol{\xi}=\mathbf{x}$ for the compatibility with the other terms $(D_{\mathbf{x}_1}H^{\boldsymbol{\xi}}_{l,k}(s,\mathbf{x}))_{k\in\{2,3\}}$ for which this specific choice is indeed needed.

the centering properties of Lemma 8:

$$\begin{aligned}
&\left|D_{\mathbf{x}_1} H^{\boldsymbol{\xi}}_{l,2}(s,\mathbf{x})\right|_{\boldsymbol{\xi}=\mathbf{x}} \\
&= \left|D_{\mathbf{x}_1} D_{\mathbf{x}_l}\left[\tilde{P}^{\boldsymbol{\xi}}_{s,t}\left(\left\langle \mathbf{F}_1(s,\cdot) - \mathbf{F}_1(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})), D_1 u(s,\cdot)\right\rangle\right)\right](\mathbf{x})\right|\Big|_{\boldsymbol{\xi}=\mathbf{x}} \\
&\leq \left|D_{\mathbf{x}_1} D_{\mathbf{x}_l}\left[\tilde{P}^{\boldsymbol{\xi}}_{s,t}\left(\left\langle \mathbf{F}_1(s,\cdot) - \mathbf{F}_1(s,\cdot_{1:l-1},\boldsymbol{\theta}^{l:n}_{s,t}(\boldsymbol{\xi})), D_1 u(s,\cdot)\right\rangle\right)\right](\mathbf{x})\right|\Big|_{\boldsymbol{\xi}=\mathbf{x}} \\
&\quad + \left|D_{\mathbf{x}_1} D_{\mathbf{x}_l}\left[\tilde{P}^{\boldsymbol{\xi}}_{s,t}\left(\left\langle \mathbf{F}_1(s,\cdot_{1:l-1},\boldsymbol{\theta}^{l:n}_{s,t}(\boldsymbol{\xi})) - \mathbf{F}_1(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))\big(D_1 u(s,\cdot) - D_1 u(s,\cdot_{1:l-1},\boldsymbol{\theta}^{l:n}_{s,t}(\boldsymbol{\xi}))\big)\right\rangle\right)\right](\mathbf{x})\right|\Big|_{\boldsymbol{\xi}=\mathbf{x}}.
\end{aligned}$$

Expanding the semi-group yields

$$\begin{aligned}
&\left|D_{\mathbf{x}_1} H^{\boldsymbol{\xi}}_{l,2}(s,\mathbf{x})\right|_{\boldsymbol{\xi}=\mathbf{x}} \\
&\leq \int_{\mathbb{R}^{nd}} d\mathbf{y} |D_{\mathbf{x}_1} D_{\mathbf{x}_l} \tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y})|_{\boldsymbol{\xi}=\mathbf{x}} \left|\mathbf{F}_1(s,\mathbf{y}) - \mathbf{F}_1(s,\mathbf{y}_{1:l-1},\boldsymbol{\theta}^{l:n}_{s,t}(\mathbf{x}))\right| |D_1 u(s,\mathbf{y})| d\mathbf{y} \\
&\quad + \int_{\mathbb{R}^{nd}} d\mathbf{y} |D_{\mathbf{x}_1} D_{\mathbf{x}_l} \tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y})|_{\boldsymbol{\xi}=\mathbf{x}} \left|\mathbf{F}_1(s,\mathbf{y}_{1:l-1},\boldsymbol{\theta}^{l:n}_{s,t}(\mathbf{x})) - \mathbf{F}_1(s,\boldsymbol{\theta}_{s,t}(\mathbf{x}))\right| \left|D_1 u(s,\mathbf{y}) - D_1 u(s,\mathbf{y}_{1:l-1},\boldsymbol{\theta}^{l:n}_{s,t}(\mathbf{x}))\right|.
\end{aligned}$$

Hence,

$$\begin{aligned}
&\left|D_{\mathbf{x}_1} H^{\boldsymbol{\xi}}_{l,2}(s,\mathbf{x})\right|_{\boldsymbol{\xi}=\mathbf{x}} \\
&\leq \Bigg(\sum_{j=l}^{n} \|D_1 u\|_\infty [(\mathbf{F}_1)_j(s,\cdot)]_{\beta_j} \int_{\mathbb{R}^{nd}} d\mathbf{y} |D_{\mathbf{x}_l} D_{\mathbf{x}_1} \tilde{p}^{\boldsymbol{\xi}}(s,t,\mathbf{x},\mathbf{y})|_{\boldsymbol{\xi}=\mathbf{x}} |(\boldsymbol{\theta}_{s,t}(\mathbf{x}) - \mathbf{y})_j|^{\beta_j} \\
&\quad + \sum_{j=l}^{n} \sum_{k=1}^{l-1} \|D_1 D_j u\|_\infty [(\mathbf{F}_1)_k(s,\cdot)]_{\beta_k} \int_{\mathbb{R}^{nd}} d\mathbf{y} |D_{\mathbf{x}_l} D_{\mathbf{x}_1} \tilde{p}^{\boldsymbol{\xi}}(s,t,\mathbf{x},\mathbf{y})|_{\boldsymbol{\xi}=\mathbf{x}} |(\boldsymbol{\theta}_{s,t}(\mathbf{x}) - \mathbf{y})_k|^{\beta_k} |(\boldsymbol{\theta}_{s,t}(\mathbf{x}) - \mathbf{y})_j| \Bigg).
\end{aligned}$$

We then conclude from equation (2.15) 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

$$\begin{aligned}
&\left|D_{\mathbf{x}_1} H^{\boldsymbol{\xi}}_{l,2}(s,\mathbf{x})\right|_{\boldsymbol{\xi}=\mathbf{x}} \qquad (2.24) \\
&\leq C\Bigg(\sum_{j=l}^{n} \|D_1 u\|_\infty [(\mathbf{F}_1)_j(s,\cdot)]_{\beta_j} (s-t)^{-l+\beta_j(j-\frac{1}{2})} + \sum_{j=l}^{n} \sum_{k=1}^{l-1} \|D_1 D_j u\|_\infty [(\mathbf{F}_1)_k(s,\cdot)]_{\beta_k} (s-t)^{-l+\beta_k(k-\frac{1}{2})+(j-\frac{1}{2})} \Bigg),
\end{aligned}$$

leading precisely to the same integrability thresholds of equation (2.23) and assumption $(\mathbf{T}_\beta)$ (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:

$$\begin{aligned}
\left|D_{\mathbf{x}_1} H^{\boldsymbol{\xi}}_{l,3}(s,\mathbf{x})\right|_{\boldsymbol{\xi}=\mathbf{x}} &= \left|D_{\mathbf{x}_1} D_{\mathbf{x}_l}\left[\tilde{P}^{\boldsymbol{\xi}}_{s,t}\left(\frac{1}{2}\mathrm{Tr}\left[\big(a(s,\cdot) - a(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))\big) D_1^2 u(s,\cdot)\right]\right)\right](\mathbf{x})\right|\Big|_{\boldsymbol{\xi}=\mathbf{x}} \\
&\leq \left|D_{\mathbf{x}_1} D_{\mathbf{x}_l}\left[\tilde{P}^{\boldsymbol{\xi}}_{s,t}\left(\frac{1}{2}\mathrm{Tr}\left[\big(a(s,\cdot) - a(s,\cdot_{1:l-1},\boldsymbol{\theta}^{l:n}_{s,t}(\boldsymbol{\xi}))\big) D_1^2 u(s,\cdot)\right]\right)\right](\mathbf{x})\right|\Big|_{\boldsymbol{\xi}=\mathbf{x}} \\
&\quad + \left|D_{\mathbf{x}_1} D_{\mathbf{x}_l}\left[\tilde{P}^{\boldsymbol{\xi}}_{s,t}\left(\frac{1}{2}\mathrm{Tr}\left[\big(a(s,\cdot_{1:l-1},\boldsymbol{\theta}^{l:n}_{s,t}(\boldsymbol{\xi})) - a(s,\boldsymbol{\theta}_{s,t}(\xi))\big) D_1^2 u(s,\cdot)\right]\right)\right](\mathbf{x})\right|\Big|_{\boldsymbol{\xi}=\mathbf{x}} \\
&=: \left|D_{\mathbf{x}_1} H^{\boldsymbol{\xi}}_{l,31}(s,\mathbf{x})\right|_{\boldsymbol{\xi}=\mathbf{x}} + \left|D_{\mathbf{x}_1} H^{\boldsymbol{\xi}}_{l,32}(s,\mathbf{x})\right|_{\boldsymbol{\xi}=\mathbf{x}}. \qquad (2.25)
\end{aligned}$$

The term $|D_{\mathbf{x}_1}H^{\boldsymbol{\xi}}_{l,31}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}}$ is already centered at the *appropriate* scales, i.e. from variables l to n. Recalling that a is Lipschitz continuous, write:

$$\begin{aligned}
|D_{\mathbf{x}_1}H^{\boldsymbol{\xi}}_{l,31}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}} &= \left|\int_{\mathbb{R}^d} d\mathbf{y} D_{\mathbf{x}_1}D_{\mathbf{x}_l}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y})\frac{1}{2}\mathrm{Tr}\left[\big(a(s,\mathbf{y})-a(s,\mathbf{y}_{1:l-1},\boldsymbol{\theta}^{l:n}_{s,t}(\boldsymbol{\xi}))\big)D_1^2u(s,\mathbf{y})\right]\right|_{\boldsymbol{\xi}=\mathbf{x}}\\
&\leq \frac{1}{2}\int_{\mathbb{R}^d} d\mathbf{y}|D_{\mathbf{x}_1}D_{\mathbf{x}_l}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y})|_{\boldsymbol{\xi}=\mathbf{x}}\ \big|\big(a(s,\mathbf{y})-a(s,\mathbf{y}_{1:l-1},\boldsymbol{\theta}^{l:n}_{s,t}(\mathbf{x}))\big|\ |D_1^2u(s,\mathbf{y})|\\
&\leq \frac{1}{2}[a(s,\cdot)]_1\|D_1^2u\|_\infty\int_{\mathbb{R}^d} d\mathbf{y}|D_{\mathbf{x}_1}D_{\mathbf{x}_l}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y})|_{\boldsymbol{\xi}=\mathbf{x}}\ |\mathbf{y}_{l:n}-\boldsymbol{\theta}^{l:n}_{s,t}(\mathbf{x})|.
\end{aligned}$$

We can apply (2.15) of Lemma 7 with $\gamma_1=1,\gamma_2=0$ to any of the entries $|\mathbf{y}_j-\boldsymbol{\theta}^j_{s,t}(\mathbf{x})|,\ j\in[\![l,n]\!]$ to obtain:

$$|D_{\mathbf{x}_1}H^{\boldsymbol{\xi}}_{l,31}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}}\leq C\|D_1^2u\|_\infty[a(s,\cdot)]_1\sum_{j=l}^n(s-t)^{-l+(j-\frac{1}{2})}\leq C\|D_1^2u\|_\infty[a(s,\cdot)]_1(s-t)^{-\frac{1}{2}},\tag{2.26}$$

which does not give a *critical* contribution w.r.t. the previously exhibited thresholds in (2.23) and $(\mathbf{T}_\beta)$. For the contribution $|D_{\mathbf{x}_1}H^{\boldsymbol{\xi}}_{l,32}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}}$ we use, in the same spirit as for $|D_{\mathbf{x}_1}H^{\boldsymbol{\xi}}_{l,2}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}}$, a centering argument and an integration by parts to obtain:

$$\begin{aligned}
&|D_{\mathbf{x}_1}H^{\boldsymbol{\xi}}_{l,32}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}}\\
=\ &\left|D_{\mathbf{x}_1}D_{\mathbf{x}_l}\Big[\tilde{P}^{\boldsymbol{\xi}}_{s,t}\Big(\frac{1}{2}\mathrm{Tr}\big[\big(a(s,\cdot_{1:l-1},\boldsymbol{\theta}^{l:n}_{s,t}(\boldsymbol{\xi}))-a(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))\big)(D_1^2u(s,\cdot)-D_1^2u(s,\cdot_{1:l-1},\boldsymbol{\theta}^{l:n}_{s,t}(\boldsymbol{\xi})))\big]\Big)\Big](\mathbf{x})\right|_{\boldsymbol{\xi}=\mathbf{x}}\\
=\ &\left|\frac{1}{2}D_{\mathbf{x}_1}D_{\mathbf{x}_l}\left(\sum_{j=1}^d\int_{\mathbb{R}^{nd}}d\mathbf{y}\Big\langle\partial_{\mathbf{y}_1^j}\Big(\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y})\big(a_{j\cdot}(s,\mathbf{y}_{1:l-1},\boldsymbol{\theta}^{l:n}_{s,t}(\boldsymbol{\xi}))-a_{j\cdot}(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))\big)\Big),\right.\right.\\
&\left.\left.\Big(D_1u(s,\mathbf{y})-D_1u(s,\mathbf{y}_{1:l-1},\boldsymbol{\theta}^{l:n}_{s,t}(\boldsymbol{\xi}))\Big)\Big\rangle\right)\right|_{\boldsymbol{\xi}=\mathbf{x}},
\end{aligned}$$

where $\mathbf{y}_1=(\mathbf{y}_1^1,\ldots,\mathbf{y}_1^d)$, $\partial_{\mathbf{y}_1^j}$ denotes the derivative w.r.t. the j^{th} scalar entry of the non-degenerate variable $\mathbf{y}_1$ and $a_{j\cdot}$ denotes the j^{th} row of the diffusion matrix a. We therefore derive from the proof of Lemma 7 and the smoothness of a (by the Rademacher theorem, a is differentiable almost everywhere):

$$\begin{aligned}
&|D_{\mathbf{x}_1}H^{\boldsymbol{\xi}}_{l,32}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}}\\
\leq\ &C\|\mathbf{D}D_1u(s,\cdot)\|_\infty\int_{\mathbb{R}^{nd}}\frac{d\mathbf{y}}{(s-t)^l}\hat{p}^{\mathbf{x}}_{C^{-1}}(t,s,\mathbf{x},\mathbf{y})\Big(\frac{|(\mathbf{y}-\boldsymbol{\theta}_{s,t}(\mathbf{x}))_{1:l-1}|}{(s-t)^{\frac{1}{2}}}+1\Big)[a(s,\cdot)]_1|(\mathbf{y}-\boldsymbol{\theta}_{s,t}(\mathbf{x}))_{l:n}|\\
\leq\ &C\|\mathbf{D}D_1u(s,\cdot)\|_\infty(s-t)^{-\frac{1}{2}}.
\end{aligned}\tag{2.27}$$

With the notations of (2.21), plugging (2.26), (2.27) into (2.25) and together with (2.24), (2.22), we eventually derive that there exists $\delta:=\delta((\mathbf{A}))>0$:

$$\left|\int_t^T dsD_{\mathbf{x}_1}H^{\boldsymbol{\xi}}_l(s,\mathbf{x})\right|_{\boldsymbol{\xi}=\mathbf{x}}\leq CT^\delta(\|D_1u\|_\infty+\|\mathbf{D}D_1u\|_\infty).\tag{2.28}$$

Degenerate part of the operator: estimates for $(D_{\mathbf{x}_1}I^{\boldsymbol{\xi}}_l(s,\mathbf{x}))_{\boldsymbol{\xi}=\mathbf{x}}$. These are the most delicate terms to handle. Restarting from (2.21), we first write for any $l\in[\![2,n]\!]$:

$$\begin{aligned}
&|D_{\mathbf{x}_1}I^{\boldsymbol{\xi}}_l(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}}\\
=&\left|D_{\mathbf{x}_1}D_{\mathbf{x}_l}\left[\tilde{P}^{\boldsymbol{\xi}}_{s,t}\left(\sum_{i=2}^n\big\langle(\mathbf{F}_i(s,\cdot)-\mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))-D_{i-1}\mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))(\cdot-\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))_{i-1}),D_iu(s,\cdot)\big\rangle\right)\right](\mathbf{x})\right|_{\boldsymbol{\xi}=\mathbf{x}},
\end{aligned}$$

which readily yields that

$$
\begin{aligned}
&|D_{\mathbf{x}_1} I_l^{\boldsymbol{\xi}}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}} \\
&\le \Big| D_{\mathbf{x}_1} D_{\mathbf{x}_l} \Big[\tilde{P}_{s,t}^{\boldsymbol{\xi}} \Big(\sum_{i=l+1}^{n} \big\langle (\mathbf{F}_i(s,\cdot) - \mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})) - D_{i-1}\mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))(\cdot - \boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))_{i-1}), D_i u(s,\cdot) \big\rangle \Big) \Big](\mathbf{x}) \Big|_{\boldsymbol{\xi}=\mathbf{x}} \\
&\quad + \Big| D_{\mathbf{x}_1} D_{\mathbf{x}_l} \Big[\tilde{P}_{s,t}^{\boldsymbol{\xi}} \Big(\sum_{i=2}^{l} \big\langle \mathbf{F}_i(s,\cdot) - \mathbf{F}_i(s,\cdot_{1:l-1}, \boldsymbol{\theta}_{s,t}^{l:n}(\boldsymbol{\xi})), D_i u(s,\cdot) \big\rangle \Big) \Big](\mathbf{x}) \Big|_{\boldsymbol{\xi}=\mathbf{x}} \\
&\quad + \Big| D_{\mathbf{x}_1} D_{\mathbf{x}_l} \Big[\tilde{P}_{s,t}^{\boldsymbol{\xi}} \Big(\sum_{i=2}^{l} \big\langle (\mathbf{F}_i(s,\cdot_{1:l-1}, \boldsymbol{\theta}_{s,t}^{l:n}(\boldsymbol{\xi})) - \mathbf{F}_i(\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})) \\
&\qquad - D_{i-1}\mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))(\cdot - \boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))_{i-1}), D_i u(s,\cdot) \big\rangle \Big) \Big](\mathbf{x}) \Big|_{\boldsymbol{\xi}=\mathbf{x}} \\
&=: |D_{\mathbf{x}_1} I_{l,1}^{\boldsymbol{\xi}}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}} + |D_{\mathbf{x}_1} I_{l,2}^{\boldsymbol{\xi}}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}} + |D_{\mathbf{x}_1} I_{l,3}^{\boldsymbol{\xi}}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}}. \qquad (2.29)
\end{aligned}
$$

• *Control of* $|D_{\mathbf{x}_1} I_{l,1}^{\boldsymbol{\xi}}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}}$ *and* $|D_{\mathbf{x}_1} I_{l,2}^{\boldsymbol{\xi}}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}}$. We emphasize that the integrands $|D_{\mathbf{x}_1} I_{l,1}^{\boldsymbol{\xi}}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}}$ and $|D_{\mathbf{x}_1} I_{l,2}^{\boldsymbol{\xi}}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}}$ are already designed to smoothen the time singularities generated by the cross differentiation of the inhomogeneous semi-group w.r.t. the variables $\mathbf{x}_l$ and $\mathbf{x}_1$. Indeed, write:

$$
\begin{aligned}
&|D_{\mathbf{x}_1} I_{l,1}^{\boldsymbol{\xi}}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}} \\
&\le \int_{\mathbb{R}^{nd}} d\mathbf{y} |D_{\mathbf{x}_1} D_{\mathbf{x}_l} \tilde{p}^{\boldsymbol{\xi}}(s,t,\mathbf{x},\mathbf{y})|_{\boldsymbol{\xi}=\mathbf{x}} \\
&\quad \times \sum_{i=l+1}^{n} \Big| \big\langle (\mathbf{F}_i(s,\mathbf{y}) - \mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\mathbf{x})) - D_{i-1}\mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\mathbf{x}))(\mathbf{y} - \boldsymbol{\theta}_{s,t}(\mathbf{x}))_{i-1}), D_i u(s,\mathbf{y}) \big\rangle \Big| \\
&\le \|D_{l+1:n} u\|_\infty \int_{\mathbb{R}^{nd}} d\mathbf{y} |D_{\mathbf{x}_1} D_{\mathbf{x}_l} \tilde{p}^{\boldsymbol{\xi}}(s,t,\mathbf{x},\mathbf{y})|_{\boldsymbol{\xi}=\mathbf{x}} \sum_{i=l+1}^{n} \Big| \mathbf{F}_i(s,\mathbf{y}) - \mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\mathbf{x})) - D_{i-1}\mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\mathbf{x}))(\mathbf{y} - \boldsymbol{\theta}_{s,t}(\mathbf{x}))_{i-1} \Big|,
\end{aligned}
$$

where $D_{l+1:n} u := \big(D_{l+1} u, \ldots, D_n u\big)$. Applying now Lemma 9 with $\mathbf{a} = \boldsymbol{\theta}_{s,t}(\mathbf{x})$, $\ell = i$, $k = n+1$ yields:

$$
\begin{aligned}
&\sum_{i=l+1}^{n} \Big| \mathbf{F}_i(s,\mathbf{y}) - \mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\mathbf{x})) - D_{i-1}\mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\mathbf{x}))(\mathbf{y} - \boldsymbol{\theta}_{s,t}(\mathbf{x}))_{i-1}) \Big| \\
&\le \; C \sum_{j=l+1}^{n} \Big([(\mathbf{F}_{l+1:j}(s,\cdot))_j]_{\beta_j} |(\boldsymbol{\theta}_{s,t}(\mathbf{x}) - \mathbf{y})_j|^{\beta_j} + [(D_{j-1}\mathbf{F}_j(s,\cdot))_{j-1}]_\eta |(\boldsymbol{\theta}_{s,t}(\mathbf{x}) - \mathbf{y})_{j-1}|^{1+\eta} \Big),
\end{aligned}
$$

using the notation $[(\mathbf{F}_{l+1:j})_j(s,\cdot)]_{\beta_j} := \max_{i \in [\![l+1,j]\!]} [(\mathbf{F}_i)_j(s,\cdot)]_{\beta_j}$ for the last inequality. We finally derive from Lemma 7:

$$
\begin{aligned}
&|D_{\mathbf{x}_1} I_{l,1}^{\boldsymbol{\xi}}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}} \qquad (2.30) \\
&\le C \|D_{l+1:n} u\|_\infty \sum_{j=l+1}^{n} \Big\{ [(\mathbf{F}_{l+1:j}(s,\cdot))_j]_{\beta_j} (s-t)^{-l+\beta_j(j-\frac{1}{2})} + [(D_{j-1}\mathbf{F}_j(s,\cdot))_{j-1}]_\eta (s-t)^{-l+(1+\eta)(j-\frac{3}{2})} \Big\}.
\end{aligned}
$$

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

Write now for the other term,

$$
\begin{aligned}
&|D_{\mathbf{x}_1} I^{\boldsymbol{\xi}}_{l,2}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}} \\
&= \left| D_{\mathbf{x}_1} D_{\mathbf{x}_l} \left[\tilde{P}^{\boldsymbol{\xi}}_{s,t} \left(\sum_{i=2}^{l} \left\langle \mathbf{F}_i(s,\cdot) - \mathbf{F}_i(s,\cdot_{1:l-1}, \boldsymbol{\theta}^{l:n}_{s,t}(\boldsymbol{\xi})), D_i u(s,\cdot) \right\rangle \right) \right] (\mathbf{x}) \right| \Big|_{\boldsymbol{\xi}=\mathbf{x}} \\
&\leq \sum_{i=2}^{l} \int_{\mathbb{R}^{nd}} d\mathbf{y} |D_{\mathbf{x}_1} D_{\mathbf{x}_l} \tilde{p}^{\boldsymbol{\xi}}(s,t,\mathbf{x},\mathbf{y})|_{\boldsymbol{\xi}=\mathbf{x}} |\mathbf{F}_i(s,\mathbf{y}) - \mathbf{F}_i(s,\mathbf{y}_{1:l-1}, \boldsymbol{\theta}^{l:n}_{s,t}(\mathbf{x}))| |D_i u(s,\mathbf{y})| \\
&\leq \|D_{2:l} u\|_\infty \sum_{i=2}^{l} \sum_{j=l}^{n} \int_{\mathbb{R}^{nd}} d\mathbf{y} |D_{\mathbf{x}_1} D_{\mathbf{x}_l} \tilde{p}^{\boldsymbol{\xi}}(s,t,\mathbf{x},\mathbf{y})|_{\boldsymbol{\xi}=\mathbf{x}} [(\mathbf{F}_i)_j(s,\cdot)]_{\beta_j} |(\boldsymbol{\theta}_{s,t}(\mathbf{x}) - \mathbf{y})_j|^{\beta_j} \\
&\leq C \|D_{2:l} u\|_\infty \sum_{j=l}^{n} [(\mathbf{F}_{2:j})_j(s,\cdot)]_{\beta_j} (s-t)^{-l} \int_{\mathbb{R}^{nd}} d\mathbf{y} |D_{\mathbf{x}_1} D_{\mathbf{x}_l} \tilde{p}^{\boldsymbol{\xi}}(s,t,\mathbf{x},\mathbf{y})|_{\boldsymbol{\xi}=\mathbf{x}} |(\boldsymbol{\theta}_{s,t}(\mathbf{x}) - \mathbf{y})_j|^{\beta_j}.
\end{aligned}
$$

Hence, Lemma 7 also yields

$$
|D_{\mathbf{x}_1} I^{\boldsymbol{\xi}}_{l,2}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}} \leq C \|D_{2:l} u\|_\infty \sum_{j=l}^{n} [(\mathbf{F}_{2:j})_j(s,\cdot)]_{\beta_j} (s-t)^{-l+\beta_j(j-\frac{1}{2})}, \tag{2.31}
$$

and those terms are again integrable as soon as the thresholds of $(\mathbf{T}_\beta)$ hold.

• *Control of* $|D_{\mathbf{x}_1} I^{\boldsymbol{\xi}}_{l,3}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}}$. It hence remains to control the terms in $|D_{\mathbf{x}_1} I^{\boldsymbol{\xi}}_{l,3}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}}$. These terms are the tricky ones since they are, *a priori*, not designed to smoothen 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

$$
\begin{aligned}
&|D_{\mathbf{x}_1} I^{\boldsymbol{\xi}}_{l,3}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}} \\
&= \Bigg| D_{\mathbf{x}_1} D_{\mathbf{x}_l} \Bigg[\tilde{P}^{\boldsymbol{\xi}}_{s,t} \Bigg(\sum_{i=2}^{l} \Big\langle (\mathbf{F}_i(s,\cdot_{1:l-1}, \boldsymbol{\theta}^{l:n}_{s,t}(\boldsymbol{\xi})) - \mathbf{F}_i(\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})) \\
&\qquad - D_{i-1}\mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))(\cdot - \boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))_{i-1}), D_i u(s,\cdot) \Big\rangle \Bigg) \Bigg] (\mathbf{x}) \Bigg| \Bigg|_{\boldsymbol{\xi}=\mathbf{x}} \qquad \text{(I)} \\
&\leq C \|D_{2:l} u\|_\infty (s-t)^{-l} \Big(\sum_{j=2}^{l} [(\mathbf{F}_{2:j})_j(s,\cdot)]_{\beta_j} (s-t)^{(j-\frac{1}{2})\beta_j} + [(D_{j-1}\mathbf{F}_j)_{j-1}(s,\cdot)]_\eta (s-t)^{(1+\eta)(j-\frac{3}{2})} \Big) \\
&\leq C \|D_{2:l} u\|_\infty (s-t)^{-l} \Big((s-t)^{\frac{3}{2}\beta_2} + (s-t)^{\frac{1}{2}(1+\eta)} \Big),
\end{aligned}
$$

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 η is meant to be small). To overcome this problem, the idea consists in writing, thanks to the cancellation properties of Lemma 8,

$$
\begin{aligned}
&|D_{\mathbf{x}_1} I^{\boldsymbol{\xi}}_{l,3}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}} \\
&= \Big| D_{\mathbf{x}_1} D_{\mathbf{x}_l} \Big[\tilde{P}^{\boldsymbol{\xi}}_{s,t} \Big(\sum_{i=2}^{l} \big\langle (\mathbf{F}_i(s,\cdot_{1:l-1}, \boldsymbol{\theta}^{l:n}_{s,t}(\boldsymbol{\xi})) - \mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})) - D_{i-1}\mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))\big(\cdot - \boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})\big)_{i-1}), \\
&\qquad \big(D_i u(s,\cdot) - D_i u(s,\cdot_{1:l-1}, \boldsymbol{\theta}^{l:n}_{s,t}(\boldsymbol{\xi})) \big) \big\rangle \Big) \Big] (\mathbf{x}) \Big| \Big|_{\boldsymbol{\xi}=\mathbf{x}}
\end{aligned} \tag{2.32}
$$

and to take advantage of the additional smoothing effect from the solution of the regularized PDE itself through the above contribution $D_i u(s,\cdot) - D_i u(s,\cdot_{1:l-1}, \boldsymbol{\theta}^{l:n}_{s,t}(\boldsymbol{\xi}))$, $i \in [\![2,l]\!]$. This was the strategy implemented in [CdR17] which, unfortunately, cannot be repeated as this in our general framework. Roughly speaking, to control for any k in $[\![l,n]\!]$, for fixed $(\mathbf{y}_{1:k-1}, \mathbf{y}_{k+1:n}) \in \mathbb{R}^{(n-1)d}$ and any $0 \leq s < t \leq T$, the α^i_k-Hölder modulus of the partial application (see (1.14)) $\mathbf{y}_k \mapsto (D_i u)_k(s,\mathbf{y}_k)$, we need to control the α^i_k-Hölder modulus of the map $\mathbf{y}_k \mapsto (D_i \tilde{G}^{\boldsymbol{\xi}} f)_k(s,\mathbf{y}_k)$, where $\tilde{G}^{\boldsymbol{\xi}} f(s,\mathbf{y}) := \int_s^T dv \tilde{P}^{\boldsymbol{\xi}}_{v,s} f(v,\mathbf{y})$ is the Green kernel involving the source f in the indicated equation. When doing so, we are faced to the fact that we should only consider indexes $k \leq i$

in order to apply cancellation arguments. The only corresponding index in (2.32) is thus $i = k = l$. When $i = l$ and k lies in $[\![l+1, n]\!]$, we can handle the corresponding terms by interpolating the α_k^l-Hölder modulus of the partial application $(D_l u)_k$ from the α_l^l-Hölder modulus of the partial application $(D_l u)_l$, see Lemma 12 below. Otherwise, i.e. when $i < l$, we expand the gradients in (2.32) with the Taylor formula. Namely, using the Fubini theorem we write:

$$
\begin{aligned}
&|D_{\mathbf{x}_1} I^{\boldsymbol{\xi}}_{l,3}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}} \\
&\leq \Bigg| \sum_{i=2}^{l-1}\sum_{k=l}^{n} \int_0^1 d\lambda D_{\mathbf{x}_1} D_{\mathbf{x}_l} \Bigg[\tilde{P}^{\boldsymbol{\xi}}_{s,t}\Bigg(\Bigg\langle \Big(\mathbf{F}_i(s,\cdot_{1:l-1},\boldsymbol{\theta}^{l:n}_{s,t}(\boldsymbol{\xi})) - \mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})) - D_{i-1}\mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))\big(\cdot - \boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})\big)_{i-1}\Big), \\
&D_k D_i u(s,\cdot_{1:l-1},\boldsymbol{\theta}^{l:n}_{s,t}(\boldsymbol{\xi}) + \lambda(\cdot_{l:n} - \boldsymbol{\theta}^{l:n}_{s,t}(\boldsymbol{\xi})))(\cdot - \boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))_k \Big\rangle \Bigg)\Bigg](\mathbf{x})\Bigg|_{\boldsymbol{\xi}=\mathbf{x}} \\
&+ \Bigg| D_{\mathbf{x}_1} D_{\mathbf{x}_l} \Bigg[\tilde{P}^{\boldsymbol{\xi}}_{s,t}\Bigg(\Bigg\langle \Big(\mathbf{F}_l(s,\cdot_{1:l-1},\boldsymbol{\theta}^{l:n}_{s,t}(\boldsymbol{\xi})) - \mathbf{F}_l(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})) - D_{l-1}\mathbf{F}_l(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))\big(\cdot - \boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})\big)_{l-1}\Big), \\
&\big(D_l u(s,\cdot) - D_l u(s,\cdot_{1:l-1},\boldsymbol{\theta}^{l:n}_{s,t}(\boldsymbol{\xi}))\big) \Big\rangle \Bigg)\Bigg](\mathbf{x})\Bigg|_{\boldsymbol{\xi}=\mathbf{x}} \\
&=: |D_{\mathbf{x}_1} I^{\boldsymbol{\xi}}_{l,31}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}} + |D_{\mathbf{x}_1} I^{\boldsymbol{\xi}}_{l,32}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}}.
\end{aligned}
\tag{2.33}
$$

As already underlined, the term $|D_{\mathbf{x}_1} I^{\boldsymbol{\xi}}_{l,32},(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}}$ is handled thanks to interpolation type argument. It thus remains to deal with the term $|D_{\mathbf{x}_1} I^{\boldsymbol{\xi}}_{l,31}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}}$. Our main idea consists first in switching the differential operators acting on the map u therein and then in rebalancing the differential operator w.r.t. the "less degenerate" direction (namely D_i) on the remaining terms in the integrand through an integration by parts. To do so, we introduce now for all $i \in [\![2, l-1]\!], k \in [\![l,n]\!]$, $(\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}) \in \mathbb{R}^{(n-1)d}, (t,\mathbf{x}) \in [0,T]\times\mathbb{R}^{nd}$, $s \in (t,T]$ and $\mathbf{y}_i \in \mathbb{R}^d$, the function:

$$
\begin{aligned}
&\Psi^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}_{i,(l,1),k}(\mathbf{y}_i) \\
:= \;& \Big[D_{\mathbf{x}_1} D_{\mathbf{x}_l} \tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y}) \\
&\otimes \Big((\mathbf{F}_i(s,\mathbf{y}_{1:l-1},\boldsymbol{\theta}^{l:n}_{s,t}(\boldsymbol{\xi})) - \mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})) D_{i-1}\mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))\big(\mathbf{y} - \boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})\big)_{i-1})\Big)((\mathbf{y} - \boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))_k)^* \Big]_{\boldsymbol{\xi}=\mathbf{x}},
\end{aligned}
\tag{2.34}
$$

where the subscript $(l,1)$ in $\Psi^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}_{i,(l,1),k}$ is here to indicate the differentiation w.r.t. $D_{\mathbf{x}_l} D_{\mathbf{x}_1}$ 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 $|D_{\mathbf{x}_1} I^{\boldsymbol{\xi}}_{l,31}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}}$:

$$
\begin{aligned}
|D_{\mathbf{x}_1} I^{\boldsymbol{\xi}}_{l,31}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}} \leq \sum_{i=2}^{l-1}\sum_{k=l}^{n} \Bigg| & \int_0^1 d\lambda \int_{\mathbb{R}^{(n-1)d}} d(\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}) \\
& \int_{\mathbb{R}^d} d\mathbf{y}_i \Bigg\{ \big(\Psi^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}_{i,(l,1),k}(\mathbf{y}_i)\big) : D_{\mathbf{y}_k} D_{\mathbf{y}_i} u(s,\mathbf{y}_{1:l-1},\boldsymbol{\theta}^{l:n}_{s,t}(\mathbf{x}) + \lambda(\mathbf{y}_{l:n} - \boldsymbol{\theta}^{l:n}_{s,t}(\mathbf{x}))) \Bigg\} \Bigg|,
\end{aligned}
$$

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

$$
\begin{aligned}
|D_{\mathbf{x}_1} I^{\boldsymbol{\xi}}_{l,31}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}} \leq \sum_{i=2}^{l-1}\sum_{k=l}^{n} \Bigg| & \int_0^1 d\lambda \int_{\mathbb{R}^{(n-1)d}} d(\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}) \\
& \int_{\mathbb{R}^d} d\mathbf{y}_i \Bigg\{ D_{\mathbf{y}_i}\big(\Psi^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}_{i,(l,1),k}(\mathbf{y}_i)\big) : D_{\mathbf{y}_k} u(s,\mathbf{y}_{1:l-1},\boldsymbol{\theta}^{l:n}_{s,t}(\mathbf{x}) + \lambda(\mathbf{y}_{l:n} - \boldsymbol{\theta}^{l:n}_{s,t}(\mathbf{x}))) \Bigg\} \Bigg|.
\end{aligned}
\tag{2.35}
$$

Let us now explain why such an expression is well designed. From the very definition of the "Ψ" term in (2.34), one can see that the additional contribution $(\cdot - \boldsymbol{\theta}_{s,t}(\mathbf{x}))_k$ is, thanks to Lemma 7, precisely designed to smoothen the time singularity coming from the differentiation w.r.t. the variables $\mathbf{x}_l$ of the semigroup (recall that $k \geq l$). 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 $\sum_{j=i}^n(|\cdot-\boldsymbol{\theta}_{s,t}(\mathbf{x}))_j|^{\beta_j}+|(\cdot-\boldsymbol{\theta}_{s,t}(\mathbf{x}))_{j-1}|^{1+\eta}$. Therefore, it follows from Lemma 7 and $(\mathbf{T}_\beta)$ that this term allows to smoothen the differentiation coming from the Schwarz theorem and the integration by parts (pay attention that for any i in $[\![2,l-1]\!]$, the differential operator D_i now acts on the term $\Psi^{\cdot}_{i,\cdot}$). The main idea consists now in absorbing the additional singularity (of order 1/2) coming from the differentiation w.r.t. the variable "$\mathbf{x}_1$" thanks to the α_i^k-Hölder regularity of $(D_ku)_i$, provided the exponent α_i^k is large enough (see Remark 7 below). This is done by putting the two above terms "in duality" w.r.t. the variable $\mathbf{y}_i$, within the framework of Besov spaces.

Recall indeed that $C^{\alpha_i^k}(\mathbb{R}^d,\mathbb{R})=B^{\alpha_i^k}_{\infty,\infty}(\mathbb{R}^d,\mathbb{R})$, where from now on the notation $B^s_{p,q}$ stands for a Besov space with associated indexes p,q,s (see Triebel [Tri83] and the reminder at the beginning of Section 3). The indexes p,q refer to the integrability parameters and s to the smoothness one. A classical fact is that $B^{\alpha_i^k}_{\infty,\infty}$ and $B^{-\alpha_i^k}_{1,1}$ can be put in duality, see e.g. Proposition 3.6 in Lemarié-Rieusset ([LR02]). Indeed, with the notations therein $B^{\alpha_i^k}_{\infty,\infty}$ is the dual of $B^{-\alpha_i^k}_{1,1}=\tilde{B}^{-\alpha_i^k}_{1,1}$ where $\tilde{B}^{-\alpha_i^k}_{1,1}$ denotes the closure of the Schwartz class $\mathcal{S}$ in $B^{-\alpha_i^k}_{1,1}$ (see also Theorem 4.1.3 in Adams Hedberg [AH96] for the density of $\mathcal{S}$ in $B^{-\alpha_i^k}_{1,1}$). Exploiting this fact, we then derive from (2.35) and the multi-linearity of the tensors involved that:

$$
\begin{aligned}
&|D_{\mathbf{x}_1}I^{\boldsymbol{\xi}}_{l,31}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}}\\
&\le\ C\sum_{i=2}^{l-1}\sum_{k=l}^{n}\int_0^1 d\lambda\int_{\mathbb{R}^{(n-1)d}}d(\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n})\\
&\times\Big\|D_i\Psi^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}_{i,(l,1),k}(\cdot)\Big\|_{B^{-\alpha_i^k}_{1,1}}\Big\|\mathbf{y}_i\mapsto D_ku(s,\mathbf{y}_{1:l-1},\boldsymbol{\theta}^{l:n}_{s,t}(\mathbf{x})+\lambda(\mathbf{y}_{l:n}-\boldsymbol{\theta}^{l:n}_{s,t}(\mathbf{x})))\Big\|_{B^{\alpha_i^k}_{\infty,\infty}}\\
&\le C\sum_{i=2}^{l-1}\sum_{k=l}^{n}\int_{\mathbb{R}^{(n-1)d}}d(\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n})\Big\|D_i\Psi^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}_{i,(l,1),k}(\cdot)\Big\|_{B^{-\alpha_i^k}_{1,1}}\sup_{\mathbf{z}_j,j\in[\![1,n]\!],j\neq i}\Big\|D_ku(s,\mathbf{z}_{1:i-1},\cdot,\mathbf{z}_{i+1:n})\Big\|_{B^{\alpha_i^k}_{\infty,\infty}}.
\end{aligned}
$$

To handle this term and conclude the proof of the main theorem, we now need the following results whose proofs are postponed to the next section:

Lemma 10. *Let $l\in[\![2,n]\!]$, $i\in[\![2,l-1]\!]$ and $k\in[\![l,n]\!]$ and let $\Psi^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}_{i,(l,1),k}:\mathbb{R}^d\to(\mathbb{R}^d)^{\otimes 4}$ be the function defined by (2.34). There exist $C:=C((\mathbf{A}),T)>0$, $\bar{C}:=\bar{C}(C)$, $\alpha_i^k:=(1+\frac{\eta}{4})/(2i-1)$, $\gamma_i^k:=\frac{1}{2}+\eta(i-\frac{3}{2})$ such that*

$$
\Big\|\Big[D_i\Psi^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}_{i,(l,1),k}(\cdot)\Big]\Big\|_{B^{-\alpha_i^k}_{1,1}}\le\bar{C}(s-t)^{-\frac{3}{2}+\gamma_i^k}\hat{q}_{c\backslash i}(t,s,\mathbf{x},(\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n})),
$$

where $c=C^{-1}$ and with the notations of Proposition 5, exploiting as well equation (G_{NL}) of Remark 6,

$$
\hat{q}_{c\backslash i}(t,s,\mathbf{x},(\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n})):=\int_{\mathbb{R}^d}d\mathbf{y}_i\hat{p}^{\mathbf{x}}_c(t,s,\mathbf{x},\mathbf{y})=\prod_{j\in[\![1,n]\!],j\neq i}^{n}\mathcal{N}_{c(s-t)^{2j-1}}\big((\boldsymbol{\theta}_{s,t}(\mathbf{x})-\mathbf{y})_j\big),\tag{2.36}
$$

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

Lemma 11. *Let u be the solution of (2.1). There exists $C:=C((\mathbf{A}))>0$ such that for all $i\le k\in[\![2,n]\!]^2$ and $\alpha_i^k=(1+\frac{\eta}{4})/(2i-1)$,*

$$
\sup_{\mathbf{y}_j,j\in[\![1,n]\!],j\neq i}\Big\|D_ku\big(s,\mathbf{y}_{1:i-1},\cdot,\mathbf{y}_{i+1:n}\big)\Big\|_{B^{\alpha_i^k}_{\infty,\infty}}\le C(\|\mathbf{D}u\|_\infty+\|\mathbf{D}D_1u\|_\infty).\tag{2.37}
$$

Remark 7 (About the Hölder exponent α_i^k, $2\le i\le k\le n$.)**.** *We emphasize that upper bounds of the Hölder exponents $(\alpha_i^k)_{2\le i\le k\le n}$ can be derived from the analysis of the Hölder modulus of the partial application $\mathbf{y}_i\mapsto(D_k\tilde{G}^{\boldsymbol{\xi}}f)_i(s,\mathbf{y}_i)$, where $\tilde{G}^{\boldsymbol{\xi}}f(s,\mathbf{y}):=\int_s^T dv\tilde{P}^{\boldsymbol{\xi}}_{v,s}f(v,\mathbf{y})$ is the Green kernel involving the source f in the indicated equation (see Appendix C.2 or Section 5.1 in [CdRHM20] for similar issues).*
Let us briefly sketch why these upper bounds are large enough to absorb the additional time singularity of order 1/2 coming from the differentiation w.r.t. the variable $\mathbf{x}_1$. First note that the computations done in Appendix C.2 yield for each $(i\le k)$ in $[\![2,n]\!]^2$ the upper bound $[1-(1-\beta_k)(k-1/2)]/[i-1/2]$. Recall then from $(\mathbf{T}_\beta)$

that $\beta_k \in \big((2k-2)/(2k-1),1\big]$. *Thus, if* $\beta_k = \big([2k-2]/[2k-1]\big)^+ := (2k-2)/(2k-1)+\varepsilon$ *for some* $0<\varepsilon<<1$ *meant to be very small, we obtain* $(1-\beta_k)(k-1/2) = (1/2)^- := 1/2-\varepsilon$. *Therefore, the smoothing effect of the gradient of the solution of the PDE in the direction "*$\mathbf{x}_k$*" w.r.t. the* i^{th} *variable is designed to smoothen a time singularity of order* $\alpha_i^k(i-1/2) \le [(1/2)^+/(i-1/2)] \times (i-1/2) = (1/2)^+$.

To control the term $|D_{\mathbf{x}_1} I^{\boldsymbol{\xi}}_{l,32}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}}$ in (2.33), we will need the following auxiliary lemma whose proof is postponed to Appendix D.

Lemma 12 (Hölder moduli of the gradients after the differentiation index through reverse Taylor expansion)**.** *There is a constant* $C>0$ *s.t, for all* $l \in [\![2,n]\!]$, $k \in [\![l,n]\!]$, $(\mathbf{y},\boldsymbol{\xi}) \in \mathbb{R}^{nd}\times\mathbb{R}^{nd}$, $s\in[t,T]$ *:*

$$\big|D_l u(s,\mathbf{y}) - D_l u(s,\mathbf{y}_{1:l-1},\boldsymbol{\theta}^{l:n}_{s,t}(\boldsymbol{\xi}))\big| \le C(\|\mathbf{D}u\|_\infty + [(D_l u)_l(s,\cdot)]_{\alpha_l^l})\Big(|(\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})-\mathbf{y})_l|^{\alpha_l^l} + \sum_{k=l+1}^n |(\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})-\mathbf{y})_k|^{\zeta_l}\Big),$$

where $\zeta_l := (1+\eta/4)/(2l+\eta/4)$.

Now, from Proposition 5, Lemma 9 and Lemma 12,

$$\begin{aligned}
&|D_{\mathbf{x}_1} I^{\boldsymbol{\xi}}_{l,32}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}}\\
\le\ & \frac{\bar C[(D_{l-1}\mathbf{F}_l)_{l-1}(s,\cdot))]_\eta}{(s-t)^l}\int_{\mathbb{R}^{nd}} d\mathbf{y}\hat p^{\mathbf{x}}_{C^{-1}}(t,s,\mathbf{x},\mathbf{y})|(\boldsymbol{\theta}_{s,t}(\mathbf{x})-\mathbf{y})_{l-1}|^{1+\eta}\\
&\times(\|\mathbf{D}u\|_\infty + [(D_l u)_l(s,\cdot)]_{\alpha_l^l})\big(|(\boldsymbol{\theta}_{s,t}(\mathbf{x})-\mathbf{y})_l|^{\alpha_l^l} + \sum_{k=l+1}^n |(\boldsymbol{\theta}_{s,t}(\mathbf{x})-\mathbf{y})_k|^{\zeta_l}\big)\\
\le\ & C(\|\mathbf{D}u\|_\infty + [(D_l u)_l(s,\cdot)]_{\alpha_l^l})\big(\sum_{k=l}^n (s-t)^{-l+(1+\eta)(l-\frac32)+\alpha_l^l(l-\frac12)\mathbb{I}_{k=l}+\zeta_l(k-\frac12)\mathbb{I}_{k\in[\![l+1,n]\!]}}\Big),
\end{aligned} \tag{2.38}$$

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

$$|D_{\mathbf{x}_1} I^{\boldsymbol{\xi}}_{l,32}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}} \le C\big(\|\mathbf{D}u\|_\infty + \|(D_l u)_l(s,\cdot)\|_{B^{\alpha_l^l}_{\infty,\infty}}\big)\sum_{k=l}^n (s-t)^{-l+(1+\eta)(l-\frac32)+\alpha_l^l(l-\frac12)\mathbb{I}_{k=l}+\zeta_l(k-\frac12)\mathbb{I}_{k\in[\![l+1,n]\!]}}. \tag{2.39}$$

Observe now that the time exponents in the previous r.h.s. are integrable. In particular, for each $l\in[\![2,n]\!]$

$$\begin{aligned}
-l+(1+\eta)(l-\frac32)+\alpha_l^l(l-\frac12) &= -(l-\frac32)(1-(1+\eta)) - \frac32 + \alpha_l^l(l-\frac12) \ge \eta(l-\frac32) - \frac32 + \frac{1+\frac\eta4}{2l-1}(l-\frac12)\\
&\ge -\frac32 + \eta(l-\frac32) + \frac12 = -\frac32 + \gamma_l^k,
\end{aligned} \tag{2.40}$$

with the notations of Lemma 10. Also, for $k\in[\![l+1,n]\!]$:

$$\begin{aligned}
-l+(1+\eta)(l-\frac32)+\zeta_l(k-\frac12) &= -\frac32+\eta(l-\frac32)+\frac{1+\frac\eta4}{2l+\frac\eta4}(k-\frac12) \ge -\frac32+\eta(l-\frac32)+\frac12\frac{1+\frac\eta4}{l+\frac\eta8}(l+\frac12)\\
&\ge -\frac32+\eta(l-\frac32)+\frac12 = -\frac32+\gamma_l^k.
\end{aligned} \tag{2.41}$$

Observe carefully from the above computations that the exponent ζ_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, see Remark 7 above.

We can then deduce from (2.36) and (2.39), using Lemma 10 and Lemma 11 that

$$|D_{\mathbf{x}_1} I^{\boldsymbol{\xi}}_{l,3}(s,\mathbf{x})|_{\boldsymbol{\xi}=\mathbf{x}} \le C\sum_{i=2}^l\sum_{k=l}^n (s-t)^{-\frac32+\gamma_i^k}\Big(\|\mathbf{D}u\|_\infty + \|\mathbf{D}D_1 u\|_\infty\Big), \tag{2.42}$$

which are integrable terms since $\gamma_i^k > 1/2$. With the notations of (2.21), (2.29), we eventually derive from (2.42), (2.31), (2.30) that there exists $\gamma := \gamma((\mathbf{A})) > 0$ such that:

$$\Big|\int_t^T ds D_{\mathbf{x}_1} I^{\boldsymbol{\xi}}_l(s,\mathbf{x})\Big|_{\boldsymbol{\xi}=\mathbf{x}} \le CT^\gamma\Big(\|\mathbf{D}u\|_\infty + \|\mathbf{D}D_1 u\|_\infty\Big). \tag{2.43}$$

Final proof of Theorem 3. Bringing together (2.43) and (2.28) yields for all $l \in [\![1,n]\!]$ and $(t,\mathbf{x}) \in [0,T]$:

$$|D_{\mathbf{x}_l} D_{\mathbf{x}_1} u(t,\mathbf{x})| \leq C(T^\gamma + T^\delta)\Big(\|\mathbf{D}u\|_\infty + \|\mathbf{D}D_1 u\|_\infty\Big). \tag{2.44}$$

It is clear that the previous analysis can be reproduced without differentiating w.r.t. $\mathbf{x}_1$, leading to improved singularity exponents (see also the proof of Lemma 11 which somehow exactly explicit these computations). We therefore get:

$$|D_{\mathbf{x}_l} u(t,\mathbf{x})| \leq C(T^{\gamma'} + T^{\delta'})\Big(\|\mathbf{D}u\|_\infty + \|\mathbf{D}D_1 u\|_\infty\Big), \tag{2.45}$$

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

Taking the time-space supremum in the l.h.s of (2.44) and (2.45), recalling as well that T is meant to be small, i.e. s.t. $4CT^{\delta\wedge\gamma} \leq 1/2$, we derive:

$$\|\mathbf{D}u\|_\infty + \|\mathbf{D}D_1 u\|_\infty \leq 2C(T^\gamma + T^\delta).$$

This concludes the proof. □

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 on Besov spaces from Section 2.6.4 of Triebel [Tri83]. For $\alpha \in \mathbb{R}, q \in (0,+\infty], p \in (0,\infty]$, $B^\alpha_{p,q}(\mathbb{R}^d) := \{f \in \mathcal{S}'(\mathbb{R}^d) : \|f\|_{\mathcal{H}^\alpha_{p,q}} < +\infty\}$ where $\mathcal{S}(\mathbb{R}^d)$ stands for the Schwartz class and

$$\|f\|_{\mathcal{H}^\alpha_{p,q}} := \|\varphi(D)f\|_{L^p(\mathbb{R}^d)} + \Big(\int_0^1 \frac{dv}{v} v^{(m-\frac{\alpha}{2})q} \|\partial_v^m h_v \star f\|^q_{L^p(\mathbb{R}^d)}\Big)^{\frac{1}{q}}, \tag{3.1}$$

with $\varphi \in C^\infty_0(\mathbb{R}^d)$ (smooth function with compact support) s.t. $\varphi(0) \neq 0$, $\varphi(D)f := (\varphi \hat{f})^\vee$ where $\hat{f}$ and $(\varphi\hat{f})^\vee$ respectively denote the Fourier transform of f and the inverse Fourier transform of $\varphi\hat{f}$. The parameter m is an integer s.t. $m > \alpha/2$ and for $v > 0$, $z \in \mathbb{R}^d$, $h_v(z) := \frac{1}{(2\pi v)^{d/2}} \exp\Big(-\frac{|z|^2}{2v}\Big)$ 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 $\|f\|_{\mathcal{H}^s_{p,q}}$, where the subscript $\mathcal{H}$ stands to indicate the dependence on the heat-kernel, depends on the considered function φ and the chosen $m \in \mathbb{N}$. It also defines a quasi-norm on $B^s_{p,q}(\mathbb{R}^d)$. The previous definition of $B^\alpha_{p,q}(\mathbb{R}^d)$ 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 $\|f\|_{\mathcal{H}^\alpha_{p,q}} =: \|f\|_{B^\alpha_{p,q}}$.

3.1 Proof of Lemma 10

We will here exploit the thermic characterization of Besov spaces (see Chapter 2.6.4 in [Tri83]) 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 $k \in [\![l,n]\!]$:

$$\|\varphi(D) D_i \Psi^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}_{i,(l,1),k}(\cdot)\|_{L^1(\mathbb{R}^d,\mathbb{R})} + \int_0^1 \frac{dv}{v} v^{1+\frac{\alpha^k_i}{2}} \|\partial_v h_v \star D_i \Psi^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}_{i,(l,1),k}(\cdot)\|_{L^1(\mathbb{R}^d,\mathbb{R})},$$

for a C^∞ compactly supported function φ s.t. $\varphi(0) \neq 0$. From the definition of $\Psi^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}_{i,(l,1),k}$ in (2.34), using Lemmas 7 and 9, we easily deduce

$$\begin{aligned}
& \|\varphi(D) D_i \Psi^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}_{i,(l,1),k}(\cdot)\|_{L^1(\mathbb{R}^d,\mathbb{R})} \\
\leq\ & \int_{\mathbb{R}^d} dz \Big| \int_{\mathbb{R}^d} d\mathbf{y}_i D_{\mathbf{y}_i} \widehat{\varphi}(z - \mathbf{y}_i) \cdot \Psi^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}_{i,(l,1),k}(\mathbf{y}_i) \Big| \\
\leq\ & \bar{C}(s-t)^{-1/2}\Big((s-t)^{\frac{3}{2}\beta_2} + (s-t)^{\frac{1}{2}(1+\eta)}\Big) \hat{q}_{c\setminus i}(t,s,\mathbf{x},(\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n})).
\end{aligned}$$

Let us now focus on the second term in the above definition. We split therein the time integral into two parts writing:

$$\begin{aligned}
&\int_0^{(s-t)^{\rho_{i,k}}} dv v^{\frac{\alpha_i^k}{2}} \|\partial_v h_v \star D_i \Psi_{i,(l,1),k}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}(\cdot)\|_{L^1(\mathbb{R}^d,\mathbb{R})} \\
&+ \int_{(s-t)^{\rho_{i,k}}}^1 dv v^{\frac{\alpha_i^k}{2}} \|\partial_v h_v \star D_i \Psi_{i,(l,1),k}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}(\cdot)\|_{L^1(\mathbb{R}^d,\mathbb{R})} \qquad (3.2)\\
&=: \mathbf{Lower}\Big[D_i \Psi_{i,(l,1),k}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}\Big] + \mathbf{Upper}\Big[D_i \Psi_{i,(l,1),k}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}\Big],
\end{aligned}$$

for a parameter $\rho_{i,k} > 0$ to be specified. The term **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 $\rho_{i,k}$ in order to match the integrability constraint

$$\mathbf{Upper}\Big[D_i \Psi_{i,(l,1),k}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}\Big] \leq \frac{\bar{C}}{(s-t)^{1+\frac{1}{2}-\gamma_i^k}} \hat{q}_{c\setminus i}(t,s,\mathbf{x},(\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n})), \qquad (3.3)$$

where $\hat{q}_{c\setminus i}$ has been defined in (2.36) and $\gamma_i^k > 1/2$ 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 $(\mathbf{T}_\beta)$. On the other hand, the contribution **Lower** in (3.2) 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 (3.3).

We now write:

$$\begin{aligned}
&\mathbf{Upper}\Big[D_i \Psi_{i,(l,1),k}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}\Big] \\
=\;& \int_{(s-t)^{\rho_{i,k}}}^1 dv v^{\frac{\alpha_i^k}{2}} \|\partial_v h_v \star D_i \Psi_{i,(l,1),k}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}(\cdot)\|_{L^1(\mathbb{R}^d,\mathbb{R})} \\
=\;& \int_{(s-t)^{\rho_{i,k}}}^1 dv v^{\frac{\alpha_i^k}{2}} \int_{\mathbb{R}^d} dz \Big| \int_{\mathbb{R}^d} d\mathbf{y}_i D_i \partial_v h_v(z-\mathbf{y}_i) \Psi_{i,(l,1),k}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}(\mathbf{y}_i)\Big|.
\end{aligned}$$

Recall from the definition in (2.34) that $\Psi_{i,(l,1),k}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}(\mathbf{y}_i)$ is $(\mathbb{R}^d)^{\otimes 4}$-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 $\Psi_{i,(l,1),k}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}(\mathbf{y}_i)$ leads to:

$$\begin{aligned}
&\mathbf{Upper}\Big[D_i \Psi_{i,(l,1),k}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}\Big] \\
\leq\;& \int_{(s-t)^{\rho_{i,k}}}^1 dv v^{\frac{\alpha_i^k}{2}} \int_{\mathbb{R}^d} dz \Big| \int_{\mathbb{R}^d} d\mathbf{y}_i D_i \partial_v h_v(z-\mathbf{y}_i) \Big(D_{\mathbf{x}_1} D_{\mathbf{x}_l} \tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y}) \Big[((\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}) - \mathbf{y})_k) \\
&\Big(\mathbf{F}_i(s,\mathbf{y}_{1:l-1},\boldsymbol{\theta}_{s,t}^{l:n}(\boldsymbol{\xi})) - \mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})) - D_{i-1}\mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))\big(\mathbf{y} - \boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})\big)_{i-1} \Big] \Big) \Big|_{\boldsymbol{\xi}=\mathbf{x}}.
\end{aligned}$$

From Lemma 9 and Proposition 5, we derive that there exist $C := C((\mathbf{A}),T) > 0$, $\bar{C} := \bar{C}(C)$ such that introducing $\hat{q}_c(t,s,\mathbf{x},\mathbf{y}) = \hat{p}_c^{\mathbf{x}}(t,s,\mathbf{x},\mathbf{y})$, $c = C^{-1}$:

$$\begin{aligned}
&\mathbf{Upper}\Big[D_i \Psi_{i,(l,1),k}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}\Big] \\
\leq\;& \bar{C} \int_{(s-t)^{\rho_{i,k}}}^1 dv v^{\frac{\alpha_i^k}{2}} \int_{\mathbb{R}^d} dz \int_{\mathbb{R}^d} d\mathbf{y}_i \frac{h_{cv}(z-\mathbf{y}_i)}{v^{\frac{3}{2}}} \frac{\hat{q}_c(t,s,\mathbf{x},\mathbf{y})}{(s-t)^{(l-\frac{1}{2})+\frac{1}{2}}} |(\boldsymbol{\theta}_{s,t}(\mathbf{x}) - \mathbf{y})_k| \\
&\times \Big\{ \sum_{j=i}^{l-1} \Big\{ |(\boldsymbol{\theta}_{s,t}(\mathbf{x}) - \mathbf{y})_j|^{\beta_j} \Big\} + |(\boldsymbol{\theta}_{s,t}(\mathbf{x}) - \mathbf{y})_{i-1}|^{1+\eta} \Big\} \\
\underset{k\geq l}{\leq}\;& \bar{C} \int_{(s-t)^{\rho_{i,k}}}^1 dv v^{\frac{\alpha_i^k}{2}} \int_{\mathbb{R}^d} dz \int_{\mathbb{R}^d} d\mathbf{y}_i \frac{h_{cv}(z-\mathbf{y}_i)}{v^{\frac{3}{2}}} \frac{\hat{q}_c(t,s,\mathbf{x},\mathbf{y})}{(s-t)^{\frac{1}{2}}} \Big\{ \sum_{j=i}^{l-1} (s-t)^{\beta_j(j-\frac{1}{2})} + (s-t)^{(1+\eta)(i-\frac{3}{2})} \Big\} \\
\underset{\text{Fubini}}{\leq}\;& \bar{C}\hat{q}_{c\setminus i}(t,s,\mathbf{x},(\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n})) \int_{(s-t)^{\rho_{i,k}}}^1 dv v^{-\frac{3}{2}+\frac{\alpha_i^k}{2}} (s-t)^{-\frac{1}{2}} \Big\{ \sum_{j=i}^{l-1} (s-t)^{\beta_j(j-\frac{1}{2})} + (s-t)^{(1+\eta)(i-\frac{3}{2})} \Big\} \\
\leq\;& \bar{C}\hat{q}_{c\setminus i}(t,s,\mathbf{x},(\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}))(s-t)^{[-\frac{1}{2}+\frac{\alpha_i^k}{2}]\rho_{i,k}-\frac{1}{2}} \big((s-t)^{\beta_i(i-\frac{1}{2})} + (s-t)^{(1+\eta)(i-\frac{3}{2})} \big),
\end{aligned}$$

recalling that the lower bound of $\beta_j(j-1/2)$ is increasing for the last inequality (recall indeed that we assumed that $\beta_j \in \left(\frac{2j-2}{2j-1}, 1\right]$), up to modification of $C, \bar{C}$ in the previous inequalities. We now want to choose the threshold $\rho_{i,k}$ in order to match the integrability condition in (3.3). This amounts to write:

$$-1-\frac{1}{2}+\gamma_i^k=[-\frac{1}{2}+\frac{\alpha_i^k}{2}]\rho_{i,k}-\frac{1}{2}+\Big(\beta_i(i-\frac{1}{2})\wedge(1+\eta)(i-\frac{3}{2})\Big).$$

Since this condition should hold for any $\beta_i \in \left(\frac{2i-2}{2i-1}, 1\right]$ and since the parameter $\eta \in (0,1)$ is small (see $(\mathbf{H}_\eta)$) we have $\beta_i(i-1/2)\wedge(1+\eta)(i-3/2)=(1+\eta)(i-3/2)$. The above condition rewrites:

$$-1-\frac{1}{2}+\gamma_i^k=[-\frac{1}{2}+\frac{\alpha_i^k}{2}]\rho_{i,k}-\frac{1}{2}+(1+\eta)(i-\frac{3}{2}).$$

Our global integrability constraint associated with the i^{th} variable in the k^{th} derivative writes:

$$-1-\frac{1}{2}+\gamma_i^k=[-\frac{1}{2}+\frac{\alpha_i^k}{2}]\rho_{i,k}-\frac{1}{2}+(1+\eta)(i-\frac{3}{2})>-1, \tag{3.4}$$

which gives

$$\rho_{i,k}<\frac{(1+\eta)(2i-3)+1}{1-\alpha_i^k}. \tag{3.5}$$

It therefore remains to check that such a choice is compatible with the time integral part for $v \in [0,(s-t)^{\rho_{i,k}}]$ in the thermic characterisation of the Besov norm, see (3.2). 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. $D_i\partial_v h_v(z-\mathbf{y}_i)$. In order to get an integrable singularity in v, we need to decrease the crude upper-bound on $D_i\partial_v h_v(z-\mathbf{y}_i)$. This is done through cancellation techniques exploiting the smoothness properties of $\Psi_{i,(l,1),k}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}$.

To investigate $\mathbf{Lower}\Big[D_i\Psi_{i,(l,1),k}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}\Big]$ let us first recall from the definition in (2.34) that for each $k \in [\![l,n]\!]$:

$$\begin{aligned}
&\Psi_{i,(l,1),k}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}(\mathbf{y}_i)\\
=\;&\Big\{D_{\mathbf{x}_l}D_{\mathbf{x}_1}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y})\\
&\times\Big[\Big(\mathbf{F}_i(s,\mathbf{y}_{1:l-1},\boldsymbol{\theta}_{s,t}^{l:n}(\boldsymbol{\xi}))-\mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))-D_{i-1}\mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))\big(\mathbf{y}-\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})\big)_{i-1}\Big)\Big](\mathbf{y}-\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))_k^*\Big\}_{\boldsymbol{\xi}=\mathbf{x}}.
\end{aligned} \tag{3.6}$$

Let us now specify the dependence w.r.t. $\mathbf{y}_i$ 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 $\mathbf{y}_i$ in (3.6) appears for all $l \in [\![2,n]\!], i \in [\![2,l-1]\!], k \in [\![l,n]\!]$ through the term $D_{\mathbf{x}_l}D_{\mathbf{x}_1}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y})$.

For the term into brackets, since $i \le l-1$, then $k>i$ and the only bracket term containing $\mathbf{y}_i$ is the one associated with $\mathbf{F}_i(\mathbf{y}_{1:l-1},\boldsymbol{\theta}_{s,t}^{l:n}(\boldsymbol{\xi}))$.

With the notations of (3.2), we write:

$$\begin{aligned}
&\mathbf{Lower}\Big[D_i\Psi_{i,(l,1),k}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}\Big]\\
=\;&\int_0^{(s-t)^{\rho_{i,k}}}dv v^{\frac{\alpha_i^k}{2}}\int_{\mathbb{R}^d}dz\Big|\int_{\mathbb{R}^d}d\mathbf{y}_i D_i\partial_v h_{cv}(z-\mathbf{y}_i)(\Psi_{i,(l,1),k}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}(\mathbf{y}_i)-\Psi_{i,(l,1),k}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}(z))\Big|\\
=:\;&\int_0^{(s-t)^{\rho_{i,k}}}dv v^{\frac{\alpha_i^k}{2}}\int_{\mathbb{R}^d}dz|(\mathscr{T}_{1,i,(l,1),k}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}+\mathscr{T}_{2,i,(l,1),k}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})})\big(v,z\big)|,
\end{aligned} \tag{3.7}$$

where:

$$\begin{aligned}
&\mathscr{T}_{1,i,(l,1),k}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}\big(v,z\big)\\
:=\;&\int_{\mathbb{R}^d}d\mathbf{y}_i D_i\partial_v h_v(z-\mathbf{y}_i)\Big(D_{\mathbf{x}_1}D_{\mathbf{x}_l}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y})\\
&\times\Big[\Big(\mathbf{F}_i(s,\mathbf{y}_{1:l-1},\boldsymbol{\theta}_{s,t}^{l:n}(\boldsymbol{\xi}))-\mathbf{F}_i(s,\mathbf{y}_{1:i-1},z,\mathbf{y}_{i+1:l-1},\boldsymbol{\theta}_{s,t}^{l:n}(\boldsymbol{\xi}))\Big)(\mathbf{y}-\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))_k^*\Big]\Big)_{\boldsymbol{\xi}=\mathbf{x}},
\end{aligned} \tag{3.8}$$

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

$$\begin{aligned}
&\mathscr{T}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}_{2,i,(l,1),k}(v,z) \qquad (3.9)\\
&:= \int_{\mathbb{R}^d} d\mathbf{y}_i D_i \partial_v h_v(z-\mathbf{y}_i)\Big\{\Big[D_{\mathbf{x}_1}D_{\mathbf{x}_l}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y}) - D_{\mathbf{x}_1}D_{\mathbf{x}_l}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y}_{1:i-1},z,\mathbf{y}_{i+1:n})\Big]\\
&\times\Big[\Big(\mathbf{F}_i(s,\mathbf{y}_{1:i-1},z,\mathbf{y}_{i+1:l-1},\boldsymbol{\theta}^{l:n}_{s,t}(\boldsymbol{\xi})) - \mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})) - D_{i-1}\mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))(\mathbf{y}-\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))_{i-1}\Big)(\mathbf{y}-\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))^*_k\Big]\Big\}_{\boldsymbol{\xi}=\mathbf{x}}.
\end{aligned}$$

Write now from (3.9), Proposition 5 and Lemma 9:

$$\begin{aligned}
&|\mathscr{T}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}_{2,i,(l,1),k}(v,z)|\\
&\le \bar{C}\int_{\mathbb{R}^d} d\mathbf{y}_i \frac{h_{cv}(z-\mathbf{y}_i)}{v^{\frac{3}{2}}}\int_0^1 d\lambda \frac{\hat{q}_c(t,s,\mathbf{x},\mathbf{y}_{1:i-1},z+\lambda(\mathbf{y}_i-z),\mathbf{y}_{i+1:n})}{(s-t)^{(i-\frac{1}{2})+(l-\frac{1}{2})+\frac{1}{2}}}|(\mathbf{y}-\boldsymbol{\theta}_{s,t}(\mathbf{x}))_k||\mathbf{y}_i-z|\\
&\times\Big(\Big|\mathbf{F}_i(s,\mathbf{y}_{1:i-1},z,\mathbf{y}_{i+1:l-1},\boldsymbol{\theta}^{l:n}_{s,t}(\mathbf{x})) - \mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\mathbf{x})) - D_{i-1}\mathbf{F}_i(s,\boldsymbol{\theta}_{s,t}(\mathbf{x}))(\mathbf{y}-\boldsymbol{\theta}_{s,t}(\mathbf{x}))_{i-1}\Big|\Big)\\
&\le \bar{C}\int_{\mathbb{R}^d} d\mathbf{y}_i \frac{h_{cv}(z-\mathbf{y}_i)}{v}\int_0^1 d\lambda \frac{\hat{q}_c(t,s,\mathbf{x},\mathbf{y}_{1:i-1},z+\lambda(\mathbf{y}_i-z),\mathbf{y}_{i+1:n})}{(s-t)^{(i-\frac{1}{2})+\frac{1}{2}}}\\
&\times\Big\{|z-(\boldsymbol{\theta}_{s,t}(\mathbf{x}))_i|^{\beta_i} + \sum_{j=i+1}^{l-1}(s-t)^{\beta_j(j-\frac{1}{2})} + (s-t)^{(1+\eta)(i-\frac{3}{2})}\Big\},
\end{aligned}$$

where for the second inequality, we used that for $k \ge l > i$, $|(\mathbf{y}-\boldsymbol{\theta}_{s,t}(\mathbf{x}))_k|(s-t)^{-(l-1/2)} \le |(\mathbf{y}-\boldsymbol{\theta}_{s,t}(\mathbf{x}))_k|(s-t)^{-(k-1/2)}$ which can be absorbed by the k^{th} variables of $\hat{q}_c$.

Writing now for any $\lambda \in [0,1]$,

$$|z-\boldsymbol{\theta}_{s,t}(\mathbf{x})_i| \le \lambda|z-\mathbf{y}_i| + |z+\lambda(\mathbf{y}_i-z)-(\boldsymbol{\theta}_{s,t}(\mathbf{x}))_i|,$$

we thus derive

$$\begin{aligned}
&|\mathscr{T}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}_{2,i,(l,1),k}(v,z)|\\
&\le \bar{C}\int_{\mathbb{R}^d} d\mathbf{y}_i \frac{h_{cv}(z-\mathbf{y}_i)}{v}\int_0^1 d\lambda \hat{q}_c(t,s,\mathbf{x},\mathbf{y}_{1:i-1},z+\lambda(\mathbf{y}_i-z),\mathbf{y}_{i+1:n})\\
&\times\Big(v^{\frac{\beta_i}{2}}(s-t)^{-(i-\frac{1}{2})-\frac{1}{2}} + (s-t)^{-(i-\frac{1}{2})-\frac{1}{2}+\beta_i(i-\frac{1}{2})} + \sum_{j=i+1}^{l-1}(s-t)^{-(i-\frac{1}{2})-\frac{1}{2}+\beta_j(j-\frac{1}{2})} + (s-t)^{-(i-\frac{1}{2})-\frac{1}{2}+(1+\eta)(i-\frac{3}{2})}\Big)\\
&\le \bar{C}\hat{q}_{c\backslash i}(t,s,\mathbf{x},\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n})\int_0^1 d\lambda\int_{\mathbb{R}^d} d\mathbf{y}_i h_{cv}(z-\mathbf{y}_i)\mathcal{N}_{c(s-t)^{2i-1}}(z+\lambda(\mathbf{y}_i-z)-(\boldsymbol{\theta}_{s,t}(\mathbf{x}))_i)\\
&\times v^{-1}\Big(v^{\frac{\beta_i}{2}}(s-t)^{-(i-\frac{1}{2})-\frac{1}{2}} + (s-t)^{-(i-\frac{1}{2})-\frac{1}{2}+(1+\eta)(i-\frac{3}{2})}\Big), \qquad (3.10)
\end{aligned}$$

recalling for the last inequality that for any j in $[\![i,l-1]\!]$, $\beta_j(j-1/2) > j-1 > i-3/2$ and η is supposed to be a small parameter.

Let us now consider the first contribution in the r.h.s. in (3.6),

$$\begin{aligned}
|\mathscr{T}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}_{1,i,(l,1),k}(v,z)| &\le \bar{C}\int_{\mathbb{R}^d} d\mathbf{y}_i \frac{h_{cv}(z-\mathbf{y}_i)}{v^{\frac{3}{2}}}\frac{\hat{q}_c(t,s,\mathbf{x},\mathbf{y})}{(s-t)^{(l-\frac{1}{2})+\frac{1}{2}}}|z-\mathbf{y}_i|^{\beta_i}|(\mathbf{y}-\boldsymbol{\theta}_{s,t}(\mathbf{x}))_k|\\
&\le \bar{C}\int_{\mathbb{R}^d} d\mathbf{y}_i \frac{h_{cv}(z-\mathbf{y}_i)}{v^{\frac{3}{2}-\frac{\beta_i}{2}}}\frac{\hat{q}_c(t,s,\mathbf{x},\mathbf{y})}{(s-t)^{\frac{1}{2}}}. \qquad (3.11)
\end{aligned}$$

From (3.10) and (3.11) we derive, with the notation introduced in (2.36):

$$\begin{aligned}
&\|\partial_v h_v \star D_i\Psi^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}_{i,(l,1),k}\|_{L^1(\mathbb{R}^d,\mathbb{R})}\\
\le\ & \bar{C}\hat{q}_{c\backslash i}(t,s,\mathbf{x},(\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}))\Big\{\frac{1}{v^{\frac{3}{2}-\frac{\beta_i}{2}}(s-t)^{\frac{1}{2}}} + \Big(\frac{v^{-1+\frac{\beta_i}{2}}}{(s-t)^i} + \frac{v^{-1}}{(s-t)^{i-(1+\eta)(i-\frac{3}{2})}}\Big)\\
&\times\int_0^1 d\lambda\int_{\mathbb{R}^d} dz\int_{\mathbb{R}^d} d\mathbf{y}_i h_{cv}(z-\mathbf{y}_i)\mathcal{N}_{c(s-t)^{2i-1}}(z+\lambda(\mathbf{y}_i-z)-(\boldsymbol{\theta}_{s,t}(\mathbf{x}))_i)\Big\}\\
\le\ & \bar{C}\hat{q}_{c\backslash i}(t,s,\mathbf{x},(\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}))\Big(\frac{1}{v^{\frac{3}{2}-\frac{\beta_i}{2}}(s-t)^{\frac{1}{2}}} + \frac{v^{-1+\frac{\beta_i}{2}}}{(s-t)^i} + \frac{v^{-1}}{(s-t)^{i-(1+\eta)(i-\frac{3}{2})}}\Big),
\end{aligned}$$

using the change of variable $(w_1, w_2) = (z - \mathbf{y}_i, z + \lambda(\mathbf{y}_i - z) - (\boldsymbol{\theta}_{s,t}(\mathbf{x}))_i)$ for the last inequality.

For $\rho_{i,k}$ chosen as in (3.5) one gets from the definition in (3.2):

$$\begin{aligned}
&\mathbf{Lower}\Big[D_i\Psi^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})}_{i,(l,1),k}\Big] \\
&\leq \bar{C}\hat{q}_{c\backslash i}(t,s,\mathbf{x},(\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}))\int_0^{(s-t)^{\rho_{i,k}}}\frac{dv}{v}v^{\frac{\alpha_i^k}{2}}\left(\frac{1}{v^{\frac{1-\beta_i}{2}}(s-t)^{\frac{1}{2}}}+\frac{v^{\frac{\beta_i}{2}}}{(s-t)^i}+\frac{1}{(s-t)^{i-(1+\eta)(i-\frac{3}{2})}}\right) \\
&=: \bar{C}\hat{q}_{c\backslash i}(t,s,\mathbf{x},(\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}))\mathcal{B}_{1,\rho_{i,k}}(t,s).
\end{aligned}$$

It therefore remains to prove that, if $\mathcal{B}_{1,\rho_{i,k}}(t,s) \leq (s-t)^{-1-1/2+\gamma_i^k}$ for $\gamma_i^k > 0$ then:

$$-1-\frac{1}{2}+\gamma_i^k > -1 \iff \gamma_i^k > \frac{1}{2}. \tag{3.12}$$

Write:

$$\begin{aligned}
\mathcal{B}_{1,\rho_{i,k}}(t,s) &\leq \int_0^{(s-t)^{\rho_{i,k}}} dv\left(\frac{v^{-\frac{3}{2}+\frac{\alpha_i^k+\beta_i}{2}}}{(s-t)^{\frac{1}{2}}}+\frac{v^{-1+\frac{\alpha_i^k+\beta_i}{2}}}{(s-t)^i}+\frac{v^{-1+\frac{\alpha_i^k}{2}}}{(s-t)^{\frac{3}{2}-\eta(i-\frac{3}{2})}}\right) \\
&\leq C\Big((s-t)^{\rho_{i,k}\frac{\alpha_i^k+\beta_i-1}{2}-\frac{1}{2}}+(s-t)^{\rho_{i,k}\frac{\alpha_i^k+\beta_i}{2}-i}+(s-t)^{\rho_{i,k}\frac{\alpha_i^k}{2}-\frac{3}{2}+\eta(i-\frac{3}{2})}\Big),
\end{aligned} \tag{3.13}$$

assuming for a while that α_i^k is s.t. $\alpha_i^k + \beta_i > 1$ to guarantee the integrability of the first above integrand (this condition will be fulfilled for the choice below). Let us check condition (3.12) is satisfied. Actually, the first two terms in (3.13) yield negligible contributions. Recall indeed from Remark 7 that the parameter α_i^k must be chosen so that $\alpha_i^k < \frac{1-(1-\beta_k)(k-1/2)}{i-\frac{1}{2}} := \bar{\alpha}_i^k$. Since $\beta_k \in (\frac{2k-2}{2k-1}, 1]$, $\bar{\alpha}_i^k = \frac{2-(1-\beta_k)(2k-1)}{2i-1} > \frac{3-2k+(2k-2)}{2i-1} = \frac{1}{2i-1}$. Thus,

$$\alpha_i^k = \frac{1+\frac{\eta}{4}}{2i-1}, \tag{3.14}$$

is an admissible choice (recall indeed that we chose $\eta < \inf_{j\in[\![2,n]\!]}\{\beta_j - \frac{2j-2}{2j-1}\}$ in assumption $(\mathbf{H}_\eta)$). It gives in particular that $\alpha_i^k + \beta_i > \frac{1+\eta/4+(2i-2)}{2i-1} = 1 + \frac{\eta/4}{2i-1}$. Hence,

$$\rho_{i,k}\frac{\alpha_i^k+\beta_i-1}{2}-\frac{1}{2} > -\frac{1}{2}$$

which already provides a regularizing term in time for the first term in the r.h.s. of (3.13). Now from (3.5), $\rho_{i,k} < \frac{(1+\eta)(2i-3)+1}{1-\alpha_i^k} =: \bar{\rho}_{i,k}$. The previous choice for α_i^k gives

$$\bar{\rho}_{i,k} = \frac{2i-2+\eta(2i-3)}{\frac{2i-1-(1+\frac{\eta}{4})}{2i-1}} = (2i-1)\frac{2i-2+\eta(2i-3)}{2i-2-\frac{\eta}{4}} > (2i-1),$$

and

$$\rho_{i,k} = 2i-1 \tag{3.15}$$

is an admissible choice. We therefore get for the exponent of the second term in (3.13):

$$\rho_{i,k}\frac{\alpha_i^k+\beta_i}{2}-i > \frac{1}{2}\Big(\rho_{i,k}(1+\frac{\frac{\eta}{4}}{2i-1})-2i\Big) > -\frac{1}{2}.$$

Eventually, for the exponent of the third contribution in (3.13), for the previous choice of $\rho_{i,k} = 2i-1$, we get

$$\rho_{i,k}\frac{\alpha_i^k}{2}-\frac{3}{2}+\eta(i-\frac{3}{2}) = \frac{1}{2}(2i-1)\frac{1+\frac{\eta}{4}}{2i-1}-\frac{3}{2}+\eta(i-\frac{3}{2}) = -1+\frac{\eta}{8}+\eta(i-\frac{3}{2}) > -1, \tag{3.16}$$

which means that criterion (3.12) is indeed satisfied, even though if this last contribution is rather critical in order to obtain the required smoothing effect with $\gamma_i^k = 1/2 + \eta(i-3/2)$. This concludes the proof of Lemma 10. □

3.2 Proof of Lemma 11

We now tackle the Besov estimate of the derivative of the solution of (1.18). Fix $t \in [0,T]$ and $(\mathbf{x}_{1:i-1}, \mathbf{x}_{i+1:n}) \in \mathbb{R}^{(n-1)d}$. From the thermic characterization of Besov spaces recalled in equation (3.1), we actually have to control:

$$\begin{aligned} \|D_k u(t,\mathbf{x}_{1,i-1},\cdot,\mathbf{x}_{i+1,n})\|_{B^{\alpha_i^k}_{\infty,\infty}} &= \|D_k u(t,\mathbf{x}_{1,i-1},\cdot,\mathbf{x}_{i+1,n})\|_{C^{\alpha_i^k}} \\ &:= \|D_k u(t,\mathbf{x}_{1:i-1},\cdot,\mathbf{x}_{i+1:n})\|_\infty + [D_k u(t,\mathbf{x}_{1,i-1},\cdot,\mathbf{x}_{i+1,n})]_{\alpha_i^k}. \quad (3.17) \end{aligned}$$

Observe first that $\|D_k u(t,\mathbf{x}_{1:i-1},\cdot,\mathbf{x}_{i+1:n})\|_\infty \le \|\mathbf{D}u\|_\infty$. Hence, we can focus on the Hölder modulus in (3.17). For given $\mathbf{x}_i,\ z \in \mathbb{R}^d$ we want to control the difference:

$$D_{\mathbf{x}_k} u(t,\mathbf{x}_{1:i-1},\mathbf{x}_i,\mathbf{x}_{i+1:n}) - D_{\mathbf{x}_k} u(t,\mathbf{x}_{1:i-1},z,\mathbf{x}_{i+1:n}) \quad (3.18)$$

through the expansion of the gradients given by (2.21). Two cases then arise: the system is globally in the *off-diagonal* regime, i.e. the spatial points $\mathbf{x}_i$ and z are *far* w.r.t. their corresponding time scale (there exists c_0 such that $c_0|\mathbf{x}_i - z| \ge (T-t)^{i-1/2}$ or equivalently $c_0|\mathbf{x}_i - z|^{2/(2i-1)} \ge (T-t)$); the system is globally in the *diagonal* regime, i.e. the spatial points $\mathbf{x}_i$ and z are *close* w.r.t. their corresponding time scale ($c_0|\mathbf{x}_i - z|^{2/(2i-1)} < (T-t)$).

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 (2.21) taking as freezing point the associated spatial argument, i.e. $\boldsymbol{\xi} = \mathbf{x}$ for the first gradient and denoting by $\mathbf{x}' = (\mathbf{x}_{1:i-1}, z, \mathbf{x}_{i+1:n})$, $\boldsymbol{\xi}' = \mathbf{x}'$ for the second one. This allows to take advantage of the underlying smoothing properties in time of the gradient (cf. Section 3.2.1 below[8]).

On the other hand, in the *global diagonal* regime (when $c_0|\mathbf{x}_i - z|^{2/(2i-1)} \le (T-t)$), we are again faced with a regime dichotomy. Note indeed that, expanding the gradients in (3.18) from (2.21), we have to deal with a time integral associated with the source and perturbative terms. We are hence again led to separate, within this time integral, the *local* off-diagonal and diagonal regimes, i.e. w.r.t. the time integration variable s. We hence introduce the time set $\mathcal{S}_i = \{s \in [t,T] : (s-t) \le c_0|\mathbf{x}_i - z|^{2/(2i-1)}\}$ (for the same previous constant c_0) which corresponds to the *local* off-diagonal regime and the complementary set $\mathcal{S}_i^c = \{s \in [t,T] : (s-t) > c_0|\mathbf{x}_i - z|^{2/(2i-1)}\}$ which corresponds to the *local* diagonal one.

As above, in the *local off-diagonal* regime, we will not expand the difference of the gradients and we will only use their underlying smoothing properties in time, working thus with their expansion around two different freezing points associated with the corresponding spatial arguments, as suggested by (2.21) (see again Section 3.2.1).

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 (2.21), 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 Section 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 (3.18) around two different freezing points. Such a strategy was already used in the companion paper [CdRHM20] and leads to consider an additional boundary term arising precisely from the change of freezing point (see Section 3.2.3). Namely, we will expand the term $D_{\mathbf{x}_k} u(t,\mathbf{x}_{1:i-1},\mathbf{x}_i,\mathbf{x}_{i+1:n})$ with (2.21) taking $\boldsymbol{\xi} = \mathbf{x}$, whereas we will expand differently the contribution $D_{\mathbf{x}_k} u(t,\mathbf{x}_{1:i-1},z,\mathbf{x}_{i+1:n})$, depending on the considered (local) regime (off-diagonal or diagonal) for the current running time. With the previous notations, this term will be expanded as in (2.21) around the freezing point $\boldsymbol{\xi}' = \mathbf{x}'$ in the local off-diagonal regime and around $\tilde{\boldsymbol{\xi}}' = \mathbf{x}$ in the local diagonal one. Denoting by $t_0 = t + c_0|\mathbf{x}_i - z|^{2/(2i-1)}$ the transition time between the two regimes, we actually have:

$$\begin{aligned} u(t,\mathbf{x}') &= \int_t^T ds \Big(\mathbb{I}_{\mathcal{S}_i} \Big[\tilde{P}^{\boldsymbol{\xi}'}_{s,t} f(s,\cdot) \Big](\mathbf{x}') + \mathbb{I}_{\mathcal{S}_i^c} \Big[\tilde{P}^{\tilde{\boldsymbol{\xi}}'}_{s,t} f(s,\cdot) \Big](\mathbf{x}') \Big) + \Big[\tilde{P}^{\boldsymbol{\xi}'}_{t_0,t} u(t_0,\cdot) \Big](\mathbf{x}') - \Big[\tilde{P}^{\tilde{\boldsymbol{\xi}}'}_{t_0,t} u(t_0,\cdot) \Big](\mathbf{x}') \\ &\quad + \int_t^T ds \Big(\mathbb{I}_{\mathcal{S}_i} \Big[\tilde{P}^{\boldsymbol{\xi}'}_{s,t} \Big((L_s - \tilde{L}^{\boldsymbol{\xi}'}_s) u \Big)(s,\cdot) \Big](\mathbf{x}') + \mathbb{I}_{\mathcal{S}_i^c} \Big[\tilde{P}^{\tilde{\boldsymbol{\xi}}'}_{s,t} \Big((L_s - \tilde{L}^{\tilde{\boldsymbol{\xi}}'}_s) u \Big)(s,\cdot) \Big](\mathbf{x}') \Big). \quad (3.19) \end{aligned}$$

We refer to Appendix B below for a proof of expansion (3.19) (see also Section 2.4 of the "Detailed guide to the proof" in [CdRHM20]). We again emphasize that in comparison with (2.21), the term in the second line

[8] note that in this case we have that for any s in $(t,T]$ the off-diagonal regime holds.

in the r.h.s. of the above equation is 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 Section 3.2.3, Lemma 14 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 Section 3.2.1 applies.

Starting from (3.18), we expand $D_{\mathbf{x}_k} u(t,\mathbf{x}) = D_{\mathbf{x}_k} u(t,\mathbf{x}_{1:i-1},\mathbf{x}_i,\mathbf{x}_{i+1:n})$ using (2.21) with $\boldsymbol{\xi} = \mathbf{x}$ and $D_{\mathbf{x}_k} u(t,\mathbf{x}_{1:i-1},z,\mathbf{x}_{i+1:n})$ differentiating (3.19) w.r.t. $\mathbf{x}_k$ and setting then $\boldsymbol{\xi}' = \mathbf{x}', \tilde{\boldsymbol{\xi}}' = \mathbf{x}$. We rewrite:

$$\begin{aligned} D_{\mathbf{x}_k} u(t,\mathbf{x}) - D_{\mathbf{x}_k} u(t,\mathbf{x}') \quad =: \quad & \big(D_{\mathbf{x}_k} u(t,\mathbf{x}) - D_{\mathbf{x}_k} u(t,\mathbf{x}')\big)|_{\mathcal{S}_i} + \big(D_{\mathbf{x}_k} u(t,\mathbf{x}) - D_{\mathbf{x}_k} u(t,\mathbf{x}')\big)|_{\mathcal{S}_i^c} \\ & + \big(D_{\mathbf{x}_k} u(t,\mathbf{x}) - D_{\mathbf{x}_k} u(t,\mathbf{x}')\big)|_{\partial\mathcal{S}_i}, \end{aligned} \tag{3.20}$$

where, with the notation of (2.21):

$$\big(D_{\mathbf{x}_k} u(t,\mathbf{x}) - D_{\mathbf{x}_k} u(t,\mathbf{x}')\big)|_{\mathcal{S}_i} := \int_{\mathcal{S}_i} ds \Big\{ \Big[H_k^{\boldsymbol{\xi}}(s,\mathbf{x}) - H_k^{\boldsymbol{\xi}'}(s,\mathbf{x}')\Big] + \Big[I_k^{\boldsymbol{\xi}}(s,\mathbf{x}) - I_k^{\boldsymbol{\xi}'}(s,\mathbf{x}')\Big] \Big\}_{(\boldsymbol{\xi},\boldsymbol{\xi}')=(\mathbf{x},\mathbf{x}')}, \tag{3.21}$$

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

$$\big(D_{\mathbf{x}_k} u(t,\mathbf{x}) - D_{\mathbf{x}_k} u(t,\mathbf{x}')\big)|_{\mathcal{S}_i^c} := \int_{\mathcal{S}_i^c} ds \Big\{ \Big[H_k^{\boldsymbol{\xi}}(s,\mathbf{x}) - H_k^{\tilde{\boldsymbol{\xi}}'}(s,\mathbf{x}')\Big] + \Big[I_k^{\boldsymbol{\xi}}(s,\mathbf{x}) - I_k^{\tilde{\boldsymbol{\xi}}'}(s,\mathbf{x}')\Big] \Big\}_{(\boldsymbol{\xi},\tilde{\boldsymbol{\xi}}')=(\mathbf{x},\mathbf{x})}, \tag{3.22}$$

is the contribution of the diagonal regime and

$$\big(D_{\mathbf{x}_k} u(t,\mathbf{x}) - D_{\mathbf{x}_k} u(t,\mathbf{x}')\big)|_{\partial\mathcal{S}_i} := \Big\{ D_{\mathbf{x}_k} \tilde{P}^{\boldsymbol{\xi}'}_{t_0,t} u(t_0,\mathbf{x}') - D_{\mathbf{x}_k} \tilde{P}^{\tilde{\boldsymbol{\xi}}'}_{t_0,t} u(t_0,\mathbf{x}') \Big\}_{(\boldsymbol{\xi}',\tilde{\boldsymbol{\xi}}')=(\mathbf{x}',\mathbf{x})}, \tag{3.23}$$

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 (3.21).

On the time set $\mathcal{S}_i$, we cannot expect some regularization from the difference of the transition densities so that we bluntly estimate the terms appearing in (3.21), writing:

$$|D_{\mathbf{x}_k} u(t,\mathbf{x}) - D_{\mathbf{x}_k} u(t,\mathbf{x}')|\Big|_{\mathcal{S}_i} \leq \int_{\mathcal{S}_i} ds \bigg(\big|H_k^{\boldsymbol{\xi}}(s,\mathbf{x})\big|\Big|_{\boldsymbol{\xi}=\mathbf{x}} + \big|H_k^{\boldsymbol{\xi}'}(s,\mathbf{x}')\big|\Big|_{\boldsymbol{\xi}'=\mathbf{x}'} + \big|I_k^{\boldsymbol{\xi}}(s,\mathbf{x})\big|\Big|_{\boldsymbol{\xi}=\mathbf{x}} + \big|I_k^{\boldsymbol{\xi}'}(s,\mathbf{x}')\big|\Big|_{\boldsymbol{\xi}'=\mathbf{x}'} \bigg). \tag{3.24}$$

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. $\mathbf{x}_1$. This improves the exponents of the time singularities of $1/2$. Similarly to (2.22), (2.24), (2.26) and (2.27) this therefore yields for the terms $H_k^{\boldsymbol{\xi}}(s,\mathbf{x})$:

$$\Big|H_k^{\boldsymbol{\xi}}(s,\mathbf{x})\Big|\Big|_{\boldsymbol{\xi}=\mathbf{x}} \leq C(s-t)^{-1+\delta_i^k} \big(\|\mathbf{D}u\|_\infty + \|\mathbf{D}D_1 u\|_\infty\big), \tag{3.25}$$

with $1 > \delta_i^k \geq 1/2 + \eta/2 > 1/2$.

Reproducing the arguments that led to equations (2.30), (2.31), (2.36)-(2.39) and the statement of Lemma 10, exploiting again that there is now no differentiation w.r.t. $\mathbf{x}_1$ we get with the notations of (1.2):

$$\Big|I_k^{\boldsymbol{\xi}}(s,\mathbf{x})\Big|\Big|_{\boldsymbol{\xi}=\mathbf{x}} \leq C(s-t)^{-1+\gamma_i^k} \big(\|\mathbf{D}u\|_\infty + \sup_{\bar{s}\in[0,T]} [(D_k u)_i(\bar{s},\cdot)]_{\alpha_i^k}\big). \tag{3.26}$$

Similar bounds hold for $\Big|H_k^{\boldsymbol{\xi}'}(s,\mathbf{x}')\Big|\Big|_{\boldsymbol{\xi}'=\mathbf{x}'}$ and $\Big|I_k^{\boldsymbol{\xi}'}(s,\mathbf{x}')\Big|\Big|_{\boldsymbol{\xi}'=\mathbf{x}'}$. Hence, from (3.25) and (3.26),

$$
\begin{aligned}
&\int_{\mathcal{S}_i} ds\bigg(\Big|H_k^{\boldsymbol{\xi}}(s,\mathbf{x})\Big|\Big|_{\boldsymbol{\xi}=\mathbf{x}} + \Big|H_k^{\boldsymbol{\xi}'}(s,\mathbf{x}')\Big|\Big|_{\boldsymbol{\xi}'=\mathbf{x}'} + \Big|I_k^{\boldsymbol{\xi}}(s,\mathbf{x})\Big|\Big|_{\boldsymbol{\xi}=\mathbf{x}} + \Big|I_k^{\boldsymbol{\xi}'}(s,\mathbf{x}')\Big|\Big|_{\boldsymbol{\xi}'=\mathbf{x}'}\bigg)\\
&\le C\Big(\int_t^{(t+c_0|\mathbf{x}_i-z|^{\frac{2}{2i-1}})\wedge T} ds\Big((s-t)^{-1+\delta_i^k}\big(\|\mathbf{D}u\|_\infty+\|\mathbf{D}D_1u\|_\infty\big)+(s-t)^{-1+\gamma_i^k}\big(\|\mathbf{D}u\|_\infty+\sup_{\bar s\in[0,T]}[\big(D_ku\big)_i(\bar s,\cdot)]_{\alpha_i^k}\big)\Big)\\
&\le C\Big(|\mathbf{x}_i-z|^{\frac{2\delta_i^k}{2i-1}}\wedge(T-t)^{\delta_i^k}\big(\|\mathbf{D}u\|_\infty+\|\mathbf{D}D_1u\|_\infty\big)+|\mathbf{x}_i-z|^{\frac{2\gamma_i^k}{2i-1}}\wedge(T-t)^{\gamma_i^k}\big(\|\mathbf{D}u\|_\infty+\sup_{\bar s\in[0,T]}[\big(D_ku\big)_i(\bar s,\cdot)]_{\alpha_i^k}\big)\Big)\\
&\le C|\mathbf{x}_i-z|^{\alpha_i^k}T^{\delta'}\Big(\|\mathbf{D}u\|_\infty+\|\mathbf{D}D_1u\|_\infty+\sup_{\bar s\in[0,T]}[\big(D_ku\big)_i(\bar s,\cdot)]_{\alpha_i^k}\Big),
\end{aligned}
\tag{3.27}
$$

for some $\delta':=\delta'((\mathbf{A}))>0$, recalling for the last inequality that $\gamma_i^k=1/2+\eta(i-3/2)$ so that $2\gamma_i^k/(2i-1)=[1+2\eta(i-3/2)]/(2i-1)>(1+\eta/4)/(2i-1)=\alpha_i^k$ (see also the statements of Lemmas 10 and 11) and similarly for the contributions involving $\delta_i^k>1/2$. We eventually get from (3.27) and (3.24):

$$
|D_{\mathbf{x}_k}u(t,\mathbf{x})-D_{\mathbf{x}_k}u(t,\mathbf{x}')|\Big|_{\mathcal{S}_i}\le CT^{\delta'}\Big(\|\mathbf{D}u\|_\infty+\|\mathbf{D}D_1u\|_\infty+\sup_{\bar s\in[0,T]}[\big(D_ku\big)_i(\bar s,\cdot)]_{\alpha_i^k}\Big).
\tag{3.28}
$$

3.2.2 Diagonal estimates: control of the term (3.22).

We consider here the difference $D_{\mathbf{x}_k}u(t,\mathbf{x})-D_{\mathbf{x}_k}u(t,\mathbf{x}')$ of (3.18) on $\mathcal{S}_i^c$. In that case the points $\mathbf{x}_i$ and $\mathbf{x}_i'$ 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 $\boldsymbol{\xi}=\tilde{\boldsymbol{\xi}}'=\mathbf{x}$ with starting points $\mathbf{x}$ and $\mathbf{x}'$ respectively. Precisely, recalling that $\mathbf{x}$ and $\mathbf{x}'$ only differ in the i^{th} component, we can write:

$$
D_{\mathbf{x}_k}\tilde p^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y})-D_{\mathbf{x}_k}\tilde p^{\boldsymbol{\xi}}(t,s,\mathbf{x}',\mathbf{y})=-\int_0^1 d\lambda D_{\mathbf{x}_i}\Big(D_{\mathbf{x}_k}\tilde p^{\boldsymbol{\xi}}(t,s,\mathbf{x}+\lambda(\mathbf{x}'-\mathbf{x}),\mathbf{y})\Big)\cdot(\mathbf{x}'-\mathbf{x})_i.
\tag{3.29}
$$

From Lemma 8 we thus derive:

$$
|D_{\mathbf{x}_k}\tilde p^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y})-D_{\mathbf{x}_k}\tilde p^{\boldsymbol{\xi}}(t,s,\mathbf{x}',\mathbf{y})|\le\frac{C|(\mathbf{x}'-\mathbf{x})_i|}{(s-t)^{(i-\frac12)+(k-\frac12)}}\int_0^1 d\lambda\hat p^{\boldsymbol{\xi}}_{C^{-1}}(t,s,\mathbf{x}+\lambda(\mathbf{x}'-\mathbf{x}),\mathbf{y}).
\tag{3.30}
$$

Now, from the definition of $\hat p_{C^{-1}}$ in Proposition 5, recalling as well from (2.7) that $\mathbf{x}\mapsto\mathbf{m}_{s,t}^{\boldsymbol{\xi}}(\mathbf{x}):=\mathbf{m}_{s,t}^{(t,\boldsymbol{\xi})}(\mathbf{x})$ is affine, we get:

$$
\begin{aligned}
&|\hat p^{\boldsymbol{\xi}}_{C^{-1}}(t,s,\mathbf{x}+\lambda(\mathbf{x}'-\mathbf{x}),\mathbf{y})|\\
\le\;&\frac{C}{(s-t)^{\frac{n^2d}{2}}}\exp(-c(s-t)|\mathbb{T}^{-1}_{s-t}(\mathbf{m}^{\boldsymbol{\xi}}_{s,t}(\mathbf{x}+\lambda(\mathbf{x}'-\mathbf{x}))-\mathbf{y})|^2)\\
\le\;&\frac{C}{(s-t)^{\frac{n^2d}{2}}}\exp(c(s-t)|\mathbb{T}^{-1}_{s-t}\mathbf{R}^{(t,\boldsymbol{\xi})}(s,t)(\mathbf{x}-\mathbf{x}')|^2)\exp(-\frac{c}{2}(s-t)|\mathbb{T}^{-1}_{s-t}(\mathbf{m}^{\boldsymbol{\xi}}_{s,t}(\mathbf{x})-\mathbf{y})|^2).
\end{aligned}
$$

Using the rescaling arguments of the proof of Proposition 5 on the resolvent (see equation (2.13)), we then get $(s-t)^{1/2}|\mathbb{T}^{-1}_{s-t}\mathbf{R}^{(t,\boldsymbol{\xi})}(s,t)(\mathbf{x}-\mathbf{x}')|\le C(s-t)^{1/2}|\mathbb{T}^{-1}_{s-t}(\mathbf{x}-\mathbf{x}')|=C(s-t)^{-i+1/2}|(\mathbf{x}'-\mathbf{x})_i|\le C$, from the very definition of $\mathcal{S}_i^c$. Hence,

$$
|\hat p^{\boldsymbol{\xi}}_{C^{-1}}(t,s,\mathbf{x}+\lambda(\mathbf{x}'-\mathbf{x}),\mathbf{y})|\le\frac{C}{(s-t)^{\frac{n^2d}{2}}}\exp(-\frac{c}{2}(s-t)|\mathbb{T}^{-1}_{s-t}(\mathbf{m}^{\boldsymbol{\xi}}_{s,t}(\mathbf{x})-\mathbf{y})|^2),
$$

so that, from (3.30) and recalling that on $\mathcal{S}_i^c$, $|(\mathbf{x}'-\mathbf{x})_i|/(s-t)\le C(|(\mathbf{x}'-\mathbf{x})_i|/(s-t))^{\alpha_i^k}$, the following important control holds:

$$
|D_{\mathbf{x}_k}\tilde p^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y})-D_{\mathbf{x}_k}\tilde p^{\boldsymbol{\xi}}(t,s,\mathbf{x}',\mathbf{y})|\le\frac{C|(\mathbf{x}'-\mathbf{x})_i|^{\alpha_i^k}}{(s-t)^{\alpha_i^k(i-\frac12)+(k-\frac12)}}\hat p^{\boldsymbol{\xi}}_{C^{-1}}(t,s,\mathbf{x},\mathbf{y}).
\tag{3.31}
$$

Write now from (3.22), recalling that $\tilde{\boldsymbol{\xi}}' = \boldsymbol{\xi}$,

$$\big|D_{\mathbf{x}_k}u(t,\mathbf{x}) - D_{\mathbf{x}_k}u(t,\mathbf{x}')\big|\Big|_{\mathcal{S}_i^c} := \left|\Big(\int_{\mathcal{S}_i^c} ds\left[H_k^{\boldsymbol{\xi}}(s,\mathbf{x}) - H_k^{\boldsymbol{\xi}}(s,\mathbf{x}')\right] + \int_{\mathcal{S}_i^c} ds\left[I_k^{\boldsymbol{\xi}}(s,\mathbf{x}) - I_k^{\boldsymbol{\xi}}(s,\mathbf{x}')\right]\Big)_{\boldsymbol{\xi}=\mathbf{x}}\right|, \quad (3.32)$$

and let us discuss how the terms $H_k^{\boldsymbol{\xi}}(s,\mathbf{x}) - H_k^{\boldsymbol{\xi}}(s,\mathbf{x}')$, $I_k^{\boldsymbol{\xi}}(s,\mathbf{x}) - I_k^{\boldsymbol{\xi}}(s,\mathbf{x}')$ in the above equation can be handled.

We first focus on the term $I_k^{\boldsymbol{\xi}}(s,\mathbf{x}) - I_k^{\boldsymbol{\xi}}(s,\mathbf{x}')$ in (3.32). This contribution, associated with the degenerate components of perturbed operator, is again the most delicate to handle. From the definitions in (2.29) we are led to control the sum $\sum_{\ell=1}^3 [I_{k,\ell}^{\boldsymbol{\xi}}(s,\mathbf{x}) - I_{k,\ell}^{\boldsymbol{\xi}}(s,\mathbf{x}')]$. For the terms $I_{k,1}^{\boldsymbol{\xi}}(s,\mathbf{x}) - I_{k,1}^{\boldsymbol{\xi}}(s,\mathbf{x}')$, $I_{k,2}^{\boldsymbol{\xi}}(s,\mathbf{x}) - I_{k,2}^{\boldsymbol{\xi}}(s,\mathbf{x}')$ we are going to reproduce the analysis leading to (2.30), (2.31). Observe first that the above terms do not involve $D_{\mathbf{x}_1}$, therefore we gain a singularity of order $1/2$ w.r.t. the indicated equations (2.30), (2.31). On the other hand, the difference of the derivatives of the frozen densities w.r.t. $\mathbf{x}_k$ can be handled with (3.31). This leads to:

$$\begin{aligned}
&\left|\sum_{\ell=1}^2 \int_{t+c_0|(\mathbf{x}'-\mathbf{x})_i|^{\frac{2}{2i-1}}}^T ds\Big(I_{k,\ell}^{\boldsymbol{\xi}}(s,\mathbf{x}) - I_{k,\ell}^{\boldsymbol{\xi}}(s,\mathbf{x}')\Big)\right|_{\boldsymbol{\xi}=\mathbf{x}} \\
&\le C|(\mathbf{x}'-\mathbf{x})_i|^{\alpha_i^k}\|\mathbf{D}u\|_\infty \sum_{j=k}^n \int_t^T ds(s-t)^{-(k-\frac12)-\alpha_i^k(i-\frac12)}\Big((s-t)^{\beta_j(j-\frac12)} + (s-t)^{(1+\eta)(j-\frac12)}\Big),
\end{aligned}$$

changing the summation variables from (2.30) for notational simplicity.

From the very definition of $\alpha_i^k = (1+\eta/4)/(2i-1)$ in Lemma 10 and the specific choice of η in assumption **(A)** (see $(\mathbf{H}_\eta)$), which yields that $\beta_j(j-1/2) > j-1+\eta(j-1/2))$, we derive

$$\left|\sum_{\ell=1}^2 \int_{t+c_0|(\mathbf{x}'-\mathbf{x})_i|^{\frac{2}{2i-1}}}^T ds\Big(I_{k,\ell}^{\boldsymbol{\xi}}(s,\mathbf{x}) - I_{k,\ell}^{\boldsymbol{\xi}}(s,\mathbf{x}')\Big)\right|_{\boldsymbol{\xi}=\mathbf{x}} \le C|(\mathbf{x}'-\mathbf{x})_i|^{\alpha_i^k}\|\mathbf{D}u\|_\infty T^\delta, \quad (3.33)$$

for some $\delta > 0$.

From the previous analysis it is therefore sufficient to focus on the tricky term, namely $I_{k,3}(s,\mathbf{x})$ introduced in (2.29). We begin the proof considering first $I_{k,3}(s,\mathbf{x})$. Exploiting as well Lemma 8 for a centering argument w.r.t. the k^{th} variable, we write:

$$\begin{aligned}
&I_{k,3}^{\boldsymbol{\xi}}(s,\mathbf{x}) - I_{k,3}^{\boldsymbol{\xi}}(s,\mathbf{x}') \\
&= \sum_{\ell=2}^k \int_{\mathbb{R}^{nd}} d\mathbf{y}\Big\langle\Big(\mathbf{F}_\ell(s,\mathbf{y}_{1:k-1},\boldsymbol{\theta}_{s,t}^{k:n}(\boldsymbol{\xi})) - \mathbf{F}_\ell(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})) - D_{\ell-1}\mathbf{F}_\ell(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))\big(\mathbf{y}-\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})\big)_{\ell-1}\Big), \\
&\Big(D_{\mathbf{y}_\ell}u(s,\mathbf{y}) - D_{\mathbf{y}_\ell}u(s,\mathbf{y}_{1:k-1},\boldsymbol{\theta}_{s,t}^{k:n}(\boldsymbol{\xi}))\Big)\Big\rangle\big(D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y}) - D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x}',\mathbf{y})\big).
\end{aligned}$$

Let us reproduce now the arguments used in Section 2.2 to handle $I_{k,3}^{\boldsymbol{\xi}}$ (see e.g. the computations from equation (2.32) to (2.33)). For $\ell \in [\![2,k-1]\!]$, expanding with the Taylor formula the difference $\Big(D_{\mathbf{y}_\ell}u(s,\mathbf{y}) - D_{\mathbf{y}_\ell}u(s,\mathbf{y}_{1:k-1},\boldsymbol{\theta}_{s,t}^{k:n}(\boldsymbol{\xi}))\Big)$, using the Schwarz theorem to exchange the order of differentiations[9], and integrating by parts as well, we obtain:

$$\begin{aligned}
&I_{k,3}^{\boldsymbol{\xi}}(s,\mathbf{x}) - I_{k,3}^{\boldsymbol{\xi}}(s,\mathbf{x}') \\
&= \sum_{\ell=2}^{k-1}\sum_{m=k}^n \int_0^1 d\lambda \int_{\mathbb{R}^{nd}} d\mathbf{y} D_{\mathbf{y}_\ell}\Big[\Big(\mathbf{F}_\ell(s,\mathbf{y}_{1:k-1},\boldsymbol{\theta}_{s,t}^{k:n}(\boldsymbol{\xi})) - \mathbf{F}_\ell(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})) - D_{\ell-1}\mathbf{F}_\ell(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))\big(\mathbf{y}-\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})\big)_{\ell-1}\Big) \\
&\big(D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y}) - D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x}',\mathbf{y})\big)(\mathbf{y}-\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))_m\Big] D_{\mathbf{y}_m}u\big(s,\mathbf{y}_{1:k-1},\boldsymbol{\theta}_{s,t}^{k:n}(\boldsymbol{\xi}) + \lambda(\mathbf{y}-\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))_{k:n}\big) \\
&+ \int_{\mathbb{R}^{nd}} d\mathbf{y}\Big[\Big(\mathbf{F}_k(s,\mathbf{y}_{1:k-1},\boldsymbol{\theta}_{s,t}^{k:n}(\boldsymbol{\xi})) - \mathbf{F}_k(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})) - D_{k-1}\mathbf{F}_\ell(s,\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))\big(\mathbf{y}-\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})\big)_{k-1}\Big) \\
&\big(D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y}) - D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x}',\mathbf{y})\big)\Big]\Big(D_{\mathbf{y}_k}u(s,\mathbf{y}) - D_{\mathbf{y}_k}u(s,\mathbf{y}_{1:k-1},\boldsymbol{\theta}_{s,t}^{k:n}(\boldsymbol{\xi}))\Big) \\
&=: \Delta_{k,31}^{\boldsymbol{\xi}}(s,\mathbf{x},\mathbf{x}') + \Delta_{k,32}^{\boldsymbol{\xi}}(s,\mathbf{x},\mathbf{x}'). \quad (3.34)
\end{aligned}$$

[9] Recall indeed that what we are able to control is precisely the Hölder moduli of the derivatives $D_{\mathbf{y}_m}u(s,\cdot)$ w.r.t. the variables $\ell \le m$.

Let us now modify slightly the definition of (2.34) and introduce: for all $\ell \in [\![2, k-1]\!]$, $m \in [\![k, n]\!]$, $(\mathbf{y}_{1:\ell-1}, \mathbf{y}_{\ell+1:n}) \in \mathbb{R}^{(n-1)d}$, $(t, \mathbf{z}, \mathbf{x}) \in [0,T] \times (\mathbb{R}^{nd})^2$, $s \in (t, T]$ and $\mathbf{y}_i \in \mathbb{R}^d$:

$$\begin{aligned}
&\Psi_{\ell,k,m}^{(s,\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}),(t,\mathbf{z},\mathbf{x})}(\mathbf{y}_\ell) \\
= \;& \Big[D_{\mathbf{x}_k} \tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{z},\mathbf{y}) \qquad (3.35) \\
& \otimes \Big(\big(\mathbf{F}_\ell(s, \mathbf{y}_{1:k-1}, \boldsymbol{\theta}_{s,t}^{k:n}(\boldsymbol{\xi})) - \mathbf{F}_\ell(s, \boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})) - D_{\ell-1}\mathbf{F}_\ell(s, \boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))\big(\mathbf{y} - \boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})\big)_{\ell-1}\big)\Big)\big((\mathbf{y} - \boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))_m\big)^* \Big]_{\boldsymbol{\xi}=\mathbf{x}}.
\end{aligned}$$

Write now with the notations of (3.36):

$$\begin{aligned}
&(\Delta_{k,31}^{\boldsymbol{\xi}}(s,\mathbf{x},\mathbf{x}'))_{\boldsymbol{\xi}=\mathbf{x}} \\
:= \;& \sum_{\ell=2}^{k-1}\sum_{m=k}^{n} \int_0^1 d\lambda \int_{\mathbb{R}^{nd}} d\mathbf{y} D_{\mathbf{y}_\ell}\Big[\Psi_{\ell,k,m}^{(s,\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}),(t,\mathbf{x},\mathbf{x})}(\mathbf{y}_\ell) - \Psi_{\ell,k,m}^{(s,\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}),(t,\mathbf{x}',\mathbf{x})}(\mathbf{y}_\ell)\Big] \\
& D_{\mathbf{y}_m} u\big(s, \mathbf{y}_{1:k-1}, \boldsymbol{\theta}_{s,t}^{k:n}(\mathbf{x}) + \lambda(\mathbf{y} - \boldsymbol{\theta}_{s,t}(\mathbf{x}))_{k:n}\big).
\end{aligned}$$

Thus, we derive similarly to (2.36):

$$\begin{aligned}
&|\Delta_{k,31}^{\boldsymbol{\xi}}(s,\mathbf{x},\mathbf{x}')|_{\boldsymbol{\xi}=\mathbf{x}} \\
\le \;& \sum_{\ell=2}^{k-1}\sum_{m=k}^{n} \int_{\mathbb{R}^{(n-1)d}} d(\mathbf{y}_{1:\ell-1}, \mathbf{y}_{\ell+1:n}) \Big\{ \Big\| D_\ell\Big[\Psi_{\ell,k,m}^{(s,\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}),(t,\mathbf{x},\mathbf{x})}(\cdot) - \Psi_{\ell,k,m}^{(s,\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}),(t,\mathbf{x}',\mathbf{x})}(\cdot)\Big]\Big\|_{B_{1,1}^{-\alpha_\ell^m}} \\
& \qquad \times \sup_{\mathbf{z}_j, j\in[\![1,n]\!], j\neq\ell} \Big\| D_m u(s, \mathbf{z}_{1:\ell-1}, \cdot, \mathbf{z}_{\ell+1:n})\Big\|_{B_{\infty,\infty}^{\alpha_\ell^m}} \Big\}, \qquad (3.36)
\end{aligned}$$

where $\int_{\mathbb{R}^{(n-1)d}} d(\mathbf{y}_{1:\ell-1}, \mathbf{y}_{\ell+1:n})$ means that we integrate over $\mathbf{y}_{1:\ell-1}$ and $\mathbf{y}_{\ell+1:n}$. 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 13. *Let $k \in [\![2,n]\!]$, $\ell \in [\![2,k-1]\!]$ and $m \in [\![k,n]\!]$ and let $\Psi_{\ell,k,m}^{(s,\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}),(t,\mathbf{z},\mathbf{x})} : \mathbb{R}^d \to \mathbb{R}^d$ be the function defined by* (3.36). *There exist $C := C((\mathbf{A})) > 0$ and $\gamma_\ell^m := \gamma_\ell^m((\mathbf{A})) := 1/2 + \eta(\ell - 3/2) > 1/2$ such that*

$$\begin{aligned}
&\Big\| D_\ell\Big[\Psi_{\ell,k,m}^{(s,\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}),(t,\mathbf{x},\mathbf{x})}(\cdot) - \Psi_{\ell,k,m}^{(s,\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}),(t,\mathbf{x}',\mathbf{x})}(\cdot)\Big]\Big\|_{B_{1,1}^{-\alpha_\ell^m}} \\
\le \;& C\hat{q}_{c\backslash\ell}(t,s,\mathbf{x},(\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}))(s-t)^{-1-(i-\frac{1}{2})\alpha_i^k+\gamma_\ell^m}|(\mathbf{x}-\mathbf{x}')_i|^{\alpha_i^k},
\end{aligned}$$

with $\hat{q}_{c\backslash\ell}(t,s,\mathbf{x},(\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}))$ as in (2.36).

Again, for the specific choice of $\alpha_i^k = (1+\eta/4)/(2i-1)$ performed in the proof of Lemma 10, we eventually derive from Lemma 13 and (3.36) that:

$$\begin{aligned}
\int_{\mathcal{S}_i^c} ds|\Delta_{k,31}^{\boldsymbol{\xi}}(s,\mathbf{x},\mathbf{x}')|_{\boldsymbol{\xi}=\mathbf{x}} \le & C\Bigg(\sum_{\ell=2}^{k-1}\sum_{m=k}^{n} \sup_{s\in[0,T]} \|\big(D_m u\big)_\ell(s,\cdot)\|_{B_{\infty,\infty}^{\alpha_\ell^m}} \int_t^T ds(s-t)^{-1-(i-\frac{1}{2})\alpha_i^k+\gamma_\ell^m}\Bigg)|(\mathbf{x}-\mathbf{x}')_i|^{\alpha_i^k} \\
\le & CT^\delta \sup_{m\in[\![k,n]\!],\ell\in[\![1,k]\!],s\in[0,T]} \|\big(D_m u\big)_\ell(s,\cdot)\|_{B_{\infty,\infty}^{\alpha_\ell^m}} |(\mathbf{x}-\mathbf{x}')_i|^{\alpha_i^k}, \qquad (3.37)
\end{aligned}$$

for some $\delta := 3\eta/8 > 0$. On the other hand, from the definition in (3.34), using as well Proposition 5, Lemmas 9, 12 and 7, we get similarly to (2.38):

$$\begin{aligned}
|\Delta_{k,32}^{\boldsymbol{\xi}}(s,\mathbf{x},\mathbf{x}')|_{\boldsymbol{\xi}=\mathbf{x}} \le & \frac{C[(D_{k-1}\mathbf{F}_k)_{k-1}(s,\cdot))]_\eta}{(s-t)^{(k-\frac{1}{2})+(i-\frac{1}{2})}}|(\mathbf{x}-\mathbf{x}')_i| \int_{\mathbb{R}^d} \hat{p}_{C^{-1}}^{\mathbf{x}}(t,s,\mathbf{x},\mathbf{y})|(\boldsymbol{\theta}_{s,t}(\mathbf{x}) - \mathbf{y})_{k-1}|^{1+\eta} \\
& \times(\|\mathbf{D}u\|_\infty + [(D_k u)_k(s,\cdot)]_{\alpha_k^k})\big(|(\boldsymbol{\theta}_{s,t}(\mathbf{x}) - \mathbf{y})_k|^{\alpha_k^k} + \sum_{m=k+1}^{n} |(\boldsymbol{\theta}_{s,t}(\mathbf{x}) - \mathbf{y})_m|^{\zeta_k}\big) \\
\le & C(\|\mathbf{D}u\|_\infty + [(D_k u)_k(s,\cdot)]_{\alpha_k^k})|(\mathbf{x}-\mathbf{x}')_i|^{\alpha_i^k} \\
& \times\Big(\sum_{m=k}^{n}(s-t)^{-(k-\frac{1}{2})-(i-\frac{1}{2})\alpha_i^k+(1+\eta)(k-\frac{3}{2})+\alpha_k^k(k-\frac{1}{2})\mathbb{I}_{m=k}+\zeta_k(m-\frac{1}{2})\mathbb{I}_{m\in[\![k+1,n]\!]}}\Big), \qquad (3.38)
\end{aligned}$$

where we have expanded the difference of the Gaussian densities and exploited the fact that on the considered time set the diagonal regime holds for those densities. From (3.38) and recalling (2.40) and (2.41) we finally get:

$$\begin{aligned}\int_{\mathcal{S}_i^c} ds|\Delta^{\boldsymbol{\xi}}_{k,32}(s,\mathbf{x},\mathbf{x}')|_{\boldsymbol{\xi}=\mathbf{x}} &\leq C(\|\mathbf{D}u\|_\infty + [(D_k u)_k(s,\cdot)]_{\alpha_k^k})|(\mathbf{x}-\mathbf{x}')_i|^{\alpha_i^k}\sum_{m=k}^n \int_t^T ds(s-t)^{-\frac{3}{2}+\gamma_k^m} \\ &\leq CT^\delta(\|\mathbf{D}u\|_\infty + [(D_k u)_k(s,\cdot)]_{\alpha_k^k})|(\mathbf{x}-\mathbf{x}')_i|^{\alpha_i^k}, \end{aligned} \tag{3.39}$$

for δ small enough. Combining the estimates (3.37) and (3.39) together with (3.33) we eventually derive

$$\left|\int_{\mathcal{S}_i^c} ds\left[I_k^{\boldsymbol{\xi}}(s,\mathbf{x}) - I_k^{\boldsymbol{\xi}}(s,\mathbf{x}')\right]\right|_{\boldsymbol{\xi}=\mathbf{x}} \leq T^\delta C|(\mathbf{x}'-\mathbf{x})_i|^{\alpha_i^k}\left(\|\mathbf{D}u\|_\infty + \sup_{m\in[\![k,n]\!],\ell\in[\![1,k]\!],s\in[0,T]}[\big(D_m u\big)_\ell(s,\cdot)]_{\alpha_\ell^m}\right). \tag{3.40}$$

Now, the term $H_k^{\boldsymbol{\xi}}(s,\mathbf{x}) - H_k^{\boldsymbol{\xi}}(s,\mathbf{x}')$ in (3.32) (non-degenerate variables) can be handled reproducing the same previous arguments for $I_k^{\boldsymbol{\xi}}(s,\mathbf{x}) - I_k^{\boldsymbol{\xi}}(s,\mathbf{x}')$, exploiting (3.31) and following the computations performed for H_k in the proof of Theorem 3 (see e.g. (2.28)). From the definition of $\mathcal{S}_i^c$ we obtain:

$$\begin{aligned}&\left|\int_{t+c_0|(\mathbf{x}'-\mathbf{x})_i|^{\frac{2}{2i-1}}}^T ds\Big(H_k^{\boldsymbol{\xi}}(s,\mathbf{x}) - H_k^{\boldsymbol{\xi}}(s,\mathbf{x}')\Big)\right|_{\boldsymbol{\xi}=\mathbf{x}} \\ &\leq C|(\mathbf{x}'-\mathbf{x})_i|^{\alpha_i^k}\big(\|\mathbf{D}D_1 u\|_\infty + \|\mathbf{D}u\|_\infty\big)\int_t^T ds(s-t)^{-1-\alpha_i^k(i-\frac{1}{2})+\delta_i^k} \\ &\leq C|(\mathbf{x}'-\mathbf{x})_i|^{\alpha_i^k}\big(\|\mathbf{D}D_1 u\|_\infty + \|\mathbf{D}u\|_\infty\big)T^\delta, \end{aligned} \tag{3.41}$$

for some $\delta > 0$ recalling for the last inequality that, from the bound following (3.25), $\delta_i^k \geq 1/2 + \eta/2$, so that $-\alpha_i^k(i-1/2) + \delta_i^k = -(1+\eta/4)/2 + \delta_i^k > 0$. The arguments needed to control this term are actually those already exploited in [CdR17] when $n = 2$.

Gathering equations (3.40) and (3.41), we finally derive with the notations of (3.32):

$$|D_{\mathbf{x}_k}u(t,\mathbf{x}) - D_{\mathbf{x}_k}u(t,\mathbf{x}')|\Big|_{\mathcal{S}_i^c} \leq CT^\delta|(\mathbf{x}'-\mathbf{x})_i|^{\alpha_i^k}\Big(\|\mathbf{D}D_1 u\|_\infty + \|\mathbf{D}u\|_\infty + \sup_{2\leq\ell\leq m\leq n, s\in[0,T]}\|\big(D_m u\big)_\ell(s,\cdot)\|_{B^{\alpha_\ell^m}_{\infty,\infty}}\Big). \tag{3.42}$$

3.2.3 Discontinuity term associated with the regime time change: control of the term (3.23).

We here aim at handling

$$\big(D_{\mathbf{x}_k}u(t,\mathbf{x}) - D_{\mathbf{x}_k}u(t,\mathbf{x}')\big)|_{\partial\mathcal{S}_i} = \Big\{D_{\mathbf{x}_k}\tilde{P}^{\boldsymbol{\xi}'}_{t_0,t}u(t_0,\mathbf{x}') - D_{\mathbf{x}_k}\tilde{P}^{\tilde{\boldsymbol{\xi}}'}_{t_0,t}u(t_0,\mathbf{x}')\Big\}_{(\boldsymbol{\xi}',\tilde{\boldsymbol{\xi}}')=(\mathbf{x}',\mathbf{x})},$$

which we will actually handle like the off-diagonal components. Recall here that the transition time $t_0 = t + c_0|(\mathbf{x}-\mathbf{x}')_i|^{2/(2i-1)}$. From Lemma 8 (cancellation argument), we write:

$$\begin{aligned}&D_{\mathbf{x}_k}\tilde{P}^{\boldsymbol{\xi}'}_{t_0,t}u(t_0,\mathbf{x}') - D_{\mathbf{x}_k}\tilde{P}^{\tilde{\boldsymbol{\xi}}'}_{t_0,t}u(t_0,\mathbf{x}') \\ &= \int_{\mathbb{R}^{nd}} d\mathbf{y}D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}'}(t,t_0,\mathbf{x}',\mathbf{y})[u(t_0,\mathbf{y}) - u(t_0,\mathbf{y}_{1:k-1},(\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_{k:n})] \\ &\quad - \int_{\mathbb{R}^{nd}} d\mathbf{y}D_{\mathbf{x}_k}\tilde{p}^{\tilde{\boldsymbol{\xi}}'}(t,t_0,\mathbf{x}',\mathbf{y})[u(t_0,\mathbf{y}) - u(t_0,\mathbf{y}_{1:k-1},(\mathbf{m}^{\tilde{\boldsymbol{\xi}}'}_{t_0,t}(\mathbf{x}'))_k,\boldsymbol{\theta}_{t_0,t}(\mathbf{x}')_{k+1:n})]. \end{aligned}$$

We now split the above contribution into three terms:

$$D_{\mathbf{x}_k}\tilde{P}^{\boldsymbol{\xi}'}_{t_0,t}u(t_0,\mathbf{x}') - D_{\mathbf{x}_k}\tilde{P}^{\tilde{\boldsymbol{\xi}}'}_{t_0,t}u(t_0,\mathbf{x}') = (\mathscr{B}_1^{\boldsymbol{\xi}',\tilde{\boldsymbol{\xi}}'} + \mathscr{B}_2^{\boldsymbol{\xi}',\tilde{\boldsymbol{\xi}}'} + \mathscr{B}_3^{\boldsymbol{\xi}',\tilde{\boldsymbol{\xi}}'})(t_0,\mathbf{x}'), \tag{3.43}$$

where

$$\begin{aligned}\mathscr{B}_1^{\boldsymbol{\xi}',\tilde{\boldsymbol{\xi}}'}(t_0,\mathbf{x}') := \Bigg\{&\int_{\mathbb{R}^{nd}} d\mathbf{y}D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}'}(t,t_0,\mathbf{x}',\mathbf{y})[u(t_0,\mathbf{y}) - u(t_0,\mathbf{y}_{1:k},(\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_{k+1:n})] \\ &- \int_{\mathbb{R}^{nd}} d\mathbf{y}D_{\mathbf{x}_k}\tilde{p}^{\tilde{\boldsymbol{\xi}}'}(t,t_0,\mathbf{x}',\mathbf{y})\Big[u(t_0,\mathbf{y}) - u(t_0,\mathbf{y}_{1:k},(\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_{k+1:n})\Big]\Bigg\}, \end{aligned} \tag{3.44}$$

$$
\begin{aligned}
&\mathscr{B}_2^{\boldsymbol{\xi}',\tilde{\boldsymbol{\xi}}'}(t_0,\mathbf{x}')\\
:=\ &\Big\{\Big[\int_{\mathbb{R}^{nd}} d\mathbf{y} D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}'}(t,t_0,\mathbf{x}',\mathbf{y})\Big[u(t_0,\mathbf{y}_{1:k},(\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_{k+1:n})-u(t_0,\mathbf{y}_{1:k-1},(\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_{k:n})\\
&\quad-\langle D_k u(t_0,\mathbf{y}_{1:k-1},(\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_{k:n}),(\mathbf{y}-\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_k\rangle\Big]\Big]\\
&-\Big[\int_{\mathbb{R}^{nd}} d\mathbf{y} D_{\mathbf{x}_k}\tilde{p}^{\tilde{\boldsymbol{\xi}}'}(t,t_0,\mathbf{x}',\mathbf{y})[u(t_0,\mathbf{y}_{1:k},(\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_{k+1:n})-u(t_0,\mathbf{y}_{1:k-1},(\mathbf{m}_{t_0,t}^{\tilde{\boldsymbol{\xi}}'}(\mathbf{x}'))_k,\boldsymbol{\theta}_{t_0,t}(\mathbf{x}')_{k+1:n})\\
&\quad-\langle D_k u(t_0,\mathbf{y}_{1:k-1},(\mathbf{m}_{t_0,t}^{\tilde{\boldsymbol{\xi}}'}(\mathbf{x}'))_k,\boldsymbol{\theta}_{t_0,t}(\mathbf{x}')_{k+1:n}),(\mathbf{y}-\mathbf{m}_{t_0,t}^{\tilde{\boldsymbol{\xi}}'}(\mathbf{x}'))_k\rangle]\Big]\Big\},
\end{aligned}
\tag{3.45}
$$

$$
\begin{aligned}
\mathscr{B}_3^{\boldsymbol{\xi}',\tilde{\boldsymbol{\xi}}'}(t_0,\mathbf{x}') :=\ &\Big\{\int_{\mathbb{R}^{nd}} d\mathbf{y} D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}'}(t,t_0,\mathbf{x}',\mathbf{y})\langle D_k u(t_0,\mathbf{y}_{1:k-1},(\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_{k:n}),(\mathbf{y}-\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_k\rangle\\
&-\int_{\mathbb{R}^{nd}} d\mathbf{y} D_{\mathbf{x}_k}\tilde{p}^{\tilde{\boldsymbol{\xi}}'}(t,t_0,\mathbf{x}',\mathbf{y})\langle D_k u(t_0,\mathbf{y}_{1:k-1},(\mathbf{m}_{t_0,t}^{\tilde{\boldsymbol{\xi}}'}(\mathbf{x}'))_k,\boldsymbol{\theta}_{t_0,t}(\mathbf{x}')_{k+1:n}),(\mathbf{y}-\mathbf{m}_{t_0,t}^{\tilde{\boldsymbol{\xi}}'}(\mathbf{x}'))_k\rangle\Big\}.
\end{aligned}
\tag{3.46}
$$

We now exploit the Hölder regularity of $D_k u$ w.r.t. the k^{th} variable to control the terms in $\mathscr{B}_2^{\boldsymbol{\xi}',\tilde{\boldsymbol{\xi}}'}(t_0,\mathbf{x}')$ defined in (3.45). Let us first write from the previous decomposition:

$$
\begin{aligned}
&|\mathscr{B}_{2,1}^{\boldsymbol{\xi}',\tilde{\boldsymbol{\xi}}'}(t_0,\mathbf{x}')|\\
:=\ &\Big|\int_{\mathbb{R}^{nd}} d\mathbf{y} D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}'}(t,t_0,\mathbf{x}',\mathbf{y})\Big[u(t_0,\mathbf{y}_{1:k},(\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_{k+1:n})-u(t_0,\mathbf{y}_{1:k-1},(\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_{k:n})\\
&-\langle D_k u(t_0,\mathbf{y}_{1:k-1},(\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_{k:n}),(\mathbf{y}-\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_k\rangle\Big]\Big|\\
=\ &\int_{\mathbb{R}^{nd}} d\mathbf{y}\int_0^1 d\lambda\Big|D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}'}(t,t_0,\mathbf{x}',\mathbf{y})\Big|[D_k u(t_0,\mathbf{y}_{1:k-1},\cdot,(\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_{k+1:n})]_{\alpha_k^k}|(\mathbf{y}-\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_k|^{1+\alpha_k^k}.
\end{aligned}
$$

From Proposition 5, we thus derive:

$$
\begin{aligned}
&|\mathscr{B}_{2,1}^{\boldsymbol{\xi}',\tilde{\boldsymbol{\xi}}'}(t_0,\mathbf{x}')|_{(\boldsymbol{\xi}',\tilde{\boldsymbol{\xi}}')=(\mathbf{x}',\mathbf{x})}\\
&\le C(t_0-t)^{\alpha_k^k(k-\frac{1}{2})}\sup_{\mathbf{z}_j,j\in[\![1,n]\!],\ j\neq k}[D_k u(t_0,\mathbf{z}_{1:k},\cdot,\mathbf{z}_{k+1:n})]_{\alpha_k^k}\int_{\mathbb{R}^{(n-1)d}} d\mathbf{y}_{1:k-1}d\mathbf{y}_{k+1:n}\hat{q}_{C^{-1}\setminus k}(t,s,\mathbf{x},(\mathbf{y}_{1:k-1},\mathbf{y}_{k+1:n}))\\
&= C(t_0-t)^{\frac{1}{2}+\frac{\eta}{8}}[(D_k u)_k(t_0,\cdot)]_{\alpha_k^k},
\end{aligned}
\tag{3.47}
$$

recalling from (3.14) that $\alpha_k^k=(1+\eta/4)/(2k-1)$. The same arguments readily give:

$$
\begin{aligned}
&|\mathscr{B}_{2,2}^{\boldsymbol{\xi}',\tilde{\boldsymbol{\xi}}'}(t_0,\mathbf{x}')|\\
:=\ &\Big|\int_{\mathbb{R}^{nd}} d\mathbf{y} D_{\mathbf{x}_k}\tilde{p}^{\tilde{\boldsymbol{\xi}}'}(t,t_0,\mathbf{x}',\mathbf{y})[u(t_0,\mathbf{y}_{1:k},(\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_{k+1:n})-u(t_0,\mathbf{y}_{1:k-1},(\mathbf{m}_{t_0,t}^{\tilde{\boldsymbol{\xi}}'}(\mathbf{x}'))_k,\boldsymbol{\theta}_{t_0,t}(\mathbf{x}')_{k+1:n})\\
&-\langle D_k u(t_0,\mathbf{y}_{1:k-1},(\mathbf{m}_{t_0,t}^{\tilde{\boldsymbol{\xi}}'}(\mathbf{x}'))_k,\boldsymbol{\theta}_{t_0,t}(\mathbf{x}')_{k+1:n}),(\mathbf{y}-\mathbf{m}_{t_0,t}^{\tilde{\boldsymbol{\xi}}'}(\mathbf{x}'))_k\rangle]\Big)\Big|\\
\le\ &C(t_0-t)^{\frac{1}{2}+\frac{\eta}{8}}[(D_k u)_k(t_0,\cdot)]_{\alpha_k^k}.
\end{aligned}
\tag{3.48}
$$

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

Let us now deal with the contribution $\mathscr{B}_1^{\boldsymbol{\xi}',\tilde{\boldsymbol{\xi}}'}(t_0,\mathbf{x}')$ in (3.44). Observe from this definition that this term is

non-zero if and only if $k<n$. Write then

$$
\begin{aligned}
&|\mathscr{B}_1^{\boldsymbol{\xi}',\tilde{\boldsymbol{\xi}}'}(t_0,\mathbf{x}')|_{(\boldsymbol{\xi}',\tilde{\boldsymbol{\xi}}')=(\mathbf{x}',\mathbf{x})}\\
\leq\ &\int_{\mathbb{R}^{nd}} d\mathbf{y}|D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}'}(t,t_0,\mathbf{x}',\mathbf{y})|_{\boldsymbol{\xi}'=\mathbf{x}'}|u(t_0,\mathbf{y})-u(t_0,\mathbf{y}_{1:k},(\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_{k+1:n})|\\
&+\int_{\mathbb{R}^{nd}} d\mathbf{y}|D_{\mathbf{x}_k}\tilde{p}^{\tilde{\boldsymbol{\xi}}'}(t,t_0,\mathbf{x}',\mathbf{y})|_{\tilde{\boldsymbol{\xi}}'=\mathbf{x}}|u(t_0,\mathbf{y})-u(t_0,\mathbf{y}_{1:k},(\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_{k+1:n})|\\
\leq\ &C\|\mathbf{D}u\|_\infty\Big(\int_{\mathbb{R}^{nd}}\frac{d\mathbf{y}}{(t_0-t)^{k-\frac{1}{2}}}\hat{p}^{\mathbf{x}'}_{C^{-1}}(t,t_0,\mathbf{x}',\mathbf{y})|(\boldsymbol{\theta}_{t_0,t}(\mathbf{x}')-\mathbf{y})_{k+1:n}| \qquad (3.49)\\
&+\int_{\mathbb{R}^{nd}}\frac{d\mathbf{y}}{(t_0-t)^{k-\frac{1}{2}}}\hat{p}^{\mathbf{x}}_{C^{-1}}(t,t_0,\mathbf{x}',\mathbf{y})\big(|(\mathbf{m}^{\mathbf{x}}_{t_0,t}(\mathbf{x}')-\mathbf{y})_{k+1:n}|+|(\mathbf{m}^{\mathbf{x}}_{t_0,t}(\mathbf{x}')-\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_{k+1:n}|\big)\Big).
\end{aligned}
$$

To deal with the last contribution in the r.h.s., we will need some auxilliary lemmas already used in [CdRHM20] for Schauder estimates. Namely, analogously to Lemmas 1 and 3 therein, we have the following result.

Lemma 14. *There exists $\vartheta=\vartheta((\mathbf{A}))\in(0,1)$ s.t. for all $j\in[\![1,n]\!]$, $\mathbf{x}\in\mathbb{R}^{nd},\mathbf{x}'=(\mathbf{x}_{1:i-1},z,\mathbf{x}_{i+1:n})\in\mathbb{R}^{nd}$:*

$$
|(\mathbf{m}^{\mathbf{x}}_{t_0,t}(\mathbf{x}')-\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_j|\leq c_0^{\vartheta}\mathbf{d}^{2j-1}(\mathbf{x},\mathbf{x}'):=c_0^{\vartheta}|(\mathbf{x}-\mathbf{x}')_i|^{\frac{2j-1}{2i-1}}. \qquad (3.50)
$$

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

$$
\begin{aligned}
\frac{\|\mathbf{D}u\|_\infty}{(t_0-t)^{k-\frac{1}{2}}}|(\mathbf{m}^{\mathbf{x}}_{t_0,t}(\mathbf{x}')-\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_{k+1:n}| &\leq \frac{\|\mathbf{D}u\|_\infty}{c_0^{k-\frac{1}{2}}|(\mathbf{x}-\mathbf{x}')_i|^{\frac{2k-1}{2i-1}}}c_0^{\vartheta}|(\mathbf{x}-\mathbf{x}')_i|^{\frac{2(k+1)-1}{2i-1}}\\
&\leq \|\mathbf{D}u\|_\infty c_0^{\vartheta-(k-\frac{1}{2})}|(\mathbf{x}-\mathbf{x}')_i|^{\frac{2}{2i-1}}\\
&\leq \|\mathbf{D}u\|_\infty c_0^{\vartheta-(k-\frac{1}{2})}|(\mathbf{x}-\mathbf{x}')_i|^{\alpha_i^k}.
\end{aligned}
$$

Plugging the above control in (3.49) and using as well Lemma 7 and Proposition 5, we obtain:

$$
|\mathscr{B}_1^{\boldsymbol{\xi}',\tilde{\boldsymbol{\xi}}'}(t_0,\mathbf{x}')|_{(\boldsymbol{\xi}',\tilde{\boldsymbol{\xi}}')=(\mathbf{x}',\mathbf{x})}\ \leq\ C\|\mathbf{D}u\|_\infty\left\{2(t_0-t)+c_0^{\vartheta-(k-\frac{1}{2})}|(\mathbf{x}-\mathbf{x}')_i|^{\alpha_i^k}\right\}. \qquad (3.51)
$$

Let us eventually control the term $\mathscr{B}_3^{\boldsymbol{\xi}',\tilde{\boldsymbol{\xi}}'}(t_0,\mathbf{x}')$ defined in (3.46) which we rewrite in the following way:

$$
\begin{aligned}
&\mathscr{B}_3^{\boldsymbol{\xi}',\tilde{\boldsymbol{\xi}}'}(t_0,\mathbf{x}')\\
=&\Big\{\int_{\mathbb{R}^{nd}}d\mathbf{y}D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}'}(t,t_0,\mathbf{x}',\mathbf{y})\Big\langle\big[D_ku(t_0,\mathbf{y}_{1:k-1},(\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_{k:n})-D_ku(t_0,\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))\big],(\mathbf{y}-\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_k\Big\rangle\\
&-\int_{\mathbb{R}^{nd}}d\mathbf{y}D_{\mathbf{x}_k}\tilde{p}^{\tilde{\boldsymbol{\xi}}'}(t,t_0,\mathbf{x}',\mathbf{y})\Big\langle\big[D_ku(t_0,\mathbf{y}_{1:k-1},(\mathbf{m}^{\tilde{\boldsymbol{\xi}}'}_{t_0,t}(\mathbf{x}'))_k,\boldsymbol{\theta}_{t_0,t}(\mathbf{x}')_{k+1:n})\\
&-D_ku(t_0,(\mathbf{m}^{\tilde{\boldsymbol{\xi}}'}_{t_0,t}(\mathbf{x}'))_{1:k},\boldsymbol{\theta}_{t_0,t}(\mathbf{x}')_{k+1:n})\big](\mathbf{y}-\mathbf{m}^{\tilde{\boldsymbol{\xi}}'}_{t_0,t}(\mathbf{x}'))_k\Big\rangle\Big\}\\
&+\Big\{\int_{\mathbb{R}^{nd}}d\mathbf{y}D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}'}(t,t_0,\mathbf{x}',\mathbf{y})\\
&\Big\langle[D_ku(t_0,\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))-D_ku(t_0,(\mathbf{m}^{\tilde{\boldsymbol{\xi}}'}_{t_0,t}(\mathbf{x}'))_{1:k},\boldsymbol{\theta}_{t_0,t}(\mathbf{x}')_{k+1:n}),(\mathbf{y}-\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_k\Big\rangle\Big\}\\
&-\Big\{\int_{\mathbb{R}^{nd}}d\mathbf{y}D_{\mathbf{x}_k}\tilde{p}^{\tilde{\boldsymbol{\xi}}'}(t,t_0,\mathbf{x}',\mathbf{y})\Big\langle D_ku(t_0,(\mathbf{m}^{\tilde{\boldsymbol{\xi}}'}_{t_0,t}(\mathbf{x}'))_{1:k},\boldsymbol{\theta}_{t_0,t}(\mathbf{x}')_{k+1:n}),(\mathbf{y}-\mathbf{m}^{\tilde{\boldsymbol{\xi}}'}_{t_0,t}(\mathbf{x}'))_k\Big\rangle\\
&-\int_{\mathbb{R}^{nd}}d\mathbf{y}D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}'}(t,t_0,\mathbf{x}',\mathbf{y})\Big\langle D_ku(t_0,(\mathbf{m}^{\tilde{\boldsymbol{\xi}}'}_{t_0,t}(\mathbf{x}'))_{1:k},\boldsymbol{\theta}_{t_0,t}(\mathbf{x}')_{k+1:n}),(\mathbf{y}-\boldsymbol{\theta}_{t_0,t}(\mathbf{x}'))_k\Big\rangle\Big\},
\end{aligned}
$$

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 $\mathbf{x}_{1:k} \mapsto D_k u(\cdot, \mathbf{x}_{1:k}, \cdot)$, Proposition 5 and Lemma 14:

$$
\begin{aligned}
&|\mathscr{B}_3^{\boldsymbol{\xi}',\tilde{\boldsymbol{\xi}}'}(t_0,\mathbf{x}')|_{(\boldsymbol{\xi}',\tilde{\boldsymbol{\xi}}')=(\mathbf{x}',\mathbf{x})} \\
&\leq C\int_{\mathbb{R}^{nd}} d\mathbf{y}\left\{\hat{p}^{\mathbf{x}'}_{C^{-1}}(t,t_0,\mathbf{x}',\mathbf{y})+\hat{p}^{\mathbf{x}}_{C^{-1}}(t,t_0,\mathbf{x}',\mathbf{y})\right\}\sum_{j=1}^{k-1}[(D_k u)_j(t_0,\cdot)]_{\alpha_j^k}(t_0-t)^{\alpha_j^k(j-\frac{1}{2})} \\
&\quad + C\int_{\mathbb{R}^{nd}} d\mathbf{y}\hat{p}^{\mathbf{x}'}_{C^{-1}}(t,t_0,\mathbf{x}',\mathbf{y})\sum_{j=1}^{k}[(D_k u)_j(t_0,\cdot)]_{\alpha_j^k}(t_0-t)^{\alpha_j^k(j-\frac{1}{2})} \\
&\leq C\Big(\sum_{j=1}^{k}[(D_k u)_j(t_0,\cdot)]_{\alpha_j^k}\Big)(t_0-t)^{\frac{1}{2}+\frac{\eta}{8}}. \qquad (3.52)
\end{aligned}
$$

Thus, plugging estimates (3.47), (3.48), (3.51) and (3.52) in (3.43) we deduce that

$$
\begin{aligned}
&|D_{\mathbf{x}_k}\tilde{P}^{\boldsymbol{\xi}'}_{t_0,t}u(t_0,\mathbf{x}')-D_{\mathbf{x}_k}\tilde{P}^{\tilde{\boldsymbol{\xi}}'}_{t_0,t}u(t_0,\mathbf{x}')|_{(\boldsymbol{\xi}',\tilde{\boldsymbol{\xi}}')=(\mathbf{x}',\mathbf{x})} \\
&\leq C\Big[\|\mathbf{D}u\|_\infty\left\{(t_0-t)+c_0^{\vartheta-(k-\frac{1}{2})}|(\mathbf{x}-\mathbf{x}')_i|^{\alpha_i^k}\right\}+\Big(\sum_{j=1}^{k}[(D_k u)_j(t_0,\cdot)]_{\alpha_j^k}\Big)(t_0-t)^{\frac{1}{2}+\frac{\eta}{8}}\Big]. \qquad (3.53)
\end{aligned}
$$

From the definition of $t_0 = t + c_0|(\mathbf{x}-\mathbf{x}')_i|^{2/(2i-1)}$ and $\alpha_i^k = (1+\eta/4)/(2i-1)$, recalling that $t_0 - t$ is small as well (i.e. $t_0 - t \leq C(t_0-t)^{1/2+\eta/8}$), we obtain from (3.53):

$$
\begin{aligned}
&|D_{\mathbf{x}_k}\tilde{P}^{\boldsymbol{\xi}'}_{t_0,t}u(t_0,\mathbf{x}')-D_{\mathbf{x}_k}\tilde{P}^{\tilde{\boldsymbol{\xi}}'}_{t_0,t}u(t_0,\mathbf{x}')|_{(\boldsymbol{\xi}',\tilde{\boldsymbol{\xi}}')=(\mathbf{x}',\mathbf{x})} \\
&\leq C\left\{c_0^{\frac{1}{2}+\frac{\eta}{8}}\Big(\sum_{j=1}^{k}[\big(D_k u\big)_j(t_0,\cdot)]_{\alpha_j^k}\Big)+(c_0^{\vartheta-(k-\frac{1}{2})}+c_0^{\frac{1}{2}+\frac{\eta}{8}})\|\mathbf{D}u\|_\infty\right\}|(\mathbf{x}-\mathbf{x}')_i|^{\alpha_i^k}.
\end{aligned}
$$

Thus, from (3.23), there exists $\tilde{\delta} := \tilde{\delta}((\mathbf{A}))\in(0,1)$ such that

$$
\Big|\big(D_{\mathbf{x}_k}u(t,\mathbf{x})-D_{\mathbf{x}_k}u(t,\mathbf{x}')\big)|_{\partial\mathcal{S}_i}\Big| \leq |(\mathbf{x}-\mathbf{x}')_i|^{\alpha_i^k}C(c_0^{-\tilde{\delta}^{-1}}\|\mathbf{D}u\|_\infty + c_0^{\tilde{\delta}}\max_{2\leq\ell\leq m\leq n}[\big(D_m u\big)_\ell(t_0,\cdot)]_{\alpha_\ell^m}). \qquad (3.54)
$$

Conclusion: control of (3.20)**.** Plugging (3.28), (3.42) and (3.54) into (3.20), (3.18), we eventually derive that for some positive $\tilde{\delta} := \tilde{\delta}((\mathbf{A})) > 0$:

$$
\|D_{\mathbf{x}_k}u(t,\mathbf{x}_{1:i-1},\cdot,\mathbf{x}_{i+1,n})\|_{B^{\alpha_i^k}_{\infty,\infty}} \leq C\Big(c_0^{-\tilde{\delta}^{-1}}\|\mathbf{D}u\|_\infty+\|\mathbf{D}D_1u\|_\infty+\big(c_0^{\tilde{\delta}}+T^{\tilde{\delta}}\big)\sup_{s\in[0,T],2\leq\ell\leq m\leq n}\|\big(D_m u\big)_\ell(s,\cdot)\|_{B^{\alpha_\ell^m}_{\infty,\infty}}\Big).
$$

The main point to close our circular argument then consists in taking the supremums w.r.t. $\mathbf{x}_{1:i-1},\mathbf{x}_{i+1:n}$, 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 $C\big(c_0^{\tilde{\delta}}+T^{\tilde{\delta}}\big)\leq 1/2$. We then derive that for all $2\leq i\leq k\leq n$:

$$
\sup_{t\in[0,T]}\|\big(D_k u\big)_i(t,\cdot)\|_{B^{\alpha_i^k}_{\infty,\infty}} \leq C\{\|\mathbf{D}u\|_\infty+\|\mathbf{D}D_1u\|_\infty\}, \qquad (3.55)
$$

which concludes the proof of Lemma 11. □.

3.3 Proof of of Lemma 13

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:

$$
D_\ell\left[\Big[\Psi^{(s,\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}),(t,\mathbf{x},\mathbf{x})}_{\ell,k,m}(\cdot)-\Psi^{(s,\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}),(t,\mathbf{x}',\mathbf{x})}_{\ell,k,m}(\cdot)\Big]\right]. \qquad (3.56)
$$

Splitting the thermic part of the Besov norm as in (3.2), we obtain the same kind estimate for the non-singular in time part. Indeed, we point out that the difference (3.56) does not involve $D_{\mathbf{x}_1}$, therefore we gain a singularity of order 1/2 w.r.t. equation (3.3). On the other hand, the difference of the derivatives of the frozen densities

w.r.t. $\mathbf{x}_k$ can be handled with (3.31). Choosing $\rho_{\ell,m} = 2\ell - 1$ as in the proof of Lemma 10, and recalling that $\alpha_i^k(i-1/2) = (1+\eta/4)/2$ (see (3.14)), it is plain to check that:

$$\begin{aligned}
&\mathbf{Upper}\Big[D_\ell\Big[\Psi_{\ell,k,m}^{(s,\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}),(t,\mathbf{x},\mathbf{x})}(\cdot) - \Psi_{\ell,k,m}^{(s,\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}),(t,\mathbf{x}',\mathbf{x})}(\cdot)\Big]\Big] \\
\leq\ & \frac{\tilde{C}|(\mathbf{x}-\mathbf{x}')_i|^{\alpha_i^k}}{(s-t)^{1+\alpha_i^k(i-\frac{1}{2})-\gamma_\ell^m}}\hat{q}_{c\backslash\ell}(t,s,\mathbf{x},(\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n})),
\end{aligned} \tag{3.57}$$

where from $\mathbf{(H_\eta)}$, $1+\alpha_i^k(i-1/2)-\gamma_\ell^m < 1$. Turning to the singular in time contribution of the thermic part of the Besov norm of (3.56) we decompose with the notations of Lemma 10 (see e.g. (3.8), (3.9) which exhibit an additional spatial derivative):

$$\begin{aligned}
&\mathbf{Lower}\Big[D_\ell\Big[\Psi_{\ell,k,m}^{(s,\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}),(t,\mathbf{x},\mathbf{x})}(\cdot) - \Psi_{\ell,k,m}^{(s,\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}),(t,\mathbf{x}',\mathbf{x})}(\cdot)\Big]\Big] \\
=:\ & \int_0^{(s-t)^{\rho_{\ell,m}}}\frac{dw}{w}w^{\frac{\alpha_\ell^m}{2}+1}\int_{\mathbb{R}^d}d\tilde{z}|(\mathscr{T}_{1,\ell,k,m}^{(s,\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}),(t,\mathbf{x},\mathbf{x}')} + \mathscr{T}_{2,\ell,k,m}^{(s,\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}),(t,\mathbf{x},\mathbf{x}')})\big(w,\tilde{z}\big)|,
\end{aligned}$$

where $\ell < k \leq m$ and:

$$\begin{aligned}
&\mathscr{T}_{1,\ell,k,m}^{(s,\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}),(t,\mathbf{x},\mathbf{x}')}\big(w,\tilde{z}\big) \\
&= \int_{\mathbb{R}^d}d\mathbf{y}_\ell D_\ell\partial_w h_w(\tilde{z}-\mathbf{y}_\ell)\big(D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y}) - D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x}',\mathbf{y})\big)_{\boldsymbol{\xi}=\mathbf{x}} \\
&\big\langle\mathbf{F}_\ell(s,\mathbf{y}_{1:k-1},\boldsymbol{\theta}_{s,t}^{k:n}(\mathbf{x})) - \mathbf{F}_\ell(s,\mathbf{y}_{1:\ell-1},\tilde{z},\mathbf{y}_{\ell+1:k-1},\boldsymbol{\theta}_{s,t}^{k:n}(\mathbf{x}))\big),(\mathbf{y}-\boldsymbol{\theta}_{s,t}(\mathbf{x}))_m\big\rangle,
\end{aligned} \tag{3.58}$$

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

$$\begin{aligned}
&\mathscr{T}_{2,\ell,k,m}^{(s,\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}),(t,\mathbf{x},\mathbf{x}')}\big(w,\tilde{z}\big) \\
&= \int_{\mathbb{R}^d}d\mathbf{y}_\ell D_\ell\partial_w h_w(\tilde{z}-\mathbf{y}_\ell)\Big[\Big(D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y}) - D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y}_{1:\ell-1},\tilde{z},\mathbf{y}_{\ell+1:n})\Big) \\
&-\Big(D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x}',\mathbf{y}) - D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x}',\mathbf{y}_{1:\ell-1},\tilde{z},\mathbf{y}_{\ell+1:n})\Big)\Big]_{\boldsymbol{\xi}=\mathbf{x}} \\
&\Big\langle\mathbf{F}_\ell(s,\mathbf{y}_{1:\ell-1},\tilde{z},\mathbf{y}_{\ell+1:k-1},\boldsymbol{\theta}_{s,t}^{k:n}(\mathbf{x})) - \mathbf{F}_\ell(s,\boldsymbol{\theta}_{s,t}(\mathbf{x})) - D_{\ell-1}\mathbf{F}_\ell(s,\boldsymbol{\theta}_{s,t}(\mathbf{x}))(\mathbf{y}-\boldsymbol{\theta}_{s,t}(\mathbf{x}))_{\ell-1},(\mathbf{y}-\boldsymbol{\theta}_{s,t}(\mathbf{x}))_m\Big\rangle.
\end{aligned} \tag{3.59}$$

Note now that, when proceeding first as in (3.29), (3.30) and controlling then the associated difference as in (3.9), we get thanks to (3.31) that:

$$\begin{aligned}
&\Big|\Big(D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y}) - D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y}_{1:\ell-1},\tilde{z},\mathbf{y}_{\ell+1:n})\Big) - \Big(D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x}',\mathbf{y}) - D_{\mathbf{x}_k}\tilde{p}^{\boldsymbol{\xi}}(t,s,\mathbf{x}',\mathbf{y}_{1:\ell-1},\tilde{z},\mathbf{y}_{\ell+1:n})\Big)\Big| \\
&\leq \frac{\tilde{C}}{(s-t)^{\ell-\frac{1}{2}+i-\frac{1}{2}+k-\frac{1}{2}}}\int_0^1 d\lambda\hat{p}_{C^{-1}}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y}_{1:\ell-1},\tilde{z}+\lambda(\mathbf{y}_\ell-z),\mathbf{y}_{\ell+1:n})|(\mathbf{x}'-\mathbf{x})_i||\tilde{z}-\mathbf{y}_\ell| \\
&\leq \frac{\tilde{C}}{(s-t)^{(\ell-\frac{1}{2})+\alpha_i^k(i-\frac{1}{2})+(k-\frac{1}{2})}}\int_0^1 d\lambda\hat{p}_{C^{-1}}^{\boldsymbol{\xi}}(t,s,\mathbf{x},\mathbf{y}_{1:\ell-1},\tilde{z}+\lambda(\mathbf{y}_\ell-z),\mathbf{y}_{\ell+1:n})|(\mathbf{x}'-\mathbf{x})_i|^{\alpha_i^k}|\tilde{z}-\mathbf{y}_\ell|,
\end{aligned}$$

using the fact we are in the diagonal regime in the last inequality. With this control at hand, together with estimate (3.31), to handle the contributions $(\mathscr{T}_{1,\ell,k,m}^{(s,\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}),(t,\mathbf{x},\mathbf{x}')} + \mathscr{T}_{2,\ell,k,m}^{(s,\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}),(t,\mathbf{x},\mathbf{x}')})\big(w,\tilde{z}\big)$, we can mimic the proof of the estimation of the contributions $(\mathscr{T}_{1,i,(l,1),k}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})} + \mathscr{T}_{2,i,(l,1),k}^{(s,\mathbf{y}_{1:i-1},\mathbf{y}_{i+1:n}),(t,\mathbf{x})})\big(v,z\big)$ done in Lemma 10 to obtain

$$\begin{aligned}
&\mathbf{Lower}\Big[D_\ell\Big[\Psi_{\ell,k,m}^{(s,\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}),(t,\mathbf{x},\mathbf{x})}(\cdot) - \Psi_{\ell,k,m}^{(s,\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}),(t,\mathbf{x}',\mathbf{x})}(\cdot)\Big]\Big] \\
\leq\ & \tilde{C}\hat{q}_{c\backslash\ell}(t,s,\mathbf{x},(\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}))|(\mathbf{x}'-\mathbf{x})_i|^{\alpha_i^k}\int_0^{(s-t)^{\rho_{\ell,m}}}\frac{dw}{w}w^{\frac{\alpha_\ell^m}{2}+1} \\
&\left\{\frac{1}{w^{\frac{3}{2}-\frac{\beta_\ell}{2}}(s-t)^{\alpha_i^k(i-\frac{1}{2})}} + \frac{w^{-1+\frac{\beta_\ell}{2}}}{(s-t)^{\ell-\frac{1}{2}+\alpha_i^k(i-\frac{1}{2})}} + \frac{w^{-1}}{(s-t)^{\ell-\frac{1}{2}+\alpha_i^k(i-\frac{1}{2})-(1+\eta)(\ell-3/2)}}\right\} \\
\leq\ & \tilde{C}\hat{q}_{c\backslash\ell}(t,s,\mathbf{x},(\mathbf{y}_{1:\ell-1},\mathbf{y}_{\ell+1:n}))(s-t)^{-1-(i-\frac{1}{2})\alpha_i^k+\gamma_\ell^m}|(\mathbf{x}'-\mathbf{x})_i|^{\alpha_i^k},
\end{aligned}$$

where $\gamma_\ell^m = 1/2 + \eta(\ell - 3/2)$. From **($\mathbf{H}_\eta$)** and the very definition of α_i^k we hence have $1 - \alpha_i^k(i-1/2) + \gamma_\ell^m < 1$ which, together with (3.57), concludes the proof. □

A Sensitivity results for the mean: Proof of Lemma 14

In order to prove Lemma 14, we first need to establish some controls on the sensitivity of the flows, see Lemma 15 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 14 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 $(\mathbf{x}, \mathbf{x}') \in \mathbb{R}^{nd}$, we define:

$$\mathbf{d}(\mathbf{x}, \mathbf{x}') := \sum_{i=1}^{n} |(\mathbf{x} - \mathbf{x}')_i|^{\frac{1}{2i-1}}. \tag{A.1}$$

The distance is homogeneous in the sense that, for any $\lambda > 0$, $\mathbf{d}(\lambda^{-1/2}\mathbb{T}_\lambda \mathbf{x}, \lambda^{-1/2}\mathbb{T}_\lambda \mathbf{x}') = \lambda^{1/2}\mathbf{d}(\mathbf{x}, \mathbf{x}')$.

A.1 A first sensitivity result for the flow

Lemma 15 (Control of the flows)**.** *Under* **(A)**, *there exists* $C := C((\mathbf{A}), T)$ *s.t. for all spatial points* $(\mathbf{x}, \mathbf{x}') \in (\mathbb{R}^{nd})^2$, $\mathbf{d}(\mathbf{x}, \mathbf{x}') \le 1$, $0 \le t < s \le T \le 1$ *and* $i \in [\![1, n]\!]$*:*

$$|(\boldsymbol{\theta}_{s,t}(\mathbf{x}) - \boldsymbol{\theta}_{s,t}(\mathbf{x}'))_i| \le C\Big((s-t)^{i-\frac{1}{2}} + \mathbf{d}^{2i-1}(\mathbf{x} - \mathbf{x}')\Big).$$

The flow, $\boldsymbol{\theta}_{s,t}$ is, somehow, locally "almost" Lipschitz continuous in space w.r.t. the homogeneous distance $\mathbf{d}$, up to a time additive term. This time contribution is a consequence of the non-Lipschitz continuity of the drif $\mathbf{F}$. The analysis which was already done for $\mathbf{F}$ Lipschitz continuous in Proposition 4.1 of [Men18], and Appendix A.1 in [CdRHM20] with different Hölder regularity of $\mathbf{F}$. Actually, as we consider a *smoother* drift than in [CdRHM20], 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 $\mathbf{F}$ 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 $\mathbf{F}_i$ for $i \in [\![2, n]\!]$. Namely, for all $v \in [0, T]$, $\mathbf{z} \in \mathbb{R}^{nd}$, $i \in [\![2, n]\!]$, we define

$$\mathbf{F}_i^\delta(v, \mathbf{z}^{i-1:n}) := \mathbf{F}_i(v, \cdot) \star \rho_{\delta_i}(\mathbf{z}) = \int_{\mathbb{R}^d} dw \mathbf{F}_i(v, \mathbf{z}_{i-1}, \mathbf{z}_i - w, \mathbf{z}_{i+1}, \dots, \mathbf{z}_n)\rho_{\delta_i}(w), \tag{A.2}$$

with $\rho_{\delta_i}(w) := (1/\delta_i^d)\rho(w/\delta_i)$ where $\rho : \mathbb{R}^d \to \mathbb{R}_+$ is a usual mollifier, namely ρ has compact support and $\int_{\mathbb{R}^d} \rho(\mathbf{z})d\mathbf{z} = 1$. Eventually, we write $\mathbf{F}^\delta(v, \mathbf{z}) := (\mathbf{F}_1(v, \mathbf{z}), \mathbf{F}_2^\delta(v, \mathbf{z}), \dots, \mathbf{F}_n^\delta(v, \mathbf{z}))$. With a slight abuse of notation in the previous definitions, since the first component $\mathbf{F}_1$ is not mollified. The sublinearity of $\mathbf{F}_1$ 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 δ_i in order to have $C := C((\mathbf{A}), T) > 0$ s.t. for all $\mathbf{z} \in \mathbb{R}^{nd}$, $u \in [t, s]$:

$$\Big|(s-t)^{\frac{1}{2}}\mathbb{T}_{s-t}^{-1}\Big(\mathbf{F}(u, \mathbf{z}) - \mathbf{F}^\delta(u, \mathbf{z})\Big)\Big| \le C(s-t)^{-1}. \tag{A.3}$$

By the previous definition of $\mathbf{F}^\delta$ in (A.2), identity (A.3) is equivalent to:

$$\sum_{i=2}^{n}(s-t)^{\frac{1}{2}-i}\delta_i^{\frac{2i-2}{2i-1}} \le C(s-t)^{-1}. \tag{A.4}$$

Hence, we choose from now on, for each $i \in [\![2, n]\!]$:

$$\delta_i = (s-t)^{(i-\frac{3}{2})\frac{2i-1}{2i-2}}. \tag{A.5}$$

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

$$\begin{aligned}
&|(\boldsymbol{\theta}_{s,t}(\mathbf{x})-\boldsymbol{\theta}_{s,t}(\mathbf{x}'))_n|\\
\le\;& |(\mathbf{x}-\mathbf{x}')_n| + \int_t^s dv\Big(|\mathbf{F}_n^\delta(v,\boldsymbol{\theta}_{v,t}(\mathbf{x}))-\mathbf{F}_n^\delta(v,\boldsymbol{\theta}_{v,t}(\mathbf{x}'))|\\
&+|\mathbf{F}_n^\delta(v,\boldsymbol{\theta}_{v,t}(\mathbf{x}))-\mathbf{F}_n(v,\boldsymbol{\theta}_{v,t}(\mathbf{x}))| + |\mathbf{F}_n^\delta(v,\boldsymbol{\theta}_{v,t}(\mathbf{x}'))-\mathbf{F}_n(v,\boldsymbol{\theta}_{v,t}(\mathbf{x}'))|\Big)\\
\le\;& |(\mathbf{x}-\mathbf{x}')_n| + C\int_t^s dv\Big(\big|\big(\boldsymbol{\theta}_{v,t}(\mathbf{x})-\boldsymbol{\theta}_{v,t}(\mathbf{x}')\big)_{n-1}\big| + \delta_n^{-1+\frac{2n-2}{2n-1}}\big|\big(\boldsymbol{\theta}_{v,t}(\mathbf{x})-\boldsymbol{\theta}_{v,t}(\mathbf{x}')\big)_n\big|\Big) + (s-t)\delta_n^{\frac{2n-2}{2n-1}},
\end{aligned}$$

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

Hence by Grönwall's lemma, we get:

$$\begin{aligned}
&|(\boldsymbol{\theta}_{s,t}(\mathbf{x})-\boldsymbol{\theta}_{s,t}(\mathbf{x}'))_n|\\
&\le C\exp\Big(C(s-t)\delta_n^{-1+\frac{2n-2}{2n-1}}\Big)\Big(|(\mathbf{x}-\mathbf{x}')_n| + (s-t)\delta_n^{\frac{2n-2}{2n-1}} + \int_t^s dv\big|\big(\boldsymbol{\theta}_{v,t}(\mathbf{x})-\boldsymbol{\theta}_{v,t}(\mathbf{x}')\big)_{n-1}\big|\Big)\\
&\le C\exp\Big(C(s-t)^{\frac12}\Big)\Big(|(\mathbf{x}-\mathbf{x}')_n| + (s-t)^{n-\frac12} + \int_t^s dv\big|\big(\boldsymbol{\theta}_{v,t}(\mathbf{x})-\boldsymbol{\theta}_{v,t}(\mathbf{x}')\big)_{n-1}\big|\Big),
\end{aligned}\tag{A.6}$$

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

$$\begin{aligned}
&|(\boldsymbol{\theta}_{s,t}(\mathbf{x})-\boldsymbol{\theta}_{s,t}(\mathbf{x}'))_{n-1}|\\
&\le C\exp\Big(C(s-t)\delta_{n-1}^{-1+\frac{2(n-1)-2}{2(n-1)-1}}\Big)\Big(|(\mathbf{x}-\mathbf{x}')_{n-1}| + (s-t)\delta_{n-1}^{\frac{2(n-1)-2}{2(n-1)-1}}\\
&\quad + \int_t^s dv\big\{\big|\big(\boldsymbol{\theta}_{v,t}(\mathbf{x})-\boldsymbol{\theta}_{v,t}(\mathbf{x}')\big)_{n-2}\big| + \big|\big(\boldsymbol{\theta}_{v,t}(\mathbf{x})-\boldsymbol{\theta}_{v,t}(\mathbf{x}')\big)_n\big|^{\beta_n}\big\}\Big)\\
&\le C\exp(C(s-t)^{\frac12})\Big(|(\mathbf{x}-\mathbf{x}')_{n-1}| + (s-t)^{n-\frac32} + \int_t^s dv\Big\{\big|\big(\boldsymbol{\theta}_{v,t}(\mathbf{x})-\boldsymbol{\theta}_{v,t}(\mathbf{x}')\big)_{n-2}\big|\\
&\quad +|(\mathbf{x}-\mathbf{x}')_n|^{\beta_n} + (v-t)^{\beta_n(n-\frac12)} + \Big(\int_t^v dw\big|\big(\boldsymbol{\theta}_{w,t}(\mathbf{x})-\boldsymbol{\theta}_{w,t}(\mathbf{x}')\big)_{n-1}\big|\Big)^{\beta_n}\Big\}\Big)\\
&\le C\exp(C(s-t)^{\frac12})\Big(|(\mathbf{x}-\mathbf{x}')_{n-1}| + (s-t)^{n-\frac32} + \int_t^s dv\Big\{\big|\big(\boldsymbol{\theta}_{v,t}(\mathbf{x})-\boldsymbol{\theta}_{v,t}(\mathbf{x}')\big)_{n-2}\big|\\
&\quad +|(\mathbf{x}-\mathbf{x}')_n|^{\frac{2n-3}{2n-1}} + \Big(\int_t^v dw\big|\big(\boldsymbol{\theta}_{w,t}(\mathbf{x})-\boldsymbol{\theta}_{w,t}(\mathbf{x}')\big)_{n-1}\big|\Big)^{\beta_n}\Big\}\Big),
\end{aligned}\tag{A.7}$$

from our choice of δ_{n-1} in (A.5) for the second inequality. We also exploited for the last inequality that, since under $(\mathbf{T}_\beta)$, $\beta_n > (2n-2)/(2n-1)$, $\beta_n(n-\frac12) > n-1 > n-3/2$ and $0 \le t < s \le T$ where T is *small*, then $(v-t)^{\beta_n(n-1/2)} \le (v-t)^{n-3/2}$. Also, since $\mathbf{d}(\mathbf{x},\mathbf{x}') \le 1$, the same arguments yield $|(\mathbf{x}-\mathbf{x}')_n|^{\beta_n} \le |(\mathbf{x}-\mathbf{x}')_n|^{(2n-2)/(2n-1)} \le |(\mathbf{x}-\mathbf{x}')_n|^{(2n-3)/(2n-1)}$.

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

$$\begin{aligned}
&\sup_{\tilde{s}\in[t,s]}|(\boldsymbol{\theta}_{\tilde{s},t}(\mathbf{x})-\boldsymbol{\theta}_{\tilde{s},t}(\mathbf{x}'))_{n-1}|\\
\le\;& C\exp(C(s-t)^{\frac12})\Big(|(\mathbf{x}-\mathbf{x}')_{n-1}| + (s-t)^{n-\frac32} + \int_t^s dv\Big\{\big|\big(\boldsymbol{\theta}_{v,t}(\mathbf{x})-\boldsymbol{\theta}_{v,t}(\mathbf{x}')\big)_{n-2}\big|\\
&+|(\mathbf{x}-\mathbf{x}')_n|^{\frac{2n-3}{2n-1}} + \Big(\int_t^v dw\big|\big(\boldsymbol{\theta}_{w,t}(\mathbf{x})-\boldsymbol{\theta}_{w,t}(\mathbf{x}')\big)_{n-1}\big|\Big)^{\beta_n}\Big\}\Big).
\end{aligned}$$

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

$$\begin{aligned}
&\sup_{\tilde{s}\in[t,s]}|(\boldsymbol{\theta}_{\tilde{s},t}(\mathbf{x})-\boldsymbol{\theta}_{\tilde{s},t}(\mathbf{x}'))_{n-1}|\\
\le\;& C\exp(C(s-t)^{\frac12})\Big(|(\mathbf{x}-\mathbf{x}')_{n-1}| + (s-t)^{n-\frac32} + \int_t^s dv\big|\big(\boldsymbol{\theta}_{v,t}(\mathbf{x})-\boldsymbol{\theta}_{v,t}(\mathbf{x}')\big)\big)_{n-2}\big|\\
&+|(\mathbf{x}-\mathbf{x}')_n|^{\frac{2n-3}{2n-1}} + \sup_{w\in[t,s]}\big|\big(\boldsymbol{\theta}_{w,t}(\mathbf{x})-\boldsymbol{\theta}_{w,t}(\mathbf{x}')\big)\big)_{n-1}\big|^{\beta_n}(s-t)^{\beta_n+1}\Big)\Big).
\end{aligned}\tag{A.8}$$

From Young's inequality we now derive:

$$\begin{aligned}\sup_{w\in[t,s]}|\big(\boldsymbol{\theta}_{w,t}(\mathbf{x})-\boldsymbol{\theta}_{w,t}(\mathbf{x}')\big)_{n-1}|^{\beta_n}(s-t)^{\beta_n+1} &\leq \frac{1}{2}\sup_{w\in[t,s]}|\big(\boldsymbol{\theta}_{w,t}(\mathbf{x})-\boldsymbol{\theta}_{w,t}(\mathbf{x}')\big)_{n-1}|(s-t)+C(s-t)^{\frac{1}{1-\beta_n}}\\ &\leq \frac{1}{2}\sup_{w\in[t,s]}|\big(\boldsymbol{\theta}_{w,t}(\mathbf{x})-\boldsymbol{\theta}_{w,t}(\mathbf{x}')\big)_{n-1}|+C(s-t)^{n-\frac{3}{2}},\end{aligned}$$

recalling 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:

$$\begin{aligned}&\frac{1}{2}\sup_{\tilde{s}\in[t,s]}|(\boldsymbol{\theta}_{\tilde{s},t}(\mathbf{x})-\boldsymbol{\theta}_{\tilde{s},t}(\mathbf{x}'))_{n-1}| \qquad (A.9)\\ \leq\ & C\exp(C(s-t)^{\frac{1}{2}})\Big(|(\mathbf{x}-\mathbf{x}')_{n-1}|+(s-t)^{n-\frac{3}{2}}+\int_t^s dv|\big(\boldsymbol{\theta}_{v,t}(\mathbf{x})-\boldsymbol{\theta}_{v,t}(\mathbf{x}')\big)_{n-2}|+|(\mathbf{x}-\mathbf{x}')_n|^{\frac{2n-3}{2n-1}}\Big).\end{aligned}$$

We explicitly see from (A.10) that each entry of the difference of the starting points appears at its intrinsic scale for the homogeneous distance $\mathbf{d}$ introduced in (A.1).

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

$$\begin{aligned}&|(\boldsymbol{\theta}_{s,t}(\mathbf{x})-\boldsymbol{\theta}_{s,t}(\mathbf{x}'))_n|\\ &\leq C\exp\Big(C(s-t)^{\frac{1}{2}}\Big)\Big(|(\mathbf{x}-\mathbf{x}')_n|+(s-t)^{n-\frac{1}{2}}+|(\mathbf{x}-\mathbf{x}')_{n-1}|(s-t)\\ &\quad+|(\mathbf{x}-\mathbf{x}')_n|^{\frac{2n-3}{2n-1}}(s-t)+\int_t^s dv\int_t^v dw|\big(\boldsymbol{\theta}_{w,t}(\mathbf{x})-\boldsymbol{\theta}_{w,t}(\mathbf{x}')\big)_{n-2}|\Big)\\ &\leq C\exp\Big(C(s-t)^{\frac{1}{2}}\Big)\Big(|(\mathbf{x}-\mathbf{x}')_n|+(s-t)^{n-\frac{1}{2}}+|(\mathbf{x}-\mathbf{x}')_{n-1}|^{\frac{2n-1}{2n-3}}+\int_t^s dv\int_t^v dw|\big(\boldsymbol{\theta}_{w,t}(\mathbf{x})-\boldsymbol{\theta}_{w,t}(\mathbf{x}')\big)_{n-2}|\Big),\end{aligned}$$

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

$$|(\boldsymbol{\theta}_{s,t}(\mathbf{x})-\boldsymbol{\theta}_{s,t}(\mathbf{x}'))_n|\leq C\bigg((s-t)^{n-\frac{1}{2}}+\sum_{j=2}^n|(\mathbf{x}-\mathbf{x}')_j|^{\frac{2n-1}{2j-1}}+\int_t^{v_n=s}dv_{n-1}\dots\int_t^{v_2}dv_1|\big(\boldsymbol{\theta}_{v_1,t}(\mathbf{x}')-\boldsymbol{\theta}_{v_1,t}(\mathbf{x})\big)_1|\bigg).$$

Anagolously, for $i\in[\![2,n]\!]$, we obtain:

$$|(\boldsymbol{\theta}_{s,t}(\mathbf{x})-\boldsymbol{\theta}_{s,t}(\mathbf{x}'))_i|\leq C\bigg((s-t)^{i-\frac{1}{2}}+\sum_{j=2}^n|(\mathbf{x}-\mathbf{x}')_j|^{\frac{2i-1}{2j-1}}+\int_t^{v_i=s}dv_{i-1}\dots\int_t^{v_2}dv_1|\big(\boldsymbol{\theta}_{v_1,t}(\mathbf{x}')-\boldsymbol{\theta}_{v_1,t}(\mathbf{x})\big)_1|\bigg).$$

Remark 8. *Observe that equations* (A.10) *and* (A.10) *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:

$$|(\boldsymbol{\theta}_{\tilde{s},t}(\mathbf{x})-\boldsymbol{\theta}_{\tilde{s},t}(\mathbf{x}'))_1|\leq|(\mathbf{x}-\mathbf{x}')_1|+C\sum_{j=1}^n\int_t^{\tilde{s}}dv|(\boldsymbol{\theta}_{v,t}(\mathbf{x})-\boldsymbol{\theta}_{v,t}(\mathbf{x}'))_j|^{\beta_j},$$

which in turn implies from (A.10), Remark 8 and convexity inequalities:

$$\begin{aligned}&\sup_{\tilde{s}\in[t,s]}|(\boldsymbol{\theta}_{\tilde{s},t}(\mathbf{x})-\boldsymbol{\theta}_{\tilde{s},t}(\mathbf{x}'))_1|\\ \leq\ & |(\mathbf{x}-\mathbf{x}')_1|+C\Big((s-t)\sup_{v\in[t,s]}|(\boldsymbol{\theta}_{v,t}(\mathbf{x})-\boldsymbol{\theta}_{v,t}(\mathbf{x}'))_1|^{\beta_1}+\sum_{j=2}^n\int_t^s dv|(\boldsymbol{\theta}_{v,t}(\mathbf{x})-\boldsymbol{\theta}_{v,t}(\mathbf{x}'))_j|^{\beta_j}\Big)\\ \leq\ & |(\mathbf{x}-\mathbf{x}')_1|+\bigg((s-t)\Big[\frac{1}{2}\sup_{v\in[t,s]}|(\boldsymbol{\theta}_{\tilde{s},t}(\mathbf{x})-\boldsymbol{\theta}_{\tilde{s},t}(\mathbf{x}'))_1|+C\Big]\\ &+C\sum_{j=2}^n(s-t)\Big((s-t)^{j-\frac{1}{2}}+\sum_{k=2}^n|(\mathbf{x}-\mathbf{x}')_k|^{\frac{2j-1}{2k-1}}+(s-t)^{j-1}\sup_{v\in[t,s]}|(\boldsymbol{\theta}_{v,t}(\mathbf{x})-\boldsymbol{\theta}_{v,t}(\mathbf{x}'))_1|\Big)^{\beta_j}\bigg).\end{aligned}$$

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

$$\begin{aligned}
&\frac{1}{2}\sup_{\tilde{s}\in[t,s]}|(\boldsymbol{\theta}_{\tilde{s},t}(\mathbf{x})-\boldsymbol{\theta}_{\tilde{s},t}(\mathbf{x}'))_1| \\
\le\ & |(\mathbf{x}-\mathbf{x}')_1| \qquad\qquad (A.10)\\
&+C\Big((s-t)+\sum_{j=2}^{n}(s-t)\Big((s-t)^{\frac{1}{2}}+\sum_{k=2}^{n}|(\mathbf{x}-\mathbf{x}')_k|^{\frac{1}{2k-1}}+(s-t)^{(j-1)\beta_j}\sup_{v\in[t,s]}|(\boldsymbol{\theta}_{v,t}(\mathbf{x})-\boldsymbol{\theta}_{v,t}(\mathbf{x}'))_1|^{\beta_j}\Big)\Big).
\end{aligned}$$

Write now from the Young inequality:

$$C(s-t)^{1+(j-1)\beta_j}\sup_{v\in[t,s]}|(\boldsymbol{\theta}_{v,t}(\mathbf{x})-\boldsymbol{\theta}_{v,t}(\mathbf{x}'))_1|^{\beta_j} \le C(s-t)+\frac{1}{4}\sup_{v\in[t,s]}|(\boldsymbol{\theta}_{v,t}(\mathbf{x})-\boldsymbol{\theta}_{v,t}(\mathbf{x}'))_1|.$$

We eventually derive from (A.11) that:

$$\sup_{v\in[t,s]}|(\boldsymbol{\theta}_{v,t}(\mathbf{x})-\boldsymbol{\theta}_{v,t}(\mathbf{x}'))_1| \le C\big((s-t)^{\frac{1}{2}}+\mathbf{d}(\mathbf{x},\mathbf{x}')\big),$$

which gives the statement for $i=1$. Plugging now this inequality into (A.10), we get for each $i\in[\![2,n]\!]$:

$$\begin{aligned}
|(\boldsymbol{\theta}_{s,t}(\mathbf{x})-\boldsymbol{\theta}_{s,t}(\mathbf{x}'))_i| &\le C\big((s-t)^{i-\frac{1}{2}}+\mathbf{d}^{2i-1}(\mathbf{x},\mathbf{x}')+(s-t)^{i-1}\sup_{v\in[t,s]}|(\boldsymbol{\theta}_{v,t}(\mathbf{x})-\boldsymbol{\theta}_{v,t}(\mathbf{x}'))_1|\big)\\
&\le C\Big((s-t)^{i-\frac{1}{2}}+\mathbf{d}^{2i-1}(\mathbf{x},\mathbf{x}')+(s-t)^{i-1}\big((s-t)^{\frac{1}{2}}+\mathbf{d}(\mathbf{x},\mathbf{x}')\big)\Big)\\
&\le C\big((s-t)^{i-\frac{1}{2}}+\mathbf{d}^{2i-1}(\mathbf{x},\mathbf{x}')\big),
\end{aligned}$$

using again for the last identity the Young inequality to derive that $(s-t)^{i-1}\mathbf{d}(\mathbf{x},\mathbf{x}') \le C\big((s-t)^{i-1/2}+\mathbf{d}^{2i-1}(\mathbf{x},\mathbf{x}')\big)$. The proof is complete. □

A.2 Sensitivity results for the mean: final proof of Lemma 14

Again through the analysis, we assume w.l.o.g. that $\mathbf{d}(\mathbf{x},\mathbf{x}') \le 1$. The control is done with a distinction of two contributions to handle.

$$\mathbf{m}^{\mathbf{x}}_{s,t}(\mathbf{x}')-\boldsymbol{\theta}_{s,t}(\mathbf{x}') = [\mathbf{m}^{\mathbf{x}}_{s,t}(\mathbf{x}')-\boldsymbol{\theta}_{s,t}(\mathbf{x})]+[\boldsymbol{\theta}_{s,t}(\mathbf{x})-\boldsymbol{\theta}_{s,t}(\mathbf{x}')]. \tag{A.11}$$

By the *proxy* definition in (2.3), we deduce that the mean value of $\tilde{\mathbf{X}}^{m,\boldsymbol{\xi}}_v$, $\mathbf{m}^{\boldsymbol{\xi}}_{v,t}$ is s.t.

$$\mathbf{m}^{\mathbf{x}}_{s,t}(\mathbf{x}')-\boldsymbol{\theta}_{s,t}(\mathbf{x}) = \mathbf{x}'-\mathbf{x}+\int_t^s dvD\mathbf{F}(v,\boldsymbol{\theta}_{v,t}(\mathbf{x}))[\mathbf{m}^{\mathbf{x}}_{v,t}(\mathbf{x}')-\boldsymbol{\theta}_{v,t}(\mathbf{x})]. \tag{A.12}$$

The sub-triangular structure of $D\mathbf{F}$ yields that for each $i\in[\![2,n]\!]$:

$$\big(\mathbf{m}^{\mathbf{x}}_{s,t}(\mathbf{x}')-\boldsymbol{\theta}_{s,t}(\mathbf{x})\big)_i = \mathbf{x}'_i-\mathbf{x}_i+\int_t^s dvD_{i-1}\mathbf{F}_i(v,\boldsymbol{\theta}_{v,t}(\mathbf{x}))[\mathbf{m}^{\mathbf{x}}_{v,t}(\mathbf{x}')_{i-1}-\boldsymbol{\theta}_{v,t}(\mathbf{x})_{i-1}].$$

Also, since $\mathbf{m}^{\mathbf{x}}_{v,t}(\mathbf{x}')_1 = \mathbf{x}'_1+\int_t^s dv\mathbf{F}_1(v,\boldsymbol{\theta}_{v,t}(\mathbf{x}))$, so we obtain that $[\mathbf{m}^{\mathbf{x}}_{v,t}(\mathbf{x}')_1-\boldsymbol{\theta}_{v,t}(\mathbf{x})_1] = \mathbf{x}'_1-\mathbf{x}_1$, we then obtain by iteration that:

$$\big(\mathbf{m}^{\mathbf{x}}_{s,t}(\mathbf{x}')-\boldsymbol{\theta}_{s,t}(\mathbf{x})\big)_i = \mathbf{x}'_i-\mathbf{x}_i+\sum_{k=2}^{i}\Big[\int_t^{v_i=s}dv_{i-1}\ldots\int_t^{v_k}dv_{k-1}\prod_{j=k}^{i}D_{j-1}\mathbf{F}_j(v_j,\boldsymbol{\theta}_{v_j,t}(\mathbf{x}))\Big][\mathbf{x}'_{k-1}-\mathbf{x}_{k-1}],$$

with the convention that for $i=1$, $\sum_{k=2}^{i}=0$. From the above control, equation (A.11) and the dynamics of the flow, and because the starting points are the same, the contributions involving differences of the spatial points $(\mathbf{x}'-\mathbf{x})$ or flows only appear in iterated time integrals, we obtain:

$$\begin{aligned}
&|\big(\mathbf{m}^{\mathbf{x}}_{s,t}(\mathbf{x}')-\boldsymbol{\theta}_{s,t}(\mathbf{x}')\big)_i|\\
\le\ &\Big|\sum_{k=2}^{i}\Big[\int_t^{v_i=s}dv_{i-1}\ldots\int_t^{v_k}dv_{k-1}\prod_{j=k}^{i}D_{j-1}\mathbf{F}_j(v_j,\boldsymbol{\theta}_{v_j,t}(\mathbf{x}))\Big][\mathbf{x}'_{k-1}-\mathbf{x}_{k-1}]\Big|\\
&+\int_t^s|\mathbf{F}_i(v,\boldsymbol{\theta}_{v,t}(\mathbf{x}))-\mathbf{F}_i(v,\boldsymbol{\theta}_{v,t}(\mathbf{x}'))|dv\\
\le\ &C\Big(\sum_{k=2}^{i-1}(s-t)^{i-k}|\mathbf{x}_k-\mathbf{x}'_k|+\int_t^s dv\Big(\sum_{j=i}^{n}\big|\big(\boldsymbol{\theta}_{v,t}(\mathbf{x})-\boldsymbol{\theta}_{v,t}(\mathbf{x}')\big)_j\big|^{\beta_j}+\big|\big(\boldsymbol{\theta}_{v,t}(\mathbf{x})-\boldsymbol{\theta}_{v,t}(\mathbf{x}')\big)_{i-1}\big|\Big)\Big).
\end{aligned}$$

We derive from the previous Lemma 15 (control of the flows) recalling again that $\beta_j > (2j-2)/(2j-1)$ and $\mathbf{d}(\mathbf{x},\mathbf{x}') \le 1, 0 \le t < s \le T \le 1$:

$$
\begin{aligned}
&|\big(\mathbf{m}^{\mathbf{x}}_{s,t}(\mathbf{x}') - \boldsymbol{\theta}_{s,t}(\mathbf{x}')\big)_i| \\
&\le C\Big(\sum_{k=2}^{i-1}(s-t)^{i-k}|\mathbf{x}_k - \mathbf{x}'_k| + (s-t)^{\frac{2i-2}{2}+1} + \mathbf{d}^{2i-2}(\mathbf{x},\mathbf{x}')(s-t) + \big((s-t)^{(i-1)-\frac{1}{2}} + \mathbf{d}^{2(i-1)-1}(\mathbf{x},\mathbf{x}')\big)(s-t)\Big).
\end{aligned}
$$

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

$$
|\big(\mathbf{m}^{\mathbf{x}}_{t_0,t}(\mathbf{x}') - \boldsymbol{\theta}_{t_0,t}(\mathbf{x}')\big)_i| \le C\Big(c_0\mathbf{d}^{2i-1}(\mathbf{x},\mathbf{x}') + (c_0^i + c_0)\mathbf{d}^{2i-1}(\mathbf{x},\mathbf{x}') + (c_0^{i-\frac{1}{2}} + c_0)\mathbf{d}^{2i-1}(\mathbf{x},\mathbf{x}')\Big),
$$

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

$$
\mathbf{d}\big(\mathbf{m}^{\mathbf{x}}_{s,t}(\mathbf{x}'), \boldsymbol{\theta}_{s,t}(\mathbf{x}')\big) \le C c_0^{\frac{1}{2n-1}}\mathbf{d}(\mathbf{x},\mathbf{x}').
$$

This concludes the proof of Lemma 14. □

B Parametrix expansion with different freezing points

In this section we show how the parametrix expansion (3.19) involving different freezing points can be derived. This can actually been done from the Duhamel formulation up to an additional discontinuity term. Restarting from (2.20) we can indeed rewrite from the Markov property that for given $(t,\mathbf{x}') \in [0,T]\times\mathbb{R}^{nd}$ and any $r \in (t,T], \boldsymbol{\xi}' \in \mathbb{R}^{nd}$:

$$
u(t,\mathbf{x}') = \Big[\tilde{P}^{\boldsymbol{\xi}'}_{r,t}u(r,\cdot)\Big](\mathbf{x}') + \int_t^r ds\Big[\tilde{P}^{\boldsymbol{\xi}'}_{s,t}f(s,\cdot)\Big](\mathbf{x}') + \int_t^r ds\Big[\tilde{P}^{\boldsymbol{\xi}'}_{s,t}\Big((L_s - \tilde{L}^{\boldsymbol{\xi}'}_s)u\Big)(s,\cdot)\Big](\mathbf{x}'). \tag{B.1}
$$

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

$$
0 = \partial_r\Big[\tilde{P}^{\boldsymbol{\xi}'}_{r,t}u(r,\cdot)\Big](\mathbf{x}') + \Big[\tilde{P}^{\boldsymbol{\xi}'}_{r,t}f(r,\cdot)\Big](\mathbf{x}') + \Big[\tilde{P}^{\boldsymbol{\xi}'}_{r,t}\Big((L_r - \tilde{L}^{\boldsymbol{\xi}'}_r)u\Big)(r,\cdot)\Big](\mathbf{x}'). \tag{B.2}
$$

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

$$
\begin{aligned}
0 &= \Big[\tilde{P}^{\boldsymbol{\xi}'}_{t_0,t}u(t_0,\cdot)\Big](\mathbf{x}') - u(t,\mathbf{x}') + \int_t^{t_0} ds\Big[\tilde{P}^{\boldsymbol{\xi}'}_{s,t}f(s,\cdot)\Big](\mathbf{x}') + \int_t^{t_0} ds\Big[\tilde{P}^{\boldsymbol{\xi}'}_{s,t}\Big((L_s - \tilde{L}^{\boldsymbol{\xi}'}_s)u\Big)(s,\cdot)\Big](\mathbf{x}') \\
&\quad + \Big[\tilde{P}^{\tilde{\boldsymbol{\xi}}'}_{T,t}u(T,\cdot)\Big](\mathbf{x}') - \Big[\tilde{P}^{\tilde{\boldsymbol{\xi}}'}_{t_0,t}u(t_0,\cdot)\Big](\mathbf{x}') + \int_{t_0}^T ds\Big[\tilde{P}^{\tilde{\boldsymbol{\xi}}'}_{s,t}f(s,\cdot)\Big](\mathbf{x}') + \int_{t_0}^T ds\Big[\tilde{P}^{\tilde{\boldsymbol{\xi}}'}_{s,t}\Big((L_s - \tilde{L}^{\tilde{\boldsymbol{\xi}}'}_s)u\Big)(s,\cdot)\Big](\mathbf{x}').
\end{aligned}
$$

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

$$
\begin{aligned}
&u(t,\mathbf{x}') \\
&= \int_t^T ds\Big(\mathbb{I}_{s\le t_0}\Big[\tilde{P}^{\boldsymbol{\xi}'}_{s,t}f(s,\cdot)\Big](\mathbf{x}') + \mathbb{I}_{s>t_0}\Big[\tilde{P}^{\tilde{\boldsymbol{\xi}}'}_{s,t}f(s,\cdot)\Big](\mathbf{x}')\Big) + \Big[\tilde{P}^{\boldsymbol{\xi}'}_{t_0,t}u(t_0,\cdot)\Big](\mathbf{x}') - \Big[\tilde{P}^{\tilde{\boldsymbol{\xi}}'}_{t_0,t}u(t_0,\cdot)\Big](\mathbf{x}') \\
&\quad + \int_t^T ds\Big(\mathbb{I}_{s\le t_0}\Big[\tilde{P}^{\boldsymbol{\xi}'}_{s,t}\Big((L_s - \tilde{L}^{\boldsymbol{\xi}'}_s)u\Big)(s,\cdot)\Big](\mathbf{x}') + \mathbb{I}_{s>t_0}\Big[\tilde{P}^{\tilde{\boldsymbol{\xi}}'}_{s,t}\Big((L_s - \tilde{L}^{\tilde{\boldsymbol{\xi}}'}_s)u\Big)(s,\cdot)\Big](\mathbf{x}')\Big).
\end{aligned}
$$

We see that for $\boldsymbol{\xi}' \neq \tilde{\boldsymbol{\xi}}'$ we have an additional discontinuity term deriving from the change of freezing point along the time variable. Eventually, the above equation precisely gives (3.19), recalling that for $t_0 = t + c_0|(\mathbf{x}_i - z)|^{2/(2i-1)}$, $\mathbb{I}_{s\le t_0} = \mathbb{I}_{\mathcal{S}_i}$.

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 Section 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 [DM10] and those of the current article.

- Notations and Assumptions from [DM10]. Consider the Gaussian process with dynamics

$$d\mathbf{G}_t = \mathbf{L}_t dt + B\Sigma_t dW_t \tag{LIN}$$

where $(\Sigma_t)_{t\in[0,T]}$ is a measurable deterministic $\mathbb{R}^d \otimes \mathbb{R}^d$ valued family s.t. $\mathbf{A}_t := \Sigma_t\Sigma_t^*$ has uniformly non-degenerate spectrum, i.e. there exists $\Lambda \geq 1$ s.t. for any $t \in [0,T]$, $\mathrm{Spec}(\mathbf{A}_t) \in [\Lambda^{-1}, \Lambda]$, and the measurable deterministic $\mathbb{R}^{nd} \otimes \mathbb{R}^{nd}$ valued family $(\mathbf{L}_t)_{t\in[0,T]}$ is such that for any $t \in [0,T]$:

$$\mathbf{L}_t = \begin{pmatrix} U_t^{1,1} & U_t^{1,2} & \cdots & \cdots & U_t^{1,n} \\ \alpha_t^1 & U_t^{2,2} & \cdots & \cdots & U_t^{2,n} \\ \mathbf{0}_{d,d} & \alpha_t^2 & U_t^{3,3} & \cdots & U_t^{3,n} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0}_{d,d} & \cdots & \mathbf{0}_{d,d} & \alpha_t^{n-1} & U_t^{n,n} \end{pmatrix}, \tag{C.1}$$

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

This is assumption $(\mathbf{A}^{\mathrm{linear}})$ in [DM10]. Proposition 3.4 of that reference states that, whenever for all $i \in [\![1, n-1]\!]$ and $t \in [0,T]$, α_t^i belongs to $\mathcal{E}_i$ (closed convex subset of $GL_d(\mathbb{R})$), there exists a constant $c \geq 1$ depending on $\mathcal{E}_i$, Λ, κ s.t. $\max_{i\in[\![1,n-1]\!]} \sup_{t\in[0,T]} |\alpha_t^i| \leq \kappa$, n, d such that the Gaussian process $(\mathbf{G}_t)_{t\in[0,T]}$ introduced in (LIN) satisfies a good scaling property with constant c in the sense of Definition 3.2 of [DM10]. Precisely, denoting by $\big(\mathbf{R}(s,t)\big)_{0\leq t,s\leq T}$ the resolvent matrix associated with $(\mathbf{L}_t)_{t\geq 0}$, the covariance matrix

$$\mathbf{K}_t = \int_0^t ds \mathbf{R}_{t,s} B \mathbf{A}_s B^* \mathbf{R}_{t,s}^*$$

of the random variable $\mathbf{G}_t$ satisfies that for any $\mathbf{y} \in \mathbb{R}^{nd}$,

$$c^{-1}t^{-1}|\mathbb{T}_t\mathbf{y}|^2 \leq \langle \mathbf{K}_t\mathbf{y}, \mathbf{y}\rangle \leq ct^{-1}|\mathbb{T}_t\mathbf{y}|^2,$$

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

$$c^{-1}t|\mathbb{T}_t^{-1}\mathbf{y}|^2 \leq \langle \mathbf{K}_t^{-1}\mathbf{y}, \mathbf{y}\rangle \leq ct|\mathbb{T}_t^{-1}\mathbf{y}|^2.$$

- Derivation of the Proposition 4 from the previous results of [DM10]. From the dynamics of the process $(\tilde{\mathbf{X}}_v^{(\tau,\boldsymbol{\xi})})_{v\in[t,T]}$ given in (2.3) and starting from $\mathbf{x}$ at time t, see also the associated integrated expression in (2.5), it can be seen that the covariance matrix of the random variable $\tilde{\mathbf{X}}_v^{(\tau,\boldsymbol{\xi})}$ writes, with the notations of Section 2.1:

$$\tilde{\mathbf{K}}_{v,t}^{(\tau,\boldsymbol{\xi})} := \int_t^v du \tilde{\mathbf{R}}^{(\tau,\boldsymbol{\xi})}(v,u) B a(u, \boldsymbol{\theta}_{u,\tau}(\boldsymbol{\xi})) B^* \tilde{\mathbf{R}}^{(\tau,\boldsymbol{\xi})}(v,u)^*,$$

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

$$\mathbf{L}_r = \begin{pmatrix} \mathbf{0}_{d,d} & \cdots & \cdots & \cdots & \mathbf{0}_{d,d} \\ D_{\mathbf{z}_1}\mathbf{F}_2(t+r, \boldsymbol{\theta}_{t+r,\tau}(\boldsymbol{\xi})) & \mathbf{0}_{d,d} & \cdots & \cdots & \mathbf{0}_{d,d} \\ \mathbf{0}_{d,d} & D_{\mathbf{z}_2}\mathbf{F}_3(t+r, \boldsymbol{\theta}_{t+r,\tau}^{2:n}(\boldsymbol{\xi})) & \mathbf{0}_{d,d} & \mathbf{0}_{d,d} & \vdots \\ \vdots & \mathbf{0}_{d,d} & \ddots & \ddots & \vdots \\ \mathbf{0}_{d,d} & \cdots & \mathbf{0}_{d,d} & D_{\mathbf{z}_{n-1}}\mathbf{F}_n(t+r, \boldsymbol{\theta}_{t+r,\tau}^{n-1:n}(\boldsymbol{\xi})) & \mathbf{0}_{d,d} \end{pmatrix}, \tag{C.2}$$

and

$$\Sigma_r = \sigma(t+r, \boldsymbol{\theta}_{t+r,\tau}(\boldsymbol{\xi})).$$

Since the resolvent $(\mathbf{R}(r,v))_{v\in[0,s-t]}$ associated with $(\mathbf{L}_r)_{r\in[0,s-t]}$ writes $\mathbf{R}(r,v) = \tilde{\mathbf{R}}^{(\tau,\boldsymbol{\xi})}(t+r, t+v)$, one readily derives that the covariance matrix $\mathbf{K}_{s-t}$ of $\mathbf{G}_{t-s} = \mathbf{R}(t-s,0)\mathbf{x} + \int_0^{t-s} \mathbf{R}(t-s,r)B\Sigma_r dW_r$ coincides with $\tilde{\mathbf{K}}_{s,t}^{(\tau,\boldsymbol{\xi})}$. Since $\mathbf{K}_{s-t} = \tilde{\mathbf{K}}_{s,t}^{(\tau,\boldsymbol{\xi})}$ 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.

$$\forall \boldsymbol{\zeta} \in \mathbb{R}^{nd}, |\big[\widehat{\widehat{\mathbf{R}}}^{(\tau,\boldsymbol{\xi}),s,t}(1,0)\big]^* \boldsymbol{\zeta}| \leq \hat{C}_1 |\boldsymbol{\zeta}|. \tag{C.3}$$

We will proceed following the arguments of Lemma 3.6 in [DM10] 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 (LIN) and introduce the *scaled process* $\hat{\mathbf{G}}^t_s = t^{1/2}\mathbb{T}_t^{-1}\mathbf{G}_{st},\ s \in [0,1]$. Intuitively, from the previously established *good scaling property of* $\mathbf{G}$, all the components of the process $\hat{\mathbf{G}}^t$ 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 $\hat{\mathbf{G}}^t_s$ satisfies (LIN), i.e. $d\hat{\mathbf{G}}^t_s = \hat{\mathbf{L}}^t_s \hat{\mathbf{G}}^t_s + \hat{\Sigma}^t_s d\hat{W}^t_s$ with:

$$\hat{\mathbf{L}}^t_s = t\mathbb{T}_t^{-1}\mathbf{L}_{st}\mathbb{T}_t,\ \hat{\Sigma}^t_s = \Sigma_{st},\ \hat{W}^t_s = t^{-1/2}W_{st},\ s \in [0,1].$$

From the general structure of $\mathbf{L}$ in (C.1), and the definition of the scale matrix $\mathbb{T}_t$, one easily derives that $|\hat{\mathbf{L}}^t_s| \leq (1 \vee T^n)|\mathbf{L}_{st}|$. More specifically, for the special structure considered in (C.2), i.e. without upper-triangular part, we even get $\hat{\mathbf{L}}^t_s = \mathbf{L}_{st}$. In any case, we obtain $|\hat{\mathbf{L}}^t_s| \leq (1 \vee T^n)\kappa$ as soon as $\sup_{s\in[0,1]} |\mathbf{L}_{st}| \leq \kappa$.

It is then clear that denoting by $\hat{\mathbf{R}}^t$ the resolvent associated with $(\hat{\mathbf{L}}^t_s)_{s\in[0,1]}$ 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

$$\hat{\mathbf{R}}^t(s_1,s_0) = \mathbb{T}_t^{-1}\mathbf{R}(s_1 t, s_0 t)\mathbb{T}_t \iff \mathbf{R}(s_1 t, s_0 t) = \mathbb{T}_t \hat{\mathbf{R}}^t(s_1,s_0)\mathbb{T}_t^{-1} \Longrightarrow [\mathbf{R}(t,0)]^* = \mathbb{T}_t^{-1}[\hat{\mathbf{R}}^t(1,0)]^*\mathbb{T}_t,$$

where $\mathbf{R}$ stands for the resolvent associated with $\mathbf{L}$. The final bound (C.3) on the rescaled resolvent associated with the frozen process $((s-t)^{1/2}\mathbb{T}_{s-t}^{-1}\tilde{\mathbf{X}}_v^{(\tau,\boldsymbol{\xi}),(t,\mathbf{x})})_{v\in[t,s]}$ is eventually derived from the same previous correspondence exhibited to prove the good scaling property of Proposition 4 in the previous paragraph.

C.2 Control of the Hölder modulus of the frozen Green kernel.

We aim here at proving, with the notations of Section 2.2 that the Hölder regularity index α^k_i of $\mathbf{y}_i \mapsto D_{\mathbf{y}_k}\tilde{G}^{i,\boldsymbol{\xi}} f(s,\mathbf{y}_i) := D_{\mathbf{y}_k}\tilde{G}^{\boldsymbol{\xi}} f(s,\mathbf{y}) = D_{\mathbf{y}_k}\int_s^T dr \tilde{P}_{r,s} f(s,\mathbf{y})$, with $i \leq k$, can be taken equal to $\alpha^k_i = \frac{1-(k-1/2)(1-\beta_k)}{i-1/2}$.

First, we get from (2.19) in Lemma 8 that:

$$|D_{\mathbf{y}_k}\tilde{G}^{i,\boldsymbol{\xi}} f(s,\mathbf{y}_i)| = |\int_s^T dr \int_{\mathbb{R}^{nd}} d\mathbf{z} D_{\mathbf{y}_k}\tilde{p}^{\boldsymbol{\xi}}(s,r,\mathbf{y},\mathbf{z}) f(s,\mathbf{z})| \leq C\sum_{j=k}^n \int_s^T dr (r-s)^{-(k-\frac{1}{2})+\beta_j(j-\frac{1}{2})}.$$

Recall from Assumption $(\mathbf{T}_\beta)$ that $\beta_j(j-1/2) \geq \beta_k(k-1/2)$ for $j \in [\![k,n]\!]$. The above equation thus yields:

$$\begin{aligned} |D_{\mathbf{y}_k}\tilde{G}^{i,\boldsymbol{\xi}} f(s,\mathbf{y}_i)| &= |\int_s^T dr \int_{\mathbb{R}^{nd}} d\mathbf{z} D_{\mathbf{y}_k}\tilde{p}^{\boldsymbol{\xi}}(s,r,\mathbf{y},\mathbf{z}) f(s,\mathbf{z})| \\ &\leq C\int_s^T dr (r-s)^{-(k-\frac{1}{2})+\beta_k(k-\frac{1}{2})} \\ &\leq C(T-s)^{1-(k-\frac{1}{2})(1-\beta_k)}. \end{aligned} \tag{C.4}$$

From this first bound we can then first draw a control of the Hölder modulus in the so-called *off-diagonal case* w.r.t. the current time scale of the considered variable i in the oscillator chain. Namely, for $\mathbf{y}_i, \mathbf{y}'_i \in \mathbb{R}^d$ s.t. $(T-s)/2 \leq |\mathbf{y}_i - \mathbf{y}'_i|^{1/(i-1/2)}$, equation (C.4) readily gives:

$$\begin{aligned} |D_{\mathbf{y}_k}\tilde{G}^{i,\boldsymbol{\xi}} f(s,\mathbf{y}_i) - D_{\mathbf{y}_k}\tilde{G}^{i,\boldsymbol{\xi}} f(s,\mathbf{y}'_i)| &\leq |D_{\mathbf{y}_k}\tilde{G}^{i,\boldsymbol{\xi}} f(s,\mathbf{y}_i)| + |D_{\mathbf{y}_k}\tilde{G}^{i,\boldsymbol{\xi}} f(s,\mathbf{y}'_i)| \\ &\leq 2C(T-s)^{1-(k-\frac{1}{2})(1-\beta_k)} \\ &\leq \tilde{C}|\mathbf{y}_i - \mathbf{y}'_i|^{\frac{1-(k-\frac{1}{2})(1-\beta_k)}{i-\frac{1}{2}}}. \end{aligned} \tag{C.5}$$

Assuming now that $(T-s)/2 \geq |\mathbf{y}_i - \mathbf{y}'_i|^{1/(i-1/2}$, we split the integrals as follows. Denoting with a slight abuse of notation $\mathbf{y}' := (\mathbf{y}_{1:i-1}, \mathbf{y}'_i, \mathbf{y}_{i+1:n})$,

$$\begin{aligned}
&|D_{\mathbf{y}_k}\tilde{G}^{i,\boldsymbol{\xi}}f(s,\mathbf{y}_i) - D_{\mathbf{y}_k}\tilde{G}^{i,\boldsymbol{\xi}}f(s,\mathbf{y}'_i)| \\
\leq\ & \Big|\int_s^{s+|\mathbf{y}_i-\mathbf{y}'_i|^{\frac{1}{i-\frac12}}} dr \int_{\mathbb{R}^{nd}} d\mathbf{z} D_{\mathbf{y}_k}\tilde{p}^{\boldsymbol{\xi}}(s,r,\mathbf{y},\mathbf{z}) f(s,\mathbf{z})\Big| + \Big|\int_s^{s+|\mathbf{y}_i-\mathbf{y}'_i|^{\frac{1}{i-\frac12}}} dr \int_{\mathbb{R}^{nd}} d\mathbf{z} D_{\mathbf{y}_k}\tilde{p}^{\boldsymbol{\xi}}(s,r,\mathbf{y}',\mathbf{z}) f(s,\mathbf{z})\Big| \\
&+\Big|\int_0^1 d\lambda \int_{s+|\mathbf{y}_i-\mathbf{y}'_i|^{\frac{1}{i-\frac12}}}^T \int_{\mathbb{R}^{nd}} d\mathbf{z} D_{\mathbf{y}_k} D_{\mathbf{y}_i}\tilde{p}^{\boldsymbol{\xi}}(s,r,\mathbf{y}'+\lambda(\mathbf{y}-\mathbf{y}'),\mathbf{z})\cdot(\mathbf{y}_i-\mathbf{y}'_i) f(s,\mathbf{z})\Big|.
\end{aligned}$$

For the two first term, we have a *local off-diagonal regime* within the global *diagonal* one. Applying (C.4) with T replaced by $s+|\mathbf{y}_i-\mathbf{y}'_i|^{1/(i-1/2)}$, we get:

$$\begin{aligned}
&|D_{\mathbf{y}_k}\tilde{G}^{i,\boldsymbol{\xi}}f(s,\mathbf{y}_i) - D_{\mathbf{y}_k}\tilde{G}^{i,\boldsymbol{\xi}}f(s,\mathbf{y}'_i)| \\
\leq\ & C|\mathbf{y}_i-\mathbf{y}'_i|^{\frac{1-(k-\frac12)(1-\beta_k)}{i-\frac12}} + \Big|\int_0^1 d\lambda \int_{s+|\mathbf{y}_i-\mathbf{y}'_i|^{\frac{1}{i-\frac12}}}^T dr \int_{\mathbb{R}^{nd}} d\mathbf{z} D_{\mathbf{y}_k} D_{\mathbf{y}_i}\tilde{p}^{\boldsymbol{\xi}}(s,r,\mathbf{y}'+\lambda(\mathbf{y}-\mathbf{y}'),\mathbf{z}) f(s,\mathbf{z})\Big|\ |\mathbf{y}_i-\mathbf{y}'_i|.
\end{aligned}$$

Using again (2.19) from the new Lemma 8, one eventually gets:

$$\begin{aligned}
&|D_{\mathbf{y}_k}\tilde{G}^{i,\boldsymbol{\xi}}f(s,\mathbf{y}_i) - D_{\mathbf{y}_k}\tilde{G}^{i,\boldsymbol{\xi}}f(s,\mathbf{y}'_i)| \\
\leq\ & C|\mathbf{y}_i-\mathbf{y}'_i|^{\frac{1-(k-\frac12)(1-\beta_k)}{i-\frac12}} + C\Big|\sum_{j=k}^n \int_{s+|\mathbf{y}_i-\mathbf{y}'_i|^{\frac{1}{i-\frac12}}}^T dr (r-s)^{-(k-\frac12)-(i-\frac12)+\beta_j(j-\frac12)}\Big|\ |\mathbf{y}_i-\mathbf{y}'_i| \\
\leq\ & C|\mathbf{y}_i-\mathbf{y}'_i|^{\frac{1-(k-\frac12)(1-\beta_k)}{i-\frac12}} + C\int_{s+|\mathbf{y}_i-\mathbf{y}'_i|^{\frac{1}{i-\frac12}}}^T dr (r-s)^{-(i-\frac12)-(k-\frac12)(1-\beta_k)}\ |\mathbf{y}_i-\mathbf{y}'_i| \\
\leq\ & C\Big(|\mathbf{y}_i-\mathbf{y}'_i|^{\frac{1-(k-\frac12)(1-\beta_k)}{i-\frac12}} + (r-s)^{-(i-\frac12)+1-(k-\frac12)(1-\beta_k)}\Big|_{r=s+|\mathbf{y}_i-\mathbf{y}'_i|^{\frac{1}{i-\frac12}}}^{r=T} |\mathbf{y}_i-\mathbf{y}'_i|\Big) \\
\leq\ & C|\mathbf{y}_i-\mathbf{y}'_i|^{\frac{1-(k-\frac12)(1-\beta_k)}{i-\frac12}}. \qquad \text{(C.6)}
\end{aligned}$$

Equations (C.5) and (C.6) precisely give the claim. Since $D_{\mathbf{y}_k}\tilde{G}^{i,\boldsymbol{\xi}}f(s,\mathbf{y}_i)$ is bounded (see (C.4)), any $\alpha_i^k \leq \frac{1-(k-1/2)(1-\beta_k)}{i-1/2}$ actually fits.

D 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. Let us decompose the expression around the variables for which Lemma 11 applies and those for which it does not. Namely, we write:

$$\begin{aligned}
&D_l u(s,\mathbf{y}) - D_l u(s,\mathbf{y}_{1:l-1},\boldsymbol{\theta}_{s,t}^{l:n}(\boldsymbol{\xi})) \\
=\ & \big(D_l u(s,\mathbf{y}) - D_l u(s,\mathbf{y}_{1:l-1},\boldsymbol{\theta}_{s,t}^{l}(\boldsymbol{\xi}),\mathbf{y}_{l+1:n})\big) + \big(D_l u(s,\mathbf{y}_{1:l-1},\boldsymbol{\theta}_{s,t}^{l}(\boldsymbol{\xi}),\mathbf{y}_{l+1:n}) - D_l u(s,\mathbf{y}_{1:l-1},\boldsymbol{\theta}_{s,t}^{l:n}(\boldsymbol{\xi}))\big) \\
=:\ & (\Delta_l^1+\Delta_l^2)(s,\mathbf{y},\boldsymbol{\xi}) \qquad \text{(D.7)}
\end{aligned}$$

We readily get:

$$|\Delta_l^1(s,\mathbf{y},\boldsymbol{\xi})| \leq C[(D_l u)_l(s,\cdot)]_{\alpha_l^l} |(\mathbf{y}-\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))_l|^{\alpha_l^l} \qquad \text{(D.8)}$$

which is controlled from Lemma 11. On the other hand, setting for any $\mathbf{a}\in\mathbb{R}^{(n-l+1)d}$, $\boldsymbol{\Xi}_{l,\mathbf{y},s}(\mathbf{a}) := D_l u(s,\mathbf{y}_{1:l-1},\mathbf{a})$, and $\mathbf{d} := \sum_{k=l+1}^n |(\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})-\mathbf{y})_k|$ we now have for any $\delta_l>0$

$$\begin{aligned}
\Delta_l^2(s,\mathbf{y},\boldsymbol{\xi}) &= \boldsymbol{\Xi}_{l,\mathbf{y},s}\big((\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))_l,\mathbf{y}_{l+1:n}\big) - \boldsymbol{\Xi}_{l,\mathbf{y},s}\big(\boldsymbol{\theta}_{s,t}^{l:n}(\boldsymbol{\xi})\big) \\
&= \int_0^1 d\mu\Big(\{D_l u(s,\mathbf{y}_{1:l-1},(\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))_l,\mathbf{y}_{l+1:n}) - D_l u(s,\mathbf{y}_{1:l-1},(\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))_l+\mu\mathbf{d}^{\delta_l},\mathbf{y}_{l+1:n})\} \\
&\quad +\{D_l u(s,\mathbf{y}_{1:l-1},(\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))_l+\mu\mathbf{d}^{\delta_l},\boldsymbol{\theta}_{s,t}^{l+1:n}(\boldsymbol{\xi})) - D_l u(s,\mathbf{y}_{1:l-1},\boldsymbol{\theta}_{s,t}^{l:n}(\boldsymbol{\xi}))\} \\
&\quad +\{D_l u(s,\mathbf{y}_{1:l-1},(\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))_l+\mu\mathbf{d}^{\delta_l},\mathbf{y}_{l+1:n}) - D_l u(s,\mathbf{y}_{1:l-1},(\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))_l+\mu\mathbf{d}^{\delta_l},\boldsymbol{\theta}_{s,t}^{l+1:n}(\boldsymbol{\xi}))\}\Big) \\
&=: \sum_{\ell=1}^3 \Delta_l^{2,\ell}(s,\mathbf{y},\boldsymbol{\xi}). \qquad \text{(D.9)}
\end{aligned}$$

The first two terms can be dealt directly.

$$|\Delta_l^{2,1}(s,\mathbf{y},\boldsymbol{\xi})| + |\Delta_l^{2,2}(s,\mathbf{y},\boldsymbol{\xi})| \leq 2[(D_l u)_l(s,\cdot)]_{\alpha_l^l}\mathbf{d}^{\delta_l\alpha_l^l}. \tag{D.10}$$

For $\Delta_l^{2,3}(s,\mathbf{y},\boldsymbol{\xi})$, we use an explicit reverse Taylor expansion which yields together with the smoothness assumption of $\mathbf{F}_i$ in **(A)**:

$$\begin{aligned}|\Delta_l^{2,3}(s,\mathbf{y},\boldsymbol{\xi})| &= \mathbf{d}^{-\delta_l}\Big|\Big[u(s,\mathbf{y}_{1:l-1},(\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))_l + \mathbf{d}^{\delta_l},\mathbf{y}_{l+1:n}) - u(s,\mathbf{y}_{1:l-1},(\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))_l,\mathbf{y}_{l+1:n}) \\ &\quad -\big(u(s,\mathbf{y}_{1:l-1},(\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))_l + \mathbf{d}^{\delta_l},(\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))^{l+1:n}) - u(s,\mathbf{y}_{1:l-1},(\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))_l,(\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi}))^{l+1:n})\big)\Big]\Big| \\ &\leq 2\|\mathbf{D}u\|_\infty \mathbf{d}^{1-\delta_l}, \end{aligned} \tag{D.11}$$

using the Lipschitz property of u w.r.t. the variables $l+1$ to n. Taking δ_l s.t. $\delta_l\alpha_l^l = 1-\delta_l^l$, which implies that $\delta_l = (1+\alpha_l^l)^{-1}$, gives in (D.10) and (D.11) a global bound of order $2(\|Du\|_\infty + [(D_l u)_l(s,\cdot)]_{\alpha_l^l})\mathbf{d}^{\delta_l\alpha_l^l}$. Since we now recall from Lemma 11 that $\alpha_l^l = (1+\eta/4)/(2l-1)$ so that

$$\delta_l\alpha_l^l = \frac{1}{\frac{2l-1+1+\frac{\eta}{4}}{2l-1}}\frac{1+\frac{\eta}{4}}{2l-1} = \frac{1+\frac{\eta}{4}}{2l+\frac{\eta}{4}} =: \zeta_l.$$

We then write from (D.9) and the definiton of $\mathbf{d}$ that:

$$|\Delta_l^2(s,\mathbf{y},\boldsymbol{\xi})| \leq 2(\|Du\|_\infty + [(D_l u)_l(s,\cdot)]_{\alpha_l^l})\sum_{k=l+1}^n |(\boldsymbol{\theta}_{s,t}(\boldsymbol{\xi})-\mathbf{y})_k|^{\zeta_l},$$

which together with (D.8) gives the result. □

Acknowledgments.

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.

References

[ABM20] S. Athreya, O. Butkovsky, and L. Mytnik. Strong existence and uniqueness for stable stochastic differential equations with distributional drift. *Ann. Probab.*, 48(1):178–210, 01 2020.

[AH96] D. R. Adams and L. I. Hedberg. *Function spaces and Potential Theory.* Springer, 1996.

[Bas97] R. F. Bass. *Diffusions and Elliptic Operators.* Springer, 1997.

[BC03] R. Bass and Z. Q. Chen. Brownian motion with singular drift. *Ann. Probab.*, 31– 2:791–817, 2003.

[BCLP10] M. Bramanti, G. Cupini, E. Lanconelli, and E. Priola. Global L^p estimates for degenerate Ornstein-Uhlenbeck operators. *Math. Z.*, 266(4):789–816, 2010.

[BCLP13] M. Bramanti, G. Cupini, E. Lanconelli, and E. Priola. Global L^p estimates for degenerate Ornstein-Uhlenbeck operators with variable coefficients. *Math. Nachr.*, 286(11-12):1087–1101, 2013.

[BFGM19] Lisa Beck, Franco Flandoli, Massimiliano Gubinelli, and Mario Maurelli. Stochastic odes and stochastic linear pdes with critical drift: regularity, duality and uniqueness. *Electronic Journal of Probability*, 24(0), 2019.

[BZ11] M. Bramanti and M. Zhu. L^p and Schauder estimates for nonvariational operators structured on Hörmander vector fields with drift. *Analysis of PDEs*, 2011.

[CC18] G. Cannizzaro and K. Chouk. Multidimensional sdes with singular drift and universal construction of the polymer measure with white noise potential. *Ann. Probab.*, 46(3):1710–1763, 05 2018.

[CdR17] P.-E. Chaudru de Raynal. Strong existence and uniqueness for degenerate SDE with Hölder drift. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 53(1):259–286, February 2017.

[CdR18] P.-E. Chaudru de Raynal. Weak regularization by stochastic drift: result and counter example. *Discrete and Continuous Dynamical Systems (Series A)*, 38–3:1269–1291, 2018.

[CdRHM20] P-E. Chaudru de Raynal, I. Honoré, and S. Menozzi. Sharp Schauder Estimates for some Degenerate Kolmogorov Equations. *arXiv:1810.12227 . To appear in Ann. Scuola Normale Superiore di Pisa, Scienze*, 2020.

[CdRM17] P-E. Chaudru de Raynal and S. Menozzi. Regularization effects of a noise propagating through a chain of differential equations: an almost sharp result. *arXiv:1710.03620. To appear in Trans. American Math. Society: https://doi.org/10.1090/tran/7947*, 2017.

[CdRM19] P.-E. Chaudru de Raynal and S. Menozzi. On multidimensional stable-driven stochastic differential equations with Besov drift. *arXiv 1907.12263*, 2019.

[CdRMP20a] P.-E. Chaudru de Raynal, S. Menozzi, and E. Priola. Schauder estimates for drifted fractional operators in the supercritical case. *Journal of Functional Analysis*, 278(8):108425, 2020.

[CdRMP20b] P.-E. Chaudru de Raynal, S. Menozzi, and E. Priola. Weak well-posedness of multidimensional stable driven SDEs in the critical case. *Stochastics and Dynamics*, 0(0):2040004, 2020.

[CG16] R. Catellier and M. Gubinelli. Averaging along irregular curves and regularisation of ODEs. *Stoch. Proc and Appl.*, 126–8:2323–2366, 2016.

[CZZ17] Z.-Q. Chen, X. Zhang, and G. Zhao. Well-posedness of supercritical SDE driven by L\'evy processes with irregular drifts. *arXiv:1709.04632*, Sep 2017.

[Dav07] A. M. Davie. Uniqueness of Solutions of Stochastic Differential Equations. *International Mathematics Research Notices*, 2007, 01 2007.

[DD15] F. Delarue and R. Diel. Rough paths and 1d SDE with a time dependent distributional drift: application to polymers. *Probability Theory and Related Fields*, pages 1–63, 2015.

[DF14] F. Delarue and F. Flandoli. The transition point in the zero noise limit for a 1d Peano example. *Discrete and Continuous Dynamical Systems*, 34(10):4071–4083, April 2014.

[DM10] F. Delarue and S. Menozzi. Density estimates for a random noise propagating through a chain of differential equations. *Journal of Functional Analysis*, 259–6:1577–1630, 2010.

[DPL89] R. Di Perna and P. L. Lions. Ordinary Differential Equations, transport theory and Sobolev Spaces. *Inv. Math.*, 98:511–547, 1989.

[FF11] E. Fedrizzi and F. Flandoli. Pathwise Uniqueness and Continuous Dependence for SDEs with Nonregular Drift. *Stochastics*, 83-3:241–257, 2011.

[FFPV17] E. Fedrizzi, F. Flandoli, E. Priola, and J. Vovelle. Regularity of stochastic kinetic equations. *Electronic Journal of Probability*, 22:♯ 48, 2017.

[FGP10] F. Flandoli, M. Gubinelli, and E. Priola. Well-posedness of the transport equation by stochastic perturbation. *Inventiones Mathematicae*, 180(1):1–53, 2010.

[FIR17] F. Flandoli, E. Issoglio, and F. Russo. Multidimensional stochastic differential equations with distributional drift. *Transactions of the American Mathematical Society*, 369(10.1090/tran/6729):1665–1688, 2017.

[Fla11a] F. Flandoli. *Random perturbation of PDEs and fluid dynamic models*, volume 2015 of *Lecture Notes in Mathematics*. Springer, Heidelberg, 2011. Lectures from the 40th Probability Summer School held in Saint-Flour, 2010.

[Fla11b] F. Flandoli. Regularizing properties of brownian paths and a result of davie. *Stochastics and Dynamics*, 11(02n03):323–331, 2011.

[Fri64] A. Friedman. *Partial differential equations of parabolic type*. Prentice-Hall, 1964.

[Fri08] A. Friedman. *Partial Differential Equations of Parabolic Type*. Dover Publications, April 2008.

[GO13] M. Gradinaru and Y. Offret. Existence and asymptotic behaviour of some time-inhomogeneous diffusions. *Annales Instit. H. Poincaré, Probabilités et Statistiques*, 49(1):182–207, 2013.

[GS69] I.I. Gikhman and A.V. Skorokhod. *Introduction to the theory of random processes*. Saunders Math. Books, 1969.

[HM16] L. Huang and S. Menozzi. A Parametrix Approach for some Degenerate Stable Driven SDEs. *Annales Instit. H. Poincaré*, 52(4):1925–1975, 2016.

[HMP19] L. Huang, S. Menozzi, and E. Priola. Lp estimates for degenerate non-local kolmogorov operators. *Journal de Mathématiques Pures et Appliquées*, 121:162 – 215, 2019.

[Hör67] L. Hörmander. Hypoelliptic second order differential equations. *Acta Mathematica*, 119:147–171, 1967.

[HP20] F.A. Harang and N. Perkowski. C^∞ regularization of ODEs perturbed by noise, 2020.

[HWZ20] Z. Hao, M. Wu, and X. Zhang. Schauder estimates for nonlocal kinetic equations and applications. *Journal de Mathématiques Pures et Appliquées*, 140:139 – 184, 2020.

[Kol34] A. N. Kolmogorov. Zufällige Bewegungen (zur Theorie der Brownschen Bewegung). *Ann. of Math.*, 2-35:116–117, 1934.

[KP10] N. V. Krylov and E. Priola. Elliptic and parabolic second-order PDEs with growing coefficients. *Comm. Partial Differential Equations*, 35(1):1–22, 2010.

[KR05] N. Krylov and M. Röckner. Strong solutions of stochastic equations with singular time dependent drift. *Prob. Theory Rel. Fields*, 131:154–196, 2005.

[Kry96] N. V. Krylov. *Lectures on elliptic and parabolic equations in Hölder spaces*. Graduate Studies in Mathematics 12. AMS, 1996.

[Kry01] N. V. Krylov. The heat equation in lq((0,t),lp)-spaces with weights. *SIAM Journal on Mathematical Analysis*, 32(5):1117–1141, 2001.

[Lor05] L. Lorenzi. 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.

[LR02] P.-G. Lemarié-Rieusset. *Recent developments in the Navier-Stokes problem*. CRC Press, 2002.

[LZ19] C. Ling and G. Zhao. Nonlocal elliptic equation in hölder space and the martingale problem, 2019.

[Mar20] L. Marino. Schauder estimates for degenerate stable Kolmogorov equations. *Bulletin des Sciences Mathématiques*, 162:102885, 2020.

[Men11] S. Menozzi. Parametrix techniques and martingale problems for some degenerate Kolmogorov equations. *Electronic Communications in Probability*, 17:234–250, 2011.

[Men18] S. Menozzi. Martingale problems for some degenerate Kolmogorov equations. *Stoc. Proc. Appl.*, 128-3:756–802, 2018.

[MP14] R. Mikulevicius and H. Pragarauskas. 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.

[MS67] H. P. McKean and I. M. Singer. Curvature and the eigenvalues of the Laplacian. *J. Differential Geometry*, 1:43–69, 1967.

[NO02] D. Nualart and Y. Ouknine. Regularization of differential equations by fractional noise. *Stochastic Processes and their Applications*, 102(1):103 – 116, 2002.

[Pri12] E. Priola. Pathwise uniqueness for singular SDEs driven by stable processes. *Osaka Journal of Mathematics*, 49(2):421–447, June 2012.

[Pri15] E. Priola. On weak uniqueness for some degenerate SDEs by global L^p estimates. *Potential Anal.*, 42(1):247–281, 2015.

[Pri18] E. Priola. Davie's type uniqueness for a class of sdes with jumps. *Ann. Inst. H. Poincaré Probab. Statist.*, 54(2):694–725, 05 2018.

[Pri19] E. Priola. On Davie's uniqueness for some degenerate SDEs, 2019.

[Sha16] A. V. Shaposhnikov. Some remarks on Davie's uniqueness theorem. *Proceedings of the Edinburgh Mathematical Society*, 59(4):1019–1035, 2016.

[Str08] D. W. Stroock. *Partial differential equations for probabilists*, volume 112 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2008.

[SV79] D.W. Stroock and S.R.S. Varadhan. *Multidimensional diffusion processes*. Springer-Verlag Berlin Heidelberg New-York, 1979.

[SW20] A. Shaposhnikov and L. Wresch. Pathwise vs. path-by-path uniqueness, 2020.

[Tri83] H. Triebel. *Theory of function spaces, II*. Birkhauser, 1983.

[TT90] D. Talay and L. Tubaro. Expansion of the global error for numerical schemes solving stochastic differential equations. *Stoch. Anal. and App.*, 8-4:94–120, 1990.

[Ver80] A. Y. Veretennikov. Strong solutions and explicit formulas for solutions of stochastic integral equations. *Matematicheski Sbornik. Novaya Seriya*, 111(153)(3):434–452, 480, 1980.

[Ver83] A. Y. Veretennikov. Stochastic equations with diffusion that degenerates with respect to part of the variables. *Izvestiya Akademii Nauk SSSR. Seriya Matematicheskaya*, 47(1):189–196, 1983.

[WZ16] F.Y. Wang and X. Zhang. Degenerate SDE with Holder-Dini drift and non-Lipschitz noise coefficient. *SIAM J. Math. Anal.*, 48(3):2189–2226, 2016.

[Zha10] X. Zhang. Well-posedness and large deviation for degenerate SDEs with Sobolev coefficients. *Revista Matematica Iberoamericana*, 29(1), 2010.

[Zha18] X. Zhang. Stochastic Hamiltonian flows with singular coefficients. *Sci. China Math.*, 61(8):1353–1384, 2018.

[Zvo74] A. K. Zvonkin. 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.

[ZZ17] X. Zhang and G. Zhao. Heat kernel and ergodicity of SDEs with distributional drifts. *arXiv:1710.10537*, 2017.