The Annals of Probability
2013, Vol. 41, No. 4, 2544–2598
DOI: 10.1214/12-AOP777

REGULARITY OF LAWS AND ERGODICITY OF HYPOELLIPTIC SDES DRIVEN BY ROUGH PATHS

BY MARTIN HAIRER[1] AND NATESH S. PILLAI[2]

University of Warwick and Harvard University

We consider differential equations driven by rough paths and study the regularity of the laws and their long time behavior. In particular, we focus on the case when the driving noise is a rough path valued fractional Brownian motion with Hurst parameter $H \in (\frac{1}{3}, \frac{1}{2}]$. Our contribution in this work is twofold.

First, when the driving vector fields satisfy Hörmander's celebrated "Lie bracket condition," we derive explicit quantitative bounds on the inverse of the Malliavin matrix. En route to this, we provide a novel "deterministic" version of Norris's lemma for differential equations driven by rough paths. This result, with the added assumption that the linearized equation has moments, will then yield that the transition laws have a smooth density with respect to Lebesgue measure.

Our second main result states that under Hörmander's condition, the solutions to rough differential equations driven by fractional Brownian motion with $H \in (\frac{1}{3}, \frac{1}{2}]$ enjoy a suitable version of the strong Feller property. Under a standard controllability condition, this implies that they admit a unique stationary solution that is physical in the sense that it does not "look into the future."

1. Introduction. In this article, we consider stochastic differential equations of the form

$$dZ_t = V_0(Z_t)\,dt + \sum_{i=1}^{d} V_i(Z_t)\,dX_t^i, \qquad Z_0 = z \in \mathbb{R}^n, \tag{1.1}$$

where X_t is a d-dimensional random rough path [16, 33, 34] and $V_0, V_i \in \mathbb{R}^n$ are smooth vector fields. While a large part of our work is deterministic and applies to a large class of rough differential equations driven by rough paths that are Hölder continuous with index greater than $\frac{1}{3}$, our probabilistic results focus on the case when X_t is a two-sided d-dimensional fractional Brownian motion with

Received April 2011; revised March 2012.

[1]Supported by EPSRC Grant EP/E002269/1, as well as by the Royal Society through a Wolfson Research Merit Award and by the Leverhulme Trust through a Philip Leverhulme Prize.

[2]Supported by NSF Grant DMS-11-07070.

MSC2010 subject classifications. Primary 60H07, 60H10; secondary 60G10, 26A33.

Key words and phrases. Hörmander's theorem, hypoellipticity, fractional Brownian motion, rough paths.

Hurst parameter $H \in (\frac{1}{3}, \frac{1}{2}]$. Recall that the fractional Brownian motion with Hurst parameter H is the centered Gaussian process such that $X_0 = 0$ and

$$\mathbb{E}|X_t - X_s|^2 = |t - s|^{2H}.$$

Differential equations driven by rough paths have been studied intensely in the past decade, and this theory has now reached a certain level of maturity; we refer to the monographs [16, 32, 34] for an overview of the theory. For driving signals that are rougher than Brownian motion, the theory of rough paths has provided a systematic way of constructing solutions to differential equations of the type (1.1) in a way that is "natural," in the sense that solutions are limits of approximate solutions where X is replaced by a smoothened version.

When the noise X_t is a replaced by a standard Brownian motion B_t, it has been well known since the groundbreaking work of Hörmander [24] that, for the laws of the Markov process Z_t to have a smooth transition density, it is sufficient that the Lie algebra formed by $\{\partial_t + V_0, V_1, \dots, V_d\}$ spans $\mathbb{R}^{n+1}$ at every point; see Assumption 2 for a precise formulation. The formalism of Malliavin calculus was invented to give a probabilistic proof of this result [28–30, 35, 36, 39]. The smoothness of transition densities coupled with some mild controllability assumptions will then yield that the system (1.1) has a unique invariant measure.

When the driving noise X is a fractional Brownian motion with $H \neq \frac{1}{2}$, solutions to (1.1) are neither a Markov process nor a semimartingale, so standard tools from stochastic calculus break down. Inspired by the results in the case of Brownian motion, two natural questions in this context are to identify conditions under which:

(1) the "transition densities" of (1.1) are smooth;
(2) the system (1.1) has a "unique invariant measure."

Since Z is not Markov in general, it does not really make sense to speak of transition probabilities, but the first question still makes sense by, for example, considering the law of the solution at some time $t > 0$, given an initial condition Z_0, conditional on the realization of $\{X_s : s \leq 0\}$. Similarly, the notion of an "invariant measure" does not make immediate sense for non-Markovian processes. This problem has been discussed extensively in [18, 19], where a notion of an invariant measure adapted to systems of the type (1.1) is introduced. Essentially, these are stationary solutions to (1.1) that are "physical" in the sense that they are independent of the innovation of X.

In recent years, the SDE (1.1) was studied when the driving noise X is a fractional Brownian motion with Hurst parameter $H > \frac{1}{2}$. In this case, the answers to both of the above questions are completely settled in a series of papers [3, 18, 22, 23, 40]. In particular, it was shown in [3, 23] that the solutions to (1.1) have smooth "transition densities" when the vector fields satisfy Hörmander's condition. It was also shown that if furthermore the control system associated to (1.1) is approximately controllable then, under suitable dissipativity and boundedness conditions on the vector fields V_i, (1.1) also admits a unique invariant measure. However, the question of smoothness of laws in the case of the driving noise X being a fractional Brownian motion with Hurst parameter $H < \frac{1}{2}$ was completely open until now, despite substantial recent progress in particular cases [10, 25]. The only general result in the context of rough paths theory was obtained in [6], where the authors obtained the existence of densities with respect to Lebesgue measure under Hörmander's condition for a large class of driving noises.

In this paper we largely settle the above two questions when X is a fractional Brownian motion with $H \in (\frac{1}{3}, \frac{1}{2}]$. An important component underlying the probabilistic proofs of the smoothness of transition densities is Norris's lemma [29, 38], which roughly speaking states that if a semimartingale is small and if one has a priori bounds on the regularity of its components, then its bounded variation part and the local martingale part are also small. In this regard this lemma can be considered as a quantitative version of the classical Doob–Meyer decomposition theorem. A version of Norris's lemma for fractional Brownian motion with $H > \frac{1}{2}$ was proved in [3]. The recent work [25], which appeared as the present article was nearing completion, contains a version of Norris's lemma for $H \in (\frac{1}{3}, \frac{1}{2})$ that is similar in spirit to the one in [3].

Our contribution in this work is twofold. First, we prove a *deterministic* version of Norris's lemma for general integrals against rough paths. This may sound strange at first since Norris's seems to be the prototype of a probabilistic statement, and the whole point of rough path theory is to get rid of stochastic calculus and replace it by a deterministic theory. We reconcile these conflicting perspectives by first showing an estimate strongly resembling that of Norris's lemma for processes of the form $Z_t = \int_0^t A_s\,dX_s + \int_0^t B_s\,ds$, where X is a rough path, and A is a rough path "controlled by X;" see Section 2 below for precise definitions. This estimate makes use of a quantity that we call the "modulus of Hölder roughness" of X, $L_\theta(X)$. See Definition 3 below for the precise definition of L_θ. In a second step, we then show that if X is fractional Brownian motion with $H \leq \frac{1}{2}$, then $L_\theta(X)$ is almost surely positive for $\theta > H$ and has inverse moments of all orders. A loose formulation of our main result is as follows (see Theorem 3.1 and Lemma 3 below for precise formulations that include the exact dependency of M on X, A and B):

THEOREM 1.1. *Let X be a γ-Hölder continuous rough path in $\mathbb{R}^n$ with $\gamma > \frac{1}{3}$, let A be a rough path in $\mathbb{R}^n$ controlled by X, let B be a γ-Hölder continuous function and set*

$$Z_t = \int_0^t A_s\,dX_s + \int_0^t B_s\,ds. \tag{1.2}$$

Then if X is θ-Hölder rough for some $\theta < 2\gamma$, there exist constants $r > 0$ and $q > 0$ such that one has the bound

$$\|A\|_\infty + \|B\|_\infty \leq M L_\theta(X)^{-q} \|Z\|_\infty^r$$

for a constant M depending polynomially on the γ-Hölder "norms" of X, A and B. Here, $\|\cdot\|_\infty$ denotes the supremum norm over the interval $[0, 1]$.

Furthermore, if X is the rough path canonically associated to fractional Brownian motion with $H \leq \frac{1}{2}$, then $\mathbb{E}L_\theta^{-p}(X) < \infty$ for every $\theta > H$ and every $p > 0$.

REMARK 1. Note that this immediately tells us that if X is Hölder rough, then it admits a kind of Doob–Meyer decomposition in the sense that the processes A and B in (1.2) are uniquely determined by Z. An interesting fact is that Hölder roughness is a *deterministic* property. In principle, one could imagine being able to check that this property holds almost surely for a number of driving noises, not even necessarily Gaussian ones.

Combined with standard arguments, this result yields quantitative bounds on the inverse of the Malliavin matrix, thus obtaining a quantitative version of the result obtained in [6], where the authors showed via a 0–1 law argument that the Malliavin matrix is almost surely invertible. If we use the additional assumption that the linearization of the RDE (1.1) has moments of all orders, our results also yield that (1.1) has *smooth densities* thus extending the work pioneered by Malliavin to the case in which the driving noise is a rough path. In fact the very recent work [8] obtains such moment bounds for the linearization of (1.1) under certain boundedness conditions, and thus our result immediately yields smoothness of densities for a large class of RDEs of the form (1.1). Even in the case $H = \frac{1}{2}$, we believe that the bounds derived in this work are new and pave the way for more quantitative versions of Hörmander's theorem.

When this work was nearing completion, we were notified of an independent work [25] showing smoothness of densities to solutions to (1.1) in the case when the driving vector fields exhibit a "nilpotent" structure, which allows one to obtain a priori bounds on the Malliavin derivatives of the solutions. The work [25] also contains a version of Norris's lemma in the context of SDEs driven by fractional Brownian motion with Hurst parameter $H \in (\frac{1}{3}, \frac{1}{2})$.

In the second half of the paper, we show that under an additional controllability assumption, the SDE (1.1) has a unique invariant measure, which follows from the strong Feller property defined in [22]. Note that, thanks to a cutoff argument, we do not need to require that the linearized equation has bounded moments to obtain this uniqueness result. If we denote by $A_{t,z}$ the closure of the set of all points that are accessible at time t for solutions to the control problem associated to (1.1) starting at z, the second major result of this paper is the following:

THEOREM 1.2. *Assume that the vector fields $\{V_i\}$ have derivatives with at most polynomial growth, and that* (1.1) *has global solutions. Then, if Hörmander's bracket condition holds at every point,* (1.1) *is strong Feller.*

In particular, if there exists $t > 0$ such that $\bigcap_{z\in\mathbb{R}^n} A_{t,z} \neq \emptyset$, then (1.1) *admits at most one invariant measure in the sense of* [18].

The above result gives us the uniqueness of the invariant measure for system (1.1). Theorem 1.2, combined with the assumption of the existence of an invariant measure, will yield that system (1.1) is ergodic.

The remainder of the article is structured as follows. In Section 2 we review the framework of controlled rough paths from [17] and set up the notation and derive some preliminary estimates. In Section 3, we then prove a general deterministic version of Norris's lemma for SDEs driven by rough paths. Furthermore we show that our assumptions are almost surely satisfied by the sample paths of fractional Brownian motion. Section 4 is a rather technical section in which we show that solutions to (1.1) are smooth in the sense of Malliavin calculus and obtain a priori bounds on their Malliavin derivatives. We then obtain quantitative bounds on the lowest eigenvalue of the Malliavin matrix in Section 5. Using the results in that section, we show that the existence of moments of the derivative of the flow implies the smoothness of the transition densities. In Section 6, we show the ergodicity of SDEs driven by fractional Brownian motion under Hörmander's condition and a standard controllability assumption. In Section 7, we finally give a few examples where our results are applicable.

2. Preliminaries.

2.1. *Notation.* Throughout this article, we will make use of the following notation. For quantities E and R, we write $E \leq \mathrm{K}(R)$ as a shorthand to mean that there exists a continuous increasing function $b : \mathbb{R}_+ \mapsto \mathbb{R}_+$ such that the bound $E \leq b(R)$ holds. Note that the function b in question is unspecified and may change from one line to the next. We also use the letter M to denote an arbitrary (possibly problem-dependent) constant whose precise value might vary from one line to the next.

2.2. *Introduction to the theory of rough paths.* In this work, we adopt the framework of [17] which offers a slightly different perspective on the pioneering work of Terry Lyons [33].

Denote by $\Omega\mathcal{C}$ the set of continuous functions from $\mathbb{R}^2$ to $\mathbb{R}$ which are 0 on the diagonal and define the "increment" operator $\delta : \mathcal{C} \mapsto \Omega\mathcal{C}$ by

$$\delta A_{st} \stackrel{\text{def}}{=} A_t - A_s. \tag{2.1}$$

For a fixed final time $T > 0$ and a continuous function $f : [0, T] \mapsto \mathbb{R}^n$, set

$$\|f\|_\infty = \sup_{t \in [0,T]} \bigl|f(t)\bigr|, \qquad \|f\|_\gamma = \sup_{s,t \in [0,T]} \frac{|\delta f_{st}|}{|t-s|^\gamma}. \tag{2.2}$$

We also define the norm

$$\|f\|_{\mathcal{C}^\gamma} = \|f\|_\infty + \|f\|_\gamma.$$

With these notation, a *rough path* on the interval $[0,T]$ consists of two parts, a continuous function $X:[0,T]\mapsto\mathbb{R}^d$ and a continuous "area process" $\mathbb{X}:[0,T]^2\mapsto\mathbb{R}^{d\times d}$, $\mathbb{X}\in\Omega\mathcal{C}$ satisfying the algebraic identity

$$\mathbb{X}^{ij}_{st}-\mathbb{X}^{ij}_{ut}-\mathbb{X}^{ij}_{su}=\delta X^i_{su}\delta X^j_{ut} \tag{2.3}$$

for all $\{s,u,t\}\in[0,T]$ and $1\le i,j\le d$. For $\mathbb{X}\in\Omega\mathcal{C}$, define

$$\|\mathbb{X}\|_{2\gamma}\stackrel{\text{def}}{=}\sup_{\substack{s,t\in[0,T]\\ s\neq t}}\frac{|\mathbb{X}_{st}|}{|t-s|^{2\gamma}}. \tag{2.4}$$

For $\gamma\in(\frac{1}{3},\frac{1}{2}]$, we denote by $\mathcal{D}^\gamma([0,T],\mathbb{R}^d)$ the space of rough paths, consisting of those pairs $(X,\mathbb{X})$ satisfying (2.3) and such that

$$\|(X,\mathbb{X})\|_\gamma\stackrel{\text{def}}{=}\|X\|_\gamma+\|\mathbb{X}\|_{2\gamma}<\infty.$$

Notice that $\|(X,\mathbb{X})\|_\gamma$ is only a semi-norm and that $\mathcal{D}^\gamma$ actually is not a vector space, due to the nonlinear constraint (2.3).

For every smooth function $X:[0,T]\to\mathbb{R}^d$, there exists a canonical representative in $\mathcal{D}^\gamma$ by choosing

$$\mathbb{X}_{s,t}=\int_s^t\delta X_{sr}\otimes dX_r.$$

We then denote by $\mathcal{D}^\gamma_g$ the closure of the set of smooth functions in $\mathcal{D}^\gamma$. (Here, g stands for "geometric.") The space $\mathcal{D}^\gamma_g$ has the nice feature of being a Polish space [16], Proposition 8.27, which will be useful in the sequel.

2.3. *Controlled rough paths*. For defining integrals with respect to rough paths, a useful notion introduced first in [17] is that of "controlled" paths:

DEFINITION 1. Let $(X,\mathbb{X})\in\mathcal{D}^\gamma([0,T],\mathbb{R}^d)$ for some $\gamma\in(\frac{1}{3},\frac{1}{2}]$. A pair (Z,Z') is said to be controlled by X if $Z\in\mathcal{C}^\gamma([0,T],\mathbb{R}^n)$, $Z'\in\mathcal{C}^\gamma([0,T],\mathbb{R}^{n\times d})$, and the "remainder" term $R^Z\in\Omega\mathcal{C}$ implicitly defined by

$$\delta Z_{st}=Z'_s\delta X_{st}+R^Z_{st}, \tag{2.5}$$

satisfies $\|R^Z\|_{2\gamma}<\infty$.

Denoting by $\mathcal{C}^\gamma_X$ the set of paths controlled by X, we endow it with the norm

$$\|(Z,Z')\|_{X,\gamma}=|Z(0)|+\|Z'\|_{\mathcal{C}^\gamma}+\|R^Z\|_{2\gamma}.$$

As noticed in [17], now we may *define* the integral of a weakly controlled path $(Z,Z')\in\mathcal{C}^\gamma_X$ with respect to a rough path $(X,\mathbb{X})$ by taking a limit of modified Riemann sums:

$$\int_0^T Z_t\otimes dX_t=\lim_{|\mathrm{P}|\to 0}\sum_{[s,t]\in\mathrm{P}}(Z_s\otimes\delta X_{st}+Z'_s\mathbb{X}_{st}), \tag{2.6}$$

where P is a finite partition of the interval $[0, T]$ into subintervals and |P| denotes the length of the largest subinterval. The following result, adapted from [17], Proposition 1, gives the continuity of the integral with respect to its integrand:

THEOREM 2.1. *Let* $(X, \mathbb{X}) \in \mathcal{D}^\gamma([0, T], \mathbb{R}^d)$ *for some* $\gamma > \frac{1}{3}$ *and* $(Y, Y') \in \mathcal{C}_X^\gamma$ *be a weakly controlled rough path. Then the map*

$$(Y, Y') \mapsto (Z, Z') \stackrel{def}{=} \left(\int_0^\cdot Y_t \otimes dX_t, Y \otimes I\right),$$

where the integral is as defined in (2.6), *is continuous from* $\mathcal{C}_X^\gamma$ *to* $\mathcal{C}_X^\gamma$, *and furthermore we have the bound*

$$\|R^Z\|_{2\gamma} \le M(\|X\|_\gamma \|R^Y\|_{2\gamma} + \|\mathbb{X}\|_{2\gamma}\|Y'\|_{\mathcal{C}^\gamma}) \tag{2.7}$$

for a constant M *independent of* X, Y.

REMARK 2. Notice that from (2.7) we deduce that

$$\|(Z, Z')\|_{X,\gamma} \le M(\|X\|_\gamma(\|R^Y\|_{2\gamma} + \|Y\|_{\mathcal{C}^\gamma}) + \|\mathbb{X}\|_{2\gamma}\|Y'\|_{\mathcal{C}^\gamma}). \tag{2.8}$$

For $(Y, Y') \in \mathcal{C}_X^\gamma$ and a C^2 function $\psi : \mathbb{R}^n \mapsto \mathbb{R}^m$, we may define a new weakly controlled rough path $(\psi(Y), \psi(Y)') \in \mathcal{C}_X^\gamma$ as

$$\psi(Y)_t = \psi(Y_t), \qquad \psi(Y)'_t = D\psi(Y_t)Y'_t. \tag{2.9}$$

Then we have the following bound from [20], Lemma 2.2:

LEMMA 1. *Let* $(Y, Y') \in \mathcal{C}_X^\gamma$ *and* $(\psi(Y), \psi(Y)')$ *be as defined in* (2.9). *Then we have*

$$\begin{aligned}&\|(\psi(Y), \psi(Y)')\|_{X,\gamma}\\ &\quad\le M(1 + \|\psi\|_\infty + \|D^2\psi\|_\infty)(1 + \|(X, \mathbb{X})\|_\gamma)(1 + \|(Y, Y')\|_{X,\gamma})^2,\end{aligned}$$

where the supremum norms of ψ *and* $D^2\psi$ *are taken over the ball of radius* $\|Y\|_\infty$, *and the constant* M *is independent of* X, Y, ψ.

2.4. *Notion of solution.* With all of these notation at hand, we give the following definition of a solution to (1.1):

DEFINITION 2. Let $\gamma > \frac{1}{3}$, and let $(X, \mathbb{X}) \in \mathcal{D}^\gamma$. Then $Z \in \mathcal{C}^\gamma$ is a solution to (1.1) if $(Z, Z') = (Z, V(Z)) \in \mathcal{C}_X^\gamma$, and the integral version of (1.1) holds, where the composition of a controlled rough path with a nonlinear function is interpreted as in (2.9) and the integral of a controlled rough path against X is defined by (2.6). Here, we denoted by V the collection $(V_1, \dots, V_d)$.

A standard fixed point argument, as given, for example, in [17, 33], then yields:

THEOREM 2.2. *For $V \in \mathcal{C}^3$, there exists a unique local solution to* (1.1).

From now on, we will refer to this notion of a solution to (1.1).

REMARK 3. In Theorem 2.2, if the vector fields V are bounded with bounded derivatives, then there exists a global solution [33].

3. A version of Norris's lemma. One of the main ingredients of the proof of Hörmander's theorem using Malliavin calculus is Norris's lemma, which is essentially a quantitative version of the Doob–Meyer decomposition theorem. Loosely speaking, it states that under certain additional regularity assumptions, if a semimartingale is "small," then both its bounded variation part and its martingale part have to be "small" separately. In other terms, if we have some a priori knowledge of the regularity of a semimartingale, then there is a limit to the amount of cancellations that can occur between the two terms in its Doob–Meyer decomposition. The intuitive reason for this is that a continuous martingale is nothing but a time-changed Brownian motion, and so it has to be very rough at every single scale.

Results of this type are usually considered to be the archetype of a probabilistic result. The aim of this section is to argue that while the probabilistic intuition described above is certainly correct, one can have a much more pathwise perspective on Norris's lemma. This was already apparent in [21], where the authors obtain a result that is similar in flavor to Norris's lemma, but where this lack of cancellations is formulated as a *deterministic* property that occurs on a *universal* "large" subset of Wiener space. Here, we take this viewpoint one step further by exhibiting a universal set on which a quantitative version of Norris's lemma holds as a deterministic property.

The main ingredient in our pathwise perspective is the following definition that makes precise what we mean by a path that is "rough at every scale":

DEFINITION 3. A path X_t with values in $\mathbb{R}^n$ is said to be *θ-Hölder rough* in the interval $[0, T]$ for $\theta \in (0, 1)$, if there exists a constant $L_\theta(X)$ such that for every $s \in [0, T]$, every $\varepsilon \in (0, T/2]$ and every $\varphi \in \mathbb{R}^n$ with $\|\varphi\| = 1$, there exists $t \in [0, T]$ such that

$$|t - s| \le \varepsilon \quad \text{and} \quad \bigl|\langle \varphi, \delta X_{s,t}\rangle\bigr| > L_\theta(X)\varepsilon^\theta. \tag{3.1}$$

We denote the largest such L_θ the "modulus of θ-Hölder roughness of X."

REMARK 4. We emphasize that the choice of quantifiers in the above definition ensures that such Hölder rough paths actually do exist. In particular, as soon as $n \ge 2$, it is essential to allow the precise location of t such that (3.1) holds to depend on the vector φ.

A first, rather straightforward consequence of this definition is that if a rough path $(X, \mathbb{X})$ happens to be Hölder rough, then the "derivative process" Z' in the decomposition (2.5) of a controlled rough path is uniquely determined by Z. This can be made quantitative in the following way:

PROPOSITION 1. *Let $(X, \mathbb{X}) \in \mathcal{D}^\gamma([0,T], \mathbb{R}^n)$ be such that X is a θ-Hölder rough path. Then there exists a constant M depending only on n and m such that the bound*

$$\|Z'\|_\infty \le \frac{M\|Z\|_\infty}{L_\theta(X)} \big(\|R^Z\|_{2\gamma}^{\theta/(2\gamma)} \|Z\|_\infty^{-\theta/(2\gamma)} \vee T^{-\theta}\big),$$

holds for every controlled rough path $(Z, Z') \in \mathcal{C}_X^\gamma([0,T], \mathbb{R}^m)$.

PROOF. Fix $s \in [0,T]$ and $\varepsilon \in (0, T/2]$, From the definition of the remainder R^Z in (2.5), it then follows that

$$\sup_{|t-s|\le\varepsilon} |Z'_s \delta X_{s,t}| \le \sup_{|t-s|\le\varepsilon} \big(|\delta Z_{s,t}| + |R^Z_{s,t}|\big) \le 2\|Z\|_\infty + \|R^Z\|_{2\gamma} \varepsilon^{2\gamma}. \tag{3.2}$$

Let now $Z'_s(j)$ denote the jth row of the matrix Z'_s. Since X is θ-Hölder rough by assumption, for every $j \le d$, there exists $v = v(j)$ with $|v - s| \le \varepsilon$ such that

$$\big|\langle Z'_s(j), \delta X_{s,v}\rangle\big| > L_\theta(X)\varepsilon^\theta |Z'_s(j)|. \tag{3.3}$$

Combining both (3.2) and (3.3), we thus obtain that

$$L_\theta(X)\varepsilon^\theta |Z'_s(j)| \le 2\|Z\|_\infty + \|R^Z\|_{2\gamma}\varepsilon^{2\gamma}.$$

Summing over the rows of Z'_s yields a universal constant C such that

$$L_\theta(X)\varepsilon^\theta |Z'_s| \le C\big(\|Z\|_\infty + \|R^Z\|_{2\gamma}\varepsilon^{2\gamma}\big).$$

Optimizing over ε, we choose $\varepsilon = \|Z\|_\infty^{1/2\gamma} \|R^Z\|_{2\gamma}^{-1/2\gamma} \wedge (T/2)$, thus deducing that

$$|Z'_s| \le \frac{M\|Z\|_\infty}{L_\theta(X)} \big(\|R^Z\|_{2\gamma}^{\theta/(2\gamma)} \|Z\|_\infty^{-\theta/(2\gamma)} \vee T^{-\theta}\big).$$

Since s was arbitrary, the stated bound follows at once. □

One way of reading Proposition 1 is to say that if $\|Z\|_\infty$ is small, then $\|Z'\|_\infty$ must also be small, provided that $(Z, Z') \in \mathcal{C}_X^\gamma([0,T], \mathbb{R}^m)$ and that X is Hölder rough. In the following theorem, we apply Proposition 1 to obtain a quantitative version of a "Doob–Meyer type decomposition" for SDEs driven by a rough path X. This is the main new technical result of this article.

THEOREM 3.1. *Let* $(X, \mathbb{X}) \in \mathcal{D}^\gamma([0,T],\mathbb{R}^n)$ *with* $\gamma > \frac{1}{3}$ *be such that* X *is* θ*-Hölder rough with* $2\gamma > \theta$. *Let* $(A, A') \in \mathcal{C}_X^\gamma([0,T],\mathbb{R}^{mn})$ *and* $B \in \mathcal{C}^\gamma([0,T],\mathbb{R}^m)$, *and set*

$$Z_t = \int_0^t A_s\, dX_s + \int_0^t B_s\, ds. \tag{3.4}$$

Then, there exist constants $r > 0$ *and* $q > 0$ *such that, setting*

$$\mathcal{R} \stackrel{def}{=} 1 + L_\theta(X)^{-1} + \|(X, \mathbb{X})\|_\gamma + \|(A, A')\|_{X,\gamma} + \|B\|_{\mathcal{C}^\gamma},$$

one has the bound

$$\|A\|_\infty + \|B\|_\infty \le M\mathcal{R}^q \|Z\|_\infty^r$$

for a constant M *depending only on* T, m *and* n.

REMARK 5. The proof provides the explicit value $q = 6$ and shows that r can be taken arbitrarily close to $(2\gamma-\theta)^2(3\gamma-1)/(4\gamma^2(1+\gamma))$, but these values are certainly not optimal.

PROOF. Note first that the definition of Z does not change if we add a constant to X. We will therefore assume without loss of generality that $X_0 = 0$, so that $\|X\|_\infty \le T^\gamma \|X\|_\gamma \le T^\gamma \mathcal{R}$. By Theorem 2.1 we deduce that the pair (Z, A) is a weakly controlled rough path, $(Z, A) \in \mathcal{C}_X^\gamma([0,T],\mathbb{R}^m)$, with

$$\delta Z = A\delta X + R^Z \tag{3.5}$$

and

$$\|R^Z\|_{2\gamma} \le M\big(\|X\|_\gamma \|R^A\|_{2\gamma} + \|\mathbb{X}\|_{2\gamma}\|A\|_{\mathcal{C}^\gamma} + \|B\|_{\mathcal{C}^\gamma}\big) \le M\mathcal{R}^2.$$

We deduce from the above that in particular, we have the a priori bound $\|Z\|_\infty \le M\mathcal{R}^2$.

It then follows from Proposition 1 that

$$\begin{aligned}\|A\|_\infty &\le M L_\theta(X)^{-1}\|Z\|_\infty^{1-\theta/(2\gamma)}\big(\|R^Z\|_{2\gamma}^{\theta/(2\gamma)} + \|Z\|_\infty^{\theta/(2\gamma)}\big)\\ &\le M\mathcal{R}^3\|Z\|_\infty^{1-\theta/(2\gamma)}.\end{aligned} \tag{3.6}$$

This is already the requested bound on A. The bound on B is slightly more difficult to obtain.

At this stage, we would like to make use of the information that $\|A\|_\infty$ is "small" to get a bound on the integral of A against X. In order to do so, it turns out to be convenient to choose a $\beta \in (\frac{1}{3}, \gamma)$ with $2\beta > \theta$, so that we can interpret $(X, \mathbb{X})$ as an element of $\mathcal{D}^\beta([0,T],\mathbb{R}^n)$ with $\|(X,\mathbb{X})\|_\beta \le M\|(X,\mathbb{X})\|_\gamma$. This will allow us to make use of interpolation inequalities to combine our a priori knowledge about the boundedness of (A, A') in $\mathcal{C}_X^\gamma$ norm with (3.6) to conclude that (A, A') is small in $\mathcal{C}_X^\beta$.

We first obtain a bound on A'. Since $(A, A') \in \mathcal{C}_X^\gamma([0,T], \mathbb{R}^{mn})$, we infer from (3.6) and Proposition 1 that

$$\|A'\|_\infty \le M L_\theta(X)^{-1} \|R^A\|_{2\beta}^{\theta/(2\gamma)} \|A\|_\infty^{1-\theta/(2\gamma)} \le M\mathcal{R}^4 \|Z\|_\infty^{(1-\theta/(2\gamma))^2}.$$

Using the inequality

$$\|A'\|_\beta \le 2\|A'\|_\gamma^{\beta/\gamma} \|A'\|_\infty^{1-\beta/\gamma}, \tag{3.7}$$

which follows immediately from the definition of the Hölder norm, we obtain the bound

$$\|A'\|_\beta \le M\|A'\|_\infty^{1-\beta/\gamma} \|A'\|_\gamma^{\beta/\gamma} \le M\mathcal{R}^4 \|Z\|_\infty^{(1-\theta/(2\gamma))^2(1-\beta/\gamma)},$$

where we used the fact that $\beta < \gamma$. Similarly, we would like to obtain a bound on $\|R^A\|_{2\beta}$. Combining the definition of R^A with (3.6), we deduce that

$$\|R^A\|_\infty \le 2\big(\|A\|_\infty + \|A'\|_\infty \|X\|_\infty\big) \le M\mathcal{R}^5 \|Z\|_\infty^{(1-\theta/(2\gamma))^2}.$$

Using the obvious equivalent to (3.7), we conclude that

$$\|R^A\|_{2\beta} \le M\|R^A\|_{2\gamma}^{\beta/\gamma} \|R^A\|_\infty^{1-\beta/\gamma} \le M\mathcal{R}^5 \|Z\|_\infty^{(1-\theta/(2\gamma))^2(1-\beta/\gamma)}.$$

We are now in a position to use Theorem 2.1 to bound the integral $\int_0^\cdot A_s\,dX_s$. Indeed, we obtain from (2.7) the bound

$$\left\|\int_0^\cdot A_s\,dX_s\right\|_\infty \le M\big(|A_0|\|X\|_\infty + \|X\|_\beta \|R^A\|_{2\beta} + \|\mathbb{X}\|_{2\beta} \|A'\|_\beta\big).$$

Inserting into this bound all of the above estimates, we conclude that

$$\left\|\int_0^\cdot A_s\,dX_s\right\|_\infty \le M\mathcal{R}^6 \|Z\|_\infty^{(1-\theta/(2\gamma))^2(1-\beta/\gamma)}. \tag{3.8}$$

This estimate, together with the definition (3.4) of Z immediately implies that we also have the bound

$$\left\|\int_0^\cdot B_s\,ds\right\|_\infty \le M\mathcal{R}^6 \|Z\|_\infty^{(1-\theta/(2\gamma))^2(1-\beta/\gamma)}.$$

Once again we use an interpolation inequality to strengthen this bound. Applying the interpolation inequality

$$\|\partial_t f\|_\infty \le M\|f\|_\infty \max\left(\frac{1}{T}, \|f\|_\infty^{-1/(\gamma+1)} \|\partial_t f\|_\gamma^{1/\gamma+1}\right)$$

(see [21], Lemma 6.14) with $f(t) = \int_0^t B_s\,ds$, it follows that

$$\|B\|_\infty \le M\mathcal{R}^6 \|Z\|_\infty^{(1-\theta/2\gamma)^2(1-\beta/\gamma)\gamma/(1+\gamma)}. \tag{3.9}$$

The claim now follows from (3.6) and (3.9), and the remark following the statement follows by choosing $\beta \approx \frac{1}{3}$. $\square$

3.1. *Hölder roughness of sample paths of fBm.* Our aim now is to show that the sample paths of a fractional Brownian motion X (with $H \le 1/2$ throughout this section) are indeed almost surely Hölder rough and to provide quantitative bounds on the tail behavior of $L_\theta(X)$ for a suitable θ. Let $\{\mathcal{F}_s^X, s \in \mathbb{R}\}$ be the natural filtration generated by the fBm X, namely $\mathcal{F}_s^X$ is the σ-algebra generated by $\{X_r\}_{r\in(-\infty,s]}$. We start with the following lemma on the small ball probability of the conditioned fBm:

LEMMA 2. *Let $\varphi \in \mathbb{R}^n$ with $\|\varphi\| = 1$ and $\delta \le 1$. Then there exist constants M and c such that the bound*

$$\mathbb{P}\Big(\inf_{\|\varphi\|=1} \sup_{s,t\in[0,\delta]} \big|\langle\varphi, \delta X_{st}\rangle\big| \le \varepsilon \big|\mathcal{F}_0^X\Big) \le M e^{-c\delta^{2H}\varepsilon^{-2}} \tag{3.10}$$

holds almost surely, for every $0 < \varepsilon \le 1$ and $H \le 1/2$.

PROOF. By the scaling properties of (conditioned) fractional Brownian motion, and since the bound is trivial for $\varepsilon > \delta^H$, we can restrict ourselves to the case $\delta = 1$ and $\varepsilon \le 1$. For the moment, let us fix an arbitrary φ with $\|\varphi\| = 1$.

Since $X_0 = 0$, we obtain

$$\mathbb{P}\Big(\sup_{s,t\in[0,1]} \big|\langle\varphi, \delta X_{st}\rangle\big| \le \varepsilon \big|\mathcal{F}_0^X\Big) \le \mathbb{P}\Big(\sup_{t\in[0,1]} \big|\langle\varphi, X_t\rangle\big| \le \varepsilon \big|\mathcal{F}_0^X\Big).$$

At this stage, we note that there exists a one-dimensional Wiener process W (depending on φ) independent of $\mathcal{F}_0^X$, a stochastic process $Y^\varphi = \langle\varphi, Y\rangle$ such that Y_t is $\mathcal{F}_0^X$-measurable for every $t \ge 0$, and a constant c such that

$$\langle\varphi, X_t\rangle = Y_t^\varphi + c\int_0^t (t-s)^{H-1/2}\, dW(s) \stackrel{\text{def}}{=} Y_t^\varphi + \hat{X}_t. \tag{3.11}$$

(See, e.g., [18, 37], as well as (4.2) below.) Furthermore, Y has almost surely bounded sample paths. Any such sample path then induces a seminorm $\|\cdot\|_Y$ on $\mathbb{R}^n$ by

$$\|\varphi\|_Y \stackrel{\text{def}}{=} \sup_{t\in[0,1]} |Y_t^\varphi|.$$

Furthermore, this is almost surely nondegenerate, so that $\|\cdot\|_Y$ is actually a norm. For the rest of the proof, we use c for a generic universal constant that will change from expression to expression.

It then follows from [27], Theorem 2 (set $\alpha = 1$), which is a refinement of Anderson's inequality [1], that we have the bound

$$\begin{aligned}\mathbb{P}\Big(\sup_{t\in[0,1]} \big|\langle\varphi, X_t\rangle\big| \le \varepsilon \big|\mathcal{F}_0^X\Big) &= \mathbb{P}\Big(\sup_{t\in[0,1]} \big|\hat{X}_t + Y_t^\varphi\big| \le \varepsilon \big|\mathcal{F}_0^X\Big) \\ &\le \exp\Big(-\inf_{\|Y^\varphi - h\|_\infty \le \varepsilon} \frac{\|h\|_{\mathcal{H}}^2}{2}\Big)\mathbb{P}\Big(\sup_{t\in[0,1]} |\hat{X}_t| \le \varepsilon\Big),\end{aligned}$$

where $\|h\|_{\mathcal{H}}$ denotes the norm of h in the Cameron–Martin space of $\hat{X}$. Since this norm is always stronger than the supremum norm, there exists a constant c such that $\|h\|_{\mathcal{H}} \geq c\|h\|_\infty \geq c(\|\varphi\|_Y - \varepsilon)$, so that

$$\mathbb{P}\Big(\sup_{t\in[0,1]} \big|\langle\varphi, X_t\rangle\big| \leq \varepsilon \big|\mathcal{F}_0^X\Big) \leq C\exp(-c\|\varphi\|_Y^2)\mathbb{P}\Big(\sup_{t\in[0,1]} |\hat{X}_t| \leq \varepsilon\Big) \tag{3.12}$$

for some positive constants c and C independent of $\varepsilon \leq 1$.

On the other hand, we can invert the expression (3.11) for $\hat{X}$, yielding

$$W_t = c\int_0^t (t-s)^{1/2-H}\,d\hat{X}_s = c\int_0^t (t-s)^{-1/2-H}\hat{X}_s\,ds.$$

In particular, provided that $H < \frac{1}{2}$, we have the bound

$$\sup_{t\in[0,1]} |W_t| \leq c \sup_{t\in[0,1]} |\hat{X}_t|,$$

so that

$$\mathbb{P}\Big(\sup_{t\in[0,1]} |\hat{X}_t| \leq \varepsilon\Big) \leq \mathbb{P}\Big(\sup_{t\in[0,\delta]} |W_t| \leq c\delta^{1/2-H}\varepsilon\Big) \leq Me^{-c\varepsilon^{-2}},$$

where the last inequality is the well-known small ball probability for the standard Brownian motion [31]. Combining this with (3.12), we conclude that

$$\mathbb{P}\Big(\sup_{t\in[0,\delta]} \big|\langle\varphi, X_t\rangle\big| \leq \varepsilon\big|\mathcal{F}_0^X\Big) \leq M\exp\Big(-\frac{c}{\varepsilon^2} - c\|\varphi\|_Y^2\Big). \tag{3.13}$$

Up to now, the calculation was performed with a fixed instance of φ. In order to conclude, we use a covering argument similar to [38], page 127. The main problem difference is that such an argument requires a priori bounds on the process X, and these are of course not uniform in Y. It turns out that, thanks to the exponentially decaying factor in (3.13), it is still possible to obtain a uniform bound, but one needs to be a little bit more careful. Our main tool is the fact that, as a consequence of John's theorem [2, 26], it is possible to perform an orthogonal change of coordinates for φ (that depends on Y) and to find constants $Y^{(1)}, \ldots, Y^{(n)}$ such that

$$\sup_j Y^{(j)}|\varphi_j| \leq \|\varphi\|_Y \leq C\sup_j Y^{(j)}|\varphi_j|, \tag{3.14}$$

were the constant C is universal and depends only on n. By equivalence of norms in $\mathbb{R}^n$ we can furthermore, for any given realization of Y, replace the Euclidean norm $\|\varphi\|$ in (3.10) by the ℓ^∞ norm $\|\varphi\|_\infty$. If we can find a finite collection $\Phi \subset \mathbb{R}^n$ such that, for every φ with $\|\varphi\|_\infty = 1$, one has the bound

$$\sup_{\|\varphi\|_\infty=1}\inf_{\tilde{\varphi}\in\Phi} \|\varphi - \tilde{\varphi}\|_Y \leq \frac{\varepsilon}{4}, \tag{3.15}$$

then we obtain the inequality

$$\begin{aligned}&\mathbb{P}\Big(\inf_{\|\varphi\|_\infty=1}\sup_{s,t\in[0,1]}|\langle\varphi,\delta X_{st}\rangle|\le\varepsilon\big|Y\Big)\\ &\qquad\le M\sum_{\tilde\varphi\in\Phi}\exp\Big(-\frac{c}{\varepsilon^2}-c\|\tilde\varphi\|_Y^2\Big)\\ &\qquad\quad+\mathbb{P}\Big(\sup_{\|\varphi\|_\infty=1}\inf_{\tilde\varphi\in\Phi}\sup_{s,t\in[0,\delta]}|\langle\varphi-\Phi,\delta\hat X_{st}\rangle|\le\frac{\varepsilon}{4}\Big).\end{aligned}\tag{3.16}$$

If we can furthermore choose the collection Φ so that

$$\sup_{\|\varphi\|_\infty=1}\inf_{\tilde\varphi\in\Phi}\|\varphi-\tilde\varphi\|_\infty\le\varepsilon^2,\tag{3.17}$$

then, due to the Gaussian tails of $\hat X$, the second term in (3.16) is bounded by $M\exp(-c/\varepsilon^2)$ as desired. It thus remains to show that, for every norm $\|\cdot\|_Y$, it is possible to choose Φ satisfying (3.15) and (3.17) and such that

$$\sum_{\varphi\in\Phi}\exp(-c\|\varphi\|_Y^2)\le\frac{C}{\varepsilon^\kappa}\tag{3.18}$$

for some constants $C>0$ and $\kappa>0$, uniformly over $\|\cdot\|_Y$. We choose $\Phi\subset\{\varphi:\|\varphi\|_\infty=1\}$ in the following way. Let

$$A_j=\varepsilon^2\wedge\frac{\varepsilon}{4Y^{(j)}},$$

and, for every $k\in\mathbf{Z}^d$, denote by Ak the element in $\mathbb{R}^n$ given by $\sum_j A_jk_je_j$, where e_j is the jth unit vector. We also write $\mathbf{Z}^d_j$ for the elements in $k\in\mathbf{Z}^d$ with $k_j=0$. We then set $\Phi=\bigcup_{j=1}^n(\Phi_j^+\cup\Phi_j^-)$, where

$$\Phi_j^\pm=\{\pm e_j+Ak:k\in\mathbf{Z}^d_j\}\cap\{\varphi:\|\varphi\|_\infty=1\}.$$

It is clear that this choice of Φ satisfies both (3.15) and (3.17), so that it remains to show that (3.18) is satisfied, uniformly over the choices of $\{Y^{(j)}\}$. For $k\in\mathbf{Z}^d_j$, denote $\varphi^\pm_{j;k}=\pm e_j+Ak$. With this notation at hand, it follows from (3.14) that there exists a constant c such that

$$\|\varphi_{j;k}\|_Y^2\ge c\varepsilon^2\sum_{i\neq j}|k_i|^2(1\wedge\varepsilon^2|Y^{(i)}|^2).$$

It follows that

$$\begin{aligned}&\sum_{\varphi\in\Phi_j^\pm}\exp(-c\|\varphi\|_Y^2)\\ &\qquad\le\prod_{i\neq j}\sum_{|k|\le A_i^{-1}}\exp(-c\varepsilon^2|k|^2(1\wedge\varepsilon^2|Y^{(i)}|^2))\end{aligned}$$

$$\begin{aligned}
&\le \prod_{i\neq j}\Bigl(2\varepsilon^{-2} + \mathbf{1}_{|4Y^{(i)}|>\varepsilon^{-1}} \sum_{|k|\le 4Y^{(j)}/\varepsilon} \exp\bigl(-c|k|^2\varepsilon^4|Y^{(i)}|^2\bigr)\Bigr)\\
&\le c\prod_{i\neq j}\Bigl(\varepsilon^{-2} + \mathbf{1}_{|4Y^{(i)}|>\varepsilon^{-1}} \int_{\mathbb{R}} e^{-c|x|^2\varepsilon^4|Y^{(i)}|^2}\,dx\Bigr)\\
&\le c\prod_{i\neq j}\Bigl(\varepsilon^{-2} + \mathbf{1}_{|4Y^{(i)}|>\varepsilon^{-1}} \frac{1}{\varepsilon^2|Y^{(i)}|}\Bigr) \le c\varepsilon^{-2n},
\end{aligned}$$

where all the constants c are independent of the choice of coefficients $Y^{(i)}$, so that the bound (3.18) does indeed hold, which concludes the proof.

Note that, for $H = 1/2$, the exact same argument goes through, but it is simplified due to the Markov property, which implies that $Y = 0$. □

We have the following corollary of Lemma 2:

COROLLARY 1. *For any interval* $I_\delta \overset{def}{=} [u_\ell, u_\ell + \delta] \subset \mathbb{R}$ *of length* δ *and any* $u \le u_\ell$, *there exist constants* M *and* c *such that the bound*

$$\mathbb{P}\Bigl(\inf_{\|\varphi\|=1} \sup_{s,t\in I_\delta} \bigl|\langle\varphi, \delta X_{st}\rangle\bigr| \le \varepsilon \big| \mathcal{F}_u^X\Bigr) \le Me^{-c\delta^{2H}\varepsilon^{-2}}$$

holds for every $0 < \varepsilon \le 1$ *and* $H \le 1/2$.

PROOF. Define the event $G \overset{\text{def}}{=} \{\inf_{\|\varphi\|=1} \sup_{s,t\in I_\delta} |\langle\varphi, \delta X_{st}\rangle| \le \varepsilon\}$. Since the increments of the fBm are stationary, by Lemma 2 we obtain the bound

$$\mathbb{E}\bigl(1_G|\mathcal{F}_{u_\ell}^X\bigr) \le Me^{-c\delta^{2H}\varepsilon^{-2}}. \tag{3.19}$$

Now notice that for any $G \in \mathcal{F}_{u_\ell}^X$ and $u \le u_\ell$, $\mathbb{E}(G|\mathcal{F}_u^X) = \mathbb{E}(\mathbb{E}(G|\mathcal{F}_{u_\ell}^X)|\mathcal{F}_u^X)$. Since the right-hand side of equation (3.19) does not depend on u_ℓ, it immediately follows that

$$\mathbb{P}\Bigl(\inf_{\|\varphi\|=1} \sup_{s,t\in I_\delta} \bigl|\langle\varphi, \delta X_{st}\rangle\bigr| \le \varepsilon \big| \mathcal{F}_u^X\Bigr) = \mathbb{E}\bigl(1_G|\mathcal{F}_u^X\bigr) \le Me^{-c\delta^{2H}\varepsilon^{-2}} \tag{3.20}$$

and the corollary follows. □

REMARK 6. Although the above proof requires that $H \le \frac{1}{2}$, one would expect that a result similar to that of Lemma 2 holds for any Gaussian process X such that

$$M_\ell^2|t-s|^{2H} \le \mathbb{E}|X_t - X_s|^2 \le M_u^2|t-s|^{2H}, \qquad s,t\in[0,T]$$

for some constants M_ℓ, M_u, even if $H > \frac{1}{2}$. For instance, a result similar to Lemma 2 can be shown to hold in the case of fBm with index $H > \frac{1}{2}$; see, for example, [3], Proposition 3.4.

Using this estimate we are now in the position to obtain bounds on the modulus of Hölder roughness for fractional Brownian motion with $H \le \frac{1}{2}$ by a type of chaining argument.

LEMMA 3. *Let X be a fBm with Hurst parameter $H \le \frac{1}{2}$. Then, for every $\theta > H$, the sample paths of X are almost surely θ-Hölder rough. Moreover, there exist constants M and c independent of X such that*

$$\mathbb{P}\big(L_\theta(X) \le \varepsilon | \mathcal{F}_0^X\big) \le M \exp\big(-c\varepsilon^{-2}\big)$$

for all $\varepsilon \in (0,1)$. In particular, $\mathbb{E}(L_\theta^{-p}(X)|\mathcal{F}_0^X) < \infty$ for every $p > 0$.

PROOF. A different way of formulating Definition 3 is given by

$$L_\theta(X) = \inf_{\|\varphi\|=1} \inf_{t\in[0,T]} \inf_{r\in[0,T/2]} \sup_{|t-s|\le r} \frac{|\langle \varphi, \delta X_{st}\rangle|}{r^\theta}.$$

We then define the "discrete analog" $D_\theta(X)$ of $L_\theta(X)$ to be

$$D_\theta(X) \stackrel{\text{def}}{=} \inf_{\|\varphi\|=1} \inf_{n\ge 1} \inf_{k\le 2^n} \sup_{s,t\in I_{k,n}} \frac{|\langle \varphi, \delta X_{st}\rangle|}{(2^{-n}T)^\theta},$$

where $I_{k,n} = [\frac{k-1}{2^n}T, \frac{k}{2^n}T]$. We first claim that

$$L_\theta(X) \ge \frac{1}{2\cdot 8^\theta} D_\theta(X). \tag{3.21}$$

Indeed, given $t \in [0,T]$ and $r \in [0, T/2]$, pick $n \in \mathbb{N}$ such that $r/8 \le 2^{-n}T < r/4$. It follows that there exists some k such that $I_{k,n}$ is included in the set $\{s : r/2 \le |t-s| \le r\}$. Then, by definition of D_θ, for any unit vector φ there exist two points $t_1, t_2 \in I_n$ such that

$$\big|\langle \varphi, \delta X_{t_2 t_1}\rangle\big| \ge 2^{-n\theta} D_\theta(X).$$

Therefore by the triangle inequality, we conclude that the magnitude of the difference between $\langle \varphi, X_t\rangle$ and one of the two terms $\langle \varphi, X_{t_i}\rangle$, $i = 1, 2$ (say t_1) is at least

$$\big|\langle \varphi, \delta X_{t_1 t}\rangle\big| \ge \tfrac{1}{2} \cdot \big(2^{-n}T\big)^\theta D_\theta(X)$$

and therefore

$$\frac{|\langle \varphi, \delta X_{t_1 t}\rangle|}{r^\theta} \ge \frac{1}{2} \cdot \frac{2^{-n\theta}}{r^\theta} D_\theta(X) \ge \frac{1}{2\cdot 8^\theta} D_\theta(X).$$

Since t, r and φ were chosen arbitrarily, claim (3.21) follows.

It follows that it is sufficient to obtain the requested bound on $\mathbb{P}(D_\theta(X) \le \varepsilon | \mathcal{F}_0^X)$. We have the straightforward bound

$$\mathbb{P}\big(D_\theta(X) \le \varepsilon | \mathcal{F}_0^X\big) \le \mathbb{P}\Big(\inf_{\|\varphi\|=1} \inf_{n\ge 1} \inf_{k \le 2^n} \sup_{s,t \in I_{k,n}} \frac{|\langle \varphi, \delta X_{st}\rangle|}{2^{-n\theta}} \le \varepsilon \Big| \mathcal{F}_0^X\Big)$$
$$\le \sum_{n=1}^\infty \sum_{k=1}^{2^n} \mathbb{P}\Big(\inf_{\|\varphi\|=1} \sup_{s,t\in I_{k,n}} \frac{|\langle \varphi, \delta X_{st}\rangle|}{2^{-n\theta}} \le \varepsilon \Big| \mathcal{F}_0^X\Big).$$

Applying Lemma 2 and noting that the bound obtained in this way is independent of k, we conclude that

$$\mathbb{P}\big(D_\theta(X) \le \varepsilon | \mathcal{F}_0^X\big) \le M \sum_{n=1}^\infty 2^n \exp\big(-c 2^{2n(\theta - H)} \varepsilon^{-2}\big) \le \tilde{M} \sum_{n=1}^\infty \exp\big(-\tilde{c} n \varepsilon^{-2}\big).$$

Here, we used the fact that we can find constants K and $\tilde{c}$ such that

$$n \log 2 - c 2^{2n(\theta-H)} \varepsilon^{-2} \le K - \tilde{c} n \varepsilon^{-2},$$

uniformly over all $\varepsilon \le 1$ and all $n \ge 1$. We deduce from this the bound

$$\mathbb{P}\big(D_\theta(X) \le \varepsilon | \mathcal{F}_0^X\big) \le M\Big(e^{-\tilde{c}\varepsilon^{-2}} + \int_1^\infty \exp\big(-\tilde{c}\varepsilon^{-2} x\big)\, dx\Big),$$

which immediately implies the result. $\square$

4. Malliavin derivatives. In this section, we derive formulas for the Malliavin derivatives of solutions to (1.1), when conditioned on the past of the driving noise. In order to clarify the meaning of this statement, we will reduce this conditioned solution to a functional of an underlying Wiener process. With this notation, the Malliavin derivative will simply be the "usual" Malliavin derivative of a random variable on Wiener space.

Before proceeding further, let us make a digression that clarifies this construction. For $\alpha \in (0, 1)$, we define the fractional integration operator $\mathcal{I}^\alpha$ and the corresponding fractional differentiation operator $\mathcal{D}^\alpha$ by

$$\begin{aligned} \mathcal{I}^\alpha f(t) &\equiv \frac{1}{\Gamma(\alpha)} \int_0^t (t-s)^{\alpha-1} f(s)\, ds, \\ \mathcal{D}^\alpha f(t) &\equiv \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_0^t (t-s)^{-\alpha} f(s)\, ds, \end{aligned} \tag{4.1}$$

with the convention that $\mathcal{I}^0$ and $\mathcal{D}^0$ denote the identity operator. The operators $\mathcal{I}^\alpha$ and $\mathcal{D}^\alpha$ are inverses of each other; see, for example, [41] for a survey of fractional integral operators.

It turns out that the operator $\mathcal{I}^{1/2-H}$ is an isometry between the Cameron–Martin space of the conditioned fBm and that of the underlying Wiener process mentioned at the beginning of this section. More precisely, given a typical instance $w_- \in \mathcal{C}(\mathbb{R}_-, \mathbb{R}^d)$ of the "past" of the fBm, it follows from the Mandelbrot–van Nesse representation of the fractional Brownian motion [18, 37] that there exists a constant α_H and a (one-sided) Wiener process W on $\mathbb{R}_+$ independent of w_- such that the future $w_+ \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$ of the fBm conditioned on the past w_- may be expressed as

$$w_+ = \mathcal{G}w_- + \alpha_H \mathcal{D}^{1/2-H} W, \tag{4.2}$$

where $\mathcal{D}^{1/2-H}$ is as defined in (4.1), and the operator $\mathcal{G}: \mathcal{C}(\mathbb{R}_-, \mathbb{R}^d) \mapsto \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$ is given by

$$(\mathcal{G}w_-)(t) \stackrel{\text{def}}{=} \gamma_H \int_0^\infty \frac{1}{r} g\left(\frac{t}{r}\right) w_-(-r)\, dr. \tag{4.3}$$

Here, the kernel g is given by

$$g(v) \stackrel{\text{def}}{=} v^{H-1/2} + (H - 3/2) v \int_0^1 \frac{(u+v)^{H-5/2}}{(1-u)^{H-1/2}}\, du, \tag{4.4}$$

and the constant γ_H is given by $\gamma_H = (H - \frac{1}{2})\alpha_H \alpha_{1-H}$, where α_H is the constant appearing in (4.2). The interpretation of the operator $\mathcal{G}$ is that $(\mathcal{G}w_-)(t)$ is the conditional expectation at time t of a two-sided fractional Brownian motion with Hurst parameter H, conditioned on coinciding with w_- for negative times.

Henceforth we will use the notation (4.2); namely we denote the past of the fBm by $w_- \in \mathcal{C}(\mathbb{R}_-, \mathbb{R}^d)$ and the future by $w_+ \in \mathcal{C}(\mathbb{R}, \mathbb{R}^d)$. At this stage, we will use a slight abuse of notation, and we will also sometimes interpret w_+ as an element in the space $\mathcal{C}_g^\gamma(\mathbb{R}_+, \mathbb{R}^d)$ of geometric rough paths that are γ-Hölder continuous, although we then usually denote it by $(X, \mathbb{X})$.

In view of (4.2), it will be useful to clarify how to interpret this identity when the future is considered as an element in the space $\mathcal{C}_g^\gamma(\mathbb{R}_+, \mathbb{R}^d)$, and for which instances of w_- the decomposition (4.2) makes sense. Recall that, for $(X, \mathbb{X}) \in \mathcal{C}_g^\gamma([0, T], \mathbb{R}^d)$ and $h \in \mathcal{C}([0, T], \mathbb{R}^d)$ a path with bounded variation, we can define a translated path $(Y, \mathbb{Y}) = \tau_h(X, \mathbb{X})$ in a natural way by

$$\begin{aligned} Y_t &= X_t + h_t, \\ \mathbb{Y}_{s,t} &= \mathbb{X}_{s,t} + \int_s^t \delta X_{s,r} \otimes dh_r + \int_s^t \delta h_{s,r} \otimes dX_r + \int_s^t \delta h_{s,r} \otimes dh_r. \end{aligned} \tag{4.5}$$

Since we assumed h to be of bounded variation, the integrals appearing in this expression should be interpreted as usual Riemann–Stieltjes integrals. Assume furthermore that h is such that there exists a constant $\|h\|_{1;\gamma}$ such that, for every $s \leq t$ in $[0, T]$, the variation of h over the interval $[s, t]$ is bounded by $\|h\|_{1;\gamma}|t - s|^\gamma$. In this case, it follows immediately that there exists M (depending on T) such that

$$\|Y - X\|_\gamma \leq \|h\|_{1;\gamma}, \qquad \|\mathbb{Y} - \mathbb{X}\|_{2\gamma} \leq M\|h\|_{1;\gamma}\big(\|h\|_{1;\gamma} + \|X\|_\gamma\big). \tag{4.6}$$

Similarly, we check that there exists a constant M such that

$$\big\|\tau_h(X,\mathbb{X}) - \tau_h(\tilde{X},\tilde{\mathbb{X}})\big\|_\gamma \le M\|h\|_{1;\gamma}\big\|(X,\mathbb{X}) - (\tilde{X},\tilde{\mathbb{X}})\big\|_\gamma, \tag{4.7}$$

so that τ_h is Lipschitz continuous as a map from $\mathcal{C}_g^\gamma([0,T],\mathbb{R}^d)$ to itself.

Denote now by $\mathcal{W}_\gamma$ the completion of $\mathcal{C}_0^\infty(\mathbb{R}_-;\mathbb{R}^d)$ with respect to the norm

$$|\!|\!|\omega|\!|\!|_\gamma \equiv \sup_{\substack{s,t\in\mathbb{R}_- \\ s\neq t}} \frac{|\omega(t)-\omega(s)|}{|t-s|^\gamma(1+|t|+|s|)^{1/2}}. \tag{4.8}$$

For $H \in (0,1)$ and $\gamma \in (0,H)$, it can be shown that there exists a probability measure $\mathbb{P}_-$ on $\mathcal{W}_\gamma$ such that the canonical process associated to $\mathbb{P}_-$ is a fractional Brownian motion with Hurst parameter H [18].

Notice now that the operator $\mathcal{G}$ given by (4.3) is actually defined on all of $\mathcal{W}_\gamma$. Indeed, similar to [22], Proposition A.2, it can be checked that the kernel g defined in equation (4.4) is smooth away from 0 and that its derivative satisfies

$$g'(t) = \mathcal{O}(1), \qquad t \ll 1, \qquad g'(t) = \mathcal{O}\big(t^{H-3/2}\big), \qquad t \gg 1. \tag{4.9}$$

It follows that, for $w_- \in \mathcal{W}_\gamma$, one has the bound

$$\big|(\mathcal{G}w_-)'(t)\big| \le C|\!|\!|w_-|\!|\!|_\gamma\big(t^{\gamma-1}+t^{\gamma-1/2}\big),$$

where $|\!|\!|\cdot|\!|\!|_\gamma$ is as in (4.8). In particular, over every finite time interval there exists a constant M such that

$$\|\mathcal{G}w_-\|_{1;\gamma} \le M|\!|\!|w_-|\!|\!|_\gamma. \tag{4.10}$$

As a consequence of this discussion, (4.2) makes sense in $\mathcal{C}_g^\gamma(\mathbb{R}_+,\mathbb{R}^d)$ for every $w_- \in \mathcal{W}_\gamma$, and this is how we will interpret this identity from now on.

4.1. *Derivatives of the solutions*. We now derive expressions for the derivatives of solutions to (1.1), both with respect to its initial condition and with respect to the driving noise. For this, we make the following assumption, which will be enforced throughout the whole article:

ASSUMPTION 1. The vector fields V_j are $\mathcal{C}^\infty$ and all of their derivatives grow at most polynomially fast. Furthermore, for every initial condition $z \in \mathbb{R}^n$, every final time T and every $(X,\mathbb{X}) \in \mathcal{D}_g^\gamma([0,T],\mathbb{R}^d)$ with $\gamma > \frac{1}{3}$, (1.1) has a solution up to time T.

REMARK 7. We assume polynomial growth so that we can bound the Malliavin derivatives in terms of moments of the Jacobian (see Theorem 4.1 below), a condition which is typically not too hard to verify. Otherwise, our conditions would be very awkward to state, for a rather minor gain in generality.

REMARK 8. Since the solution to (1.1) depends continuously on both its initial condition and the rough path $(X, \mathbb{X})$ [16], it follows from a simple compactness argument that, for every $R > 0$ and every final time $T > 0$, there exists a constant M such that, if we denote by (Z, Z') the solution to (1.1), the bound

$$\|(Z, Z')\|_{X,\gamma} \le M$$

holds uniformly over all initial conditions $|z| \le R$ and all driving noises $\|(X, \mathbb{X})\|_\gamma \le R$. Here, we use the fact that, over finite time intervals, the embedding $\mathcal{D}_g^\gamma \hookrightarrow \mathcal{D}_g^\beta$ is compact for $\beta < \gamma$ [16], Proposition 8.17, and that the continuous dependence on the driving path also holds in $\mathcal{D}_g^\beta$.

For an initial condition z and an instance of the driving noise $w = (w_-, w_+)$, let $\Phi_t(z, w_+)$ denote the solution map of (1.1),

$$Z_t = \Phi_t(z, w_+).$$

Note that for defining the solution, we only use w_+ and do not use w_-, the past of the driving noise. Define the Jacobian

$$J_{0,t} \stackrel{\text{def}}{=} \frac{\partial \Phi_t(z, w_+)}{\partial z},$$

and, for notational convenience, set $V = (V_1, V_2, \dots, V_d)$. Then the Jacobian $J_{0,t}$ and its inverse satisfy the (rough) evolution equations

$$dJ_{0,t} = DV_0(Z_t) J_{0,t}\, dt + DV(Z_t) J_{0,t}\, dX_t, \tag{4.11a}$$

$$dJ_{0,t}^{-1} = -J_{0,t}^{-1} DV_0(Z_t)\, dt - J_{0,t}^{-1} DV(Z_t)\, dX_t. \tag{4.11b}$$

Here, both J and J^{-1} are $n \times n$ matrices, and $J_{0,0} = J_{0,0}^{-1} = 1$. In order to deduce (4.11b) from (4.11a), we used the chain rule, which holds provided that $(X, \mathbb{X}) \in \mathcal{C}_g^\gamma$.

We now consider the effect on the solution of a variation, not of the initial condition, but of the driving noise itself. For this, we define the operators $\mathcal{A}_T : L^2([0, T], \mathbb{R}^d) \mapsto \mathbb{R}^n$ by

$$\mathcal{A}_T v = \int_0^T J_{0,s}^{-1} V(Z_s) v(s)\, ds. \tag{4.12}$$

A particular role will be played by $\mathcal{A}_T^* : \mathbb{R}^n \mapsto L^2([0, T], \mathbb{R}^d)$, the adjoint of $\mathcal{A}_T$, which is given by

$$(\mathcal{A}_T^* \xi)(s) = V(Z_s)^* (J_{0,s}^{-1})^* \xi, \qquad \xi \in \mathbb{R}^n. \tag{4.13}$$

It is known [16] that for every sample path w_+ of fractional Brownian motion in $\mathcal{D}_g^\gamma$ and for any fixed T, the map[3]

$$h \in \mathcal{H}_{H,+} \mapsto \Phi_T(z, w_+ + h) \tag{4.14}$$

[3] Since $\mathcal{D}_g^\gamma$ is not a linear space, the "addition" of the paths w_+ and h should be interpreted in the sense of (4.5) below.

is Fréchet differentiable, where $\mathcal{H}_{H,+}$ denotes the Cameron–Martin space of the Gaussian process w_+. In fact this Fréchet differentiability in Cameron–Martin directions holds in great generality for RDE solutions driven by rough paths [16]. Furthermore, setting $h(s) = \int_0^s v(r)\,dr$ for some v, one has the identity

$$D_h \Phi_T(z, w_+ + h)|_{h=0} = J_{0,T} \mathcal{A}_T v, \tag{4.15}$$

whenever $v \in L^2$. Note that, by (4.2), the space $\mathcal{H}_{H,+}$ consists of those paths h such that $h = \mathcal{D}^{1/2-H} \tilde{h}$ for some $\tilde{h}$ in the Cameron–Martin space of W, which in turn is equal to H^1, the space of square integrable functions with square integrable weak derivative. If $H \neq \frac{1}{2}$, the corresponding element v does not necessarily belong to L^2, so that one may wonder what the meaning of (4.15) is in general. Writing $F_s = J_{0,s}^{-1} V(Z_s)$ as a shorthand, a calculation shows that, for $1/3 < H < 1/2$, there is a constant c such that

$$\begin{aligned} &\mathcal{A}_T \mathcal{D}^{1/2-H} v \\ &\quad = c \int_0^T \biggl(\int_s^T (r-s)^{H-3/2} (F_s - F_r)\,dr + \frac{(T-s)^{H-1/2}}{1/2 - H} F_s \biggr) v(s)\,ds, \end{aligned} \tag{4.16}$$

so that $|\mathcal{A}_T \mathcal{D}^{1/2-H} v| \leq M \|F\|_{\mathcal{C}^\gamma} \|v\|_{L^2}$ for some constant M, provided that $\gamma > \frac{1}{2} - H$. Since, in our particular case, F is a rough path controlled by $(X, \mathbb{X})$, the condition $\|F\|_{\mathcal{C}^\gamma} < \infty$ for $\gamma > \frac{1}{2} - H$ can always be satisfied when $\frac{1}{2} - H < H$, namely when $H > \frac{1}{4}$. See [7] for more discussion on why a bound like this is true in general. The reason for deriving the explicit expression (4.16) in our case is that it will be useful in the next subsection.

In the sequel, we will write $\mathrm{D}_v Z_T^z$ as a shorthand for the derivative of the solution map in the direction $h = \int_0^\cdot v(s)\,ds$, that is,

$$\mathrm{D}_v Z_T^z = D_h \Phi_T(z, w_+) = J_{0,T} \mathcal{A}_T v. \tag{4.17}$$

We also set

$$\mathrm{D}_s Z_t^z = J_{s,t} V\bigl(Z_s^z\bigr) \stackrel{\text{def}}{=} J_{0,t} J_{0,s}^{-1} V\bigl(Z_s^z\bigr), \tag{4.18}$$

so that $\mathrm{D}_v Z_T^z$ is the L^2-scalar product of $\mathrm{D} Z_T^z$ with v.

4.2. *Malliavin differentiability of the solutions.* Using representation (4.2), the solution map $\Phi_t(z, w_+)$ conditioned on the past w_- of the driving noise may be viewed as a functional of an underlying Wiener process on $[0, \infty)$ which then allows us to define the Malliavin derivative of the solution map $\mathbf{D} Z_t^z = \mathbf{D}\Phi_t(z, w_+)$ in the usual way. For $H = 1/2$, we have $\mathbf{D}_s Z_t^z = \mathrm{D}_s Z_t^z$ where D is as defined in (4.18). Thus we focus on the case $H < 1/2$ below. As shown in [7], the Malliavin derivative is related to the Fréchet derivative given by (4.14) in the following way. For any $\int_0^\cdot v(s)\,ds \in \mathcal{H}_{H,+}$, define $\tilde{v} = \mathcal{I}^{1/2-H} v$. Then we have the identity

$$\mathbf{D}_{\tilde{v}} Z_T^z = \frac{1}{\alpha_H} \mathrm{D}_v Z_T^z = \frac{1}{\alpha_H} J_{0,T} \mathcal{A}_T v = \frac{1}{\alpha_H} J_{0,T} \mathcal{A}_T \mathcal{D}^{1/2-H} \tilde{v}, \tag{4.19}$$

where α_H is as in (4.2). In line with the notation from [39], we define $s \mapsto \mathbf{D}_s Z_t^z$ to be the stochastic process such that the relation

$$\mathbf{D}_{\tilde{v}} Z_t^z = \int_0^t \tilde{v}(s) \mathbf{D}_s Z_t^z \, ds$$

holds for every $\tilde{v} \in L^2$. Comparing this with (4.16), we see that one has the identity

$$\begin{aligned} \mathbf{D}_s Z_t^z &= c \int_s^t (r-s)^{H-3/2} \big(J_{s,t} V(Z_s^z) - J_{r,t} V(Z_r^z) \big) \, dr \\ &\quad + \frac{2c}{1-2H} (T-s)^{H-1/2} J_{s,t} V(Z_s^z) \end{aligned} \tag{4.20}$$

for some fixed constant c. (Furthermore, $\mathbf{D}_s Z_t^z = 0$ for $s \geq t$.) In general, we can rewrite this as

$$\mathbf{D} Z_t^z = \mathcal{D}_+^{1/2-H} \mathrm{D} Z_t^z, \tag{4.21}$$

where $\mathrm{D}_s Z_t^z$ is as in (4.18), with $\mathrm{D}_s Z_t^z = 0$ for $s \geq t$, and $\mathcal{D}_+^{1/2-H}$ is the linear operator given by

$$\big(\mathcal{D}_+^{1/2-H} f\big)(s) = c \int_s^\infty (r-s)^{H-3/2} \big(f(r) - f(s) \big) \, dr.$$

The aim of this section is to obtain a priori bounds on the higher-order Malliavin derivatives of the solution. As a first step, we obtain pointwise bounds on multiple derivatives of the solution map. In view of (4.18), we will need to compute $\mathrm{D}_s J_{0,t}$ in order to obtain such bounds. At this stage, let us put indices back into the various expressions in order to clarify the precise meaning of the various expressions that appear. We will use Einstein's convention of summation over repeated indices, and we write D_s^i for the derivative with respect to the ith component of the driving noise $(X, \mathbb{X})$. It is clear that $\mathrm{D}_s J_{0,t} = 0$ for $t < s$. Furthermore, we see from (4.11a) that

$$\mathrm{D}_s^i J_{0,s}^{k\ell} = D_m V_i^k (Z_s) J_{0,s}^{m\ell}. \tag{4.22}$$

For $t > s$, we formally differentiate (4.11a), and we use identity (4.18) to obtain the rough evolution equation

$$\begin{aligned} d\mathrm{D}_s^i J_{0,t}^{k\ell} &= D_m V_j^k (Z_t) \mathrm{D}_s^i J_{0,t}^{m\ell} \, dX^j(t) + D_m V_0^k (Z_t) \mathrm{D}_s^i J_{0,t}^{m\ell} \, dt \\ &\quad + D_{mn}^2 V_j^k (Z_t) J_{0,t}^{m\ell} J_{s,t}^{no} V_i^o (Z_t) \, dX^j(t) \\ &\quad + D_{mn}^2 V_0^k (Z_t) J_{0,t}^{m\ell} J_{s,t}^{no} V_i^o (Z_t) \, dt. \end{aligned}$$

Note now that this is a linear inhomogeneous equation for $\mathrm{D}_s^i J_{0,t}^{k\ell}$, where the linear part has exactly the same structure as (4.11a). As a consequence, we can solve it using the variation of constants formula which, when combined with (4.22), yields the expression

$$
\begin{aligned}
\mathrm{D}_s^i J_{0,t}^{k\ell} &= J_{s,t}^{kj} D_m V_i^j(Z_s) J_{0,s}^{m\ell} \\
&\quad + \int_s^t J_{r,t}^{kp} D_{mn}^2 V_j^p(Z_r) J_{0,r}^{m\ell} J_{s,r}^{nq} V_i^q(Z_r)\, dX^j(r) \qquad (4.23)\\
&\quad + \int_s^t J_{r,t}^{kp} D_{mn}^2 V_0^p(Z_r) J_{0,r}^{m\ell} J_{s,r}^{nq} V_i^q(Z_r)\, dr.
\end{aligned}
$$

A similar identity also holds for $\mathrm{D}_s J_{0,t}^{-1}$, but the precise form of this expression is unimportant as will be seen presently.

We now introduce the following notation in order to keep track of the terms appearing in the expressions for higher order Malliavin derivatives. Denote by T the space of finite rooted trees, and by F the space of finite forests (unordered finite collections of trees, allowing for repetitions). Formally, we denote by $(\tau_1, \dots, \tau_k)$ the forest consisting of the trees $\tau_1, \dots, \tau_k$. For $F = (\tau_1, \dots, \tau_k) \in \mathrm{F}$, we also write $[F] = [\tau_1, \dots, \tau_k] \in \mathrm{T}$ for the tree obtained by gluing the roots of the trees of F to a common new root.

For any forest $F \in \mathrm{F}$, we then build a sequence of subsets $\mathcal{V}_F^k \subset \mathcal{C}_X^\gamma([0,T], \mathbb{R})$ with $k \geq 1$ in the following way. For the empty forest $(\cdot)$, we set $\mathcal{V}_{(\cdot)}^k = \{1\}$ for every k. For $F = (\bullet)$, the forest consisting of one single tree, which itself consists only of a root, we set

$$
\mathcal{V}_{(\bullet)}^k = \{J_{0,\cdot}^{ij}, (J_{0,\cdot}^{-1})^{ij}, D_{i_1,\dots,i_\ell}^{(\ell)} V_m^j(Z_\cdot) : \ell \in \{0, \dots, k\}\},
$$

where $m \in \{0, \dots, d\}$, and the indices i, j and $i_1, \dots, i_\ell$ belong to $\{1, \dots, n\}$. For forests $F = (\tau_1, \dots, \tau_m)$ consisting of more than one tree, we set

$$
\mathcal{V}_{(\tau_1,\dots,\tau_m)}^k = \{Y_1 \cdots Y_m : Y_j \in \mathcal{V}_{(\tau_j)}^k, \forall j\}. \qquad (4.24)
$$

In other words, the processes contained in $\mathcal{V}_{(\tau_1,\dots,\tau_m)}^k$ are obtained by multiplying together the processes contained in $\mathcal{V}_{(\tau_j)}^k$. Finally, if F consists of a single tree consisting of more than just one root, so that $F = [G]$ for some forest G, we set

$$
\mathcal{V}_{[G]}^k = \left\{\int_0^\cdot Y_s\, dX^\ell(s), \int_0^\cdot Y_s\, ds : Y \in \mathcal{V}_G^k, \ell \in \{1, \dots, d\}\right\}. \qquad (4.25)
$$

This construction has the following feature:

LEMMA 4. *There exists a map $F \mapsto T_F$ from F to 2^{F}, the set of subsets of F, with the following properties*:

- *The set T_F is finite for every $F \in \mathrm{F}$.*

- *For every* $F \in \mathrm{F}$, $k \geq 1$, *there exist coefficients* $c^i_{Y,U,\bar{U}}$ *taking values in* $\{0, 1, -1\}$ *such that the identity*

$$\mathrm{D}^i_s Y_t = \sum_{G,\bar{G} \in T_F} \sum_{U \in \mathcal{V}^{k+1}_G} \sum_{\bar{U} \in \mathcal{V}^{k+1}_{\bar{G}}} c^i_{Y,U,\bar{U}} U_s \bar{U}_t$$

holds for every $Y \in \mathcal{V}^k_F$ *and every* $0 \leq s < t \leq T$.

REMARK 9. Objects indexed by trees arise naturally when considering higher-order expansions (in time) of solutions to differential equations [5], but this is completely unrelated to the construction of this section. In our case, the bound on the higher order Malliavin derivatives proceeds via an inductive argument, and the set of trees used here is simply one relatively easy combinatorial object having the same recursive structure as our bounds.

PROOF OF LEMMA 4. Note first that, by writing $J_{r,t} = J_{0,t} J^{-1}_{0,r}$ and similarly for $J_{s,r}$, we see from (4.23) that $\mathrm{D}_s J_{0,t}$ can indeed be written as

$$\mathrm{D}^i_s J^{k\ell}_{0,t} = \sum_{G,\bar{G} \in T_J} \sum_{U \in \mathcal{V}^2_G} \sum_{\bar{U} \in \mathcal{V}^2_{\bar{G}}} c^{ik\ell}_{U,\bar{U}} U_s \bar{U}_t$$

for some coefficients $c^{ik\ell}_{U,\bar{U}} \in \{0, 1, -1\}$, and for

$$T_J = \{(\bullet), (\bullet, \bullet, \bullet), (\bullet, [\bullet, \bullet, \bullet, \bullet, \bullet])\}.$$

The same statement holds true for $\mathrm{D}^i_s (J^{-1}_{0,t})^{k\ell}$. Furthermore, we have from (4.18) and the chain rule the identity

$$\mathrm{D}^i_s D^{(\ell)}_{i_1,\ldots,i_\ell} V^j_m(Z_t) = D^{(\ell+1)}_{i_1,\ldots,i_\ell,k} V^j_m(Z_t) J^{kp}_{s,t} V^p_i(Z_s),$$

so that we have

$$\mathrm{D}^i_s D^{(\ell)}_{i_1,\ldots,i_\ell} V^j_m(Z_t) = \sum_{G,\bar{G} \in T_D} \sum_{U \in \mathcal{V}^2_G} \sum_{\bar{U} \in \mathcal{V}^2_{\bar{G}}} \bar{c}_{U,\bar{U}} U_s \bar{U}_t$$

for some coefficients $\bar{c}_{U,\bar{U}} \in \{0, 1, -1\}$ (depending also on all the indices appearing on the left-hand side), and for

$$T_D = \{(\bullet, \bullet)\}.$$

It follows that we can indeed find a set $T_{(\bullet)} = T_J \cup T_D$ with the two properties stated in the lemma.

For more complicated forests, the claim follows by building T recursively in the following way. If $F = (\tau_1, \ldots, \tau_m)$ for trees τ_j such that $T_{(\tau_j)}$ is known, we observe that one has the identity

$$\mathrm{D}^j_s Y_1(t) \cdots Y_m(t) = \sum_{i=1}^m Y_1(t) \cdots Y_{i-1}(t) \mathrm{D}^j_s Y_i(t) Y_{i+1}(t) \cdots Y_m(t). \tag{4.26}$$

As a consequence, if we write $F \oplus G$ for the union of two forests and $F \ominus G$ for the forest obtained by removing from F its subforest G, we can set

$$T_{(\tau_1,\dots,\tau_m)} = \bigcup_{i=1}^{m} \{T_{(\tau_i)}, F \ominus (\tau_i) \oplus T_{(\tau_i)}\}. \tag{4.27}$$

It follows from (4.26) and (4.24) that this definition does indeed ensure that the requested properties are satisfied.

It remains to consider the case $F = (\tau)$ for some nontrivial tree τ. In this case, there exists a forest G such that $\tau = [G]$ and elements in $\mathcal{V}_F^k$ are given by (4.25). Note now that one has the identities

$$\mathrm{D}_s^i \int_0^t Y_r \, dX^\ell(r) = \delta_{i\ell} Y_s + \int_0^t \mathrm{D}_s^i Y_r \, dX^\ell(r),$$
$$\mathrm{D}_s^i \int_0^\cdot Y_r \, dr = \int_0^t \mathrm{D}_s^i Y_r \, dr.$$

As a consequence, if we set

$$T_{[G]} = T_G \cup \{G, (\cdot)\} \cup \{([H]) : H \in T_G\}, \tag{4.28}$$

the requested properties are again satisfied by induction. Since every forest can be built from elementary trees by the two operations considered in (4.27) and (4.28), this concludes the proof. □

For our purpose, this has the following useful consequence. For a fixed final time T, define the controlled rough path $\mathcal{J}^z \in \mathcal{D}_g^\gamma([0,T], \mathbb{R}^{2n^2+n})$ by

$$\mathcal{J}_t^z = (Z_t, J_{0,t}, J_{0,t}^{-1}). \tag{4.29}$$

(Note that both J and J^{-1} also implicitly depend on the starting point z.) We then have an a priori bound on the derivatives of the solution with respect to the driving noise in terms of $\mathcal{J}^z$:

PROPOSITION 2. *Let A_t denote any component of the vector $\mathcal{J}_t^z$. Under Assumption 1, for every multiindex $\alpha = (\alpha_1, \dots, \alpha_\ell)$ there exists a finite index set T_α and elements $F_j^k \in \mathcal{C}_X^\gamma([0,T], \mathbb{R})$ with $k \in T_\alpha$ and $j \in \{1, \dots, |\alpha| + 1\}$, such that the identity*

$$\mathrm{D}_{s_1}^{\alpha_1} \cdots \mathrm{D}_{s_\ell}^{\alpha_\ell} A_t = \sum_{k \in T_\alpha} F_1^k(s_1) \cdots F_\ell^k(s_\ell) F_{\ell+1}^k(t), \tag{4.30}$$

holds for every $0 \le s_1 < \cdots < s_\ell < t \le T$.

Furthermore, there exist constants M and p depending only on α and T such that the bound

$$\|F_j^k\|_{X,\gamma} \le M(1 + \|(\mathcal{J}^z, \mathcal{J}^{z\prime})\|_{X,\gamma})^p,$$

holds for every k and j.

PROOF. The first claim follows immediately from Lemma 4 by induction on $|\alpha|$. The second claim follows from the construction of the sets $\mathcal{V}_F^k$, combined with Lemma 1 and Theorem 2.1. □

In the particular case when X is fractional Brownian motion with Hurst parameter H, it follows from Proposition 2 that, if we consider the Malliavin derivatives $\mathbf{D}_s$ with respect to the underlying Wiener process as at the beginning of this section, we have the following bound:

THEOREM 4.1. *As above, let A_t denote any component of the vector $\mathcal{J}_t^z$, and let $(X, \mathbb{X})$ be fractional Brownian motion with Hurst parameter $H \in (\frac{1}{3}, \frac{1}{2}]$. Under Assumption 1, for every multiindex $\alpha = (\alpha_1, \dots, \alpha_\ell)$, every $\gamma \in (\frac{1}{3}, H)$, every $\delta > 0$ and every $T > 0$, there exist constants M and p such that the bound*

$$\left(|\mathbf{D}_{s_1}^{\alpha_1} \cdots \mathbf{D}_{s_\ell}^{\alpha_\ell} A_{s_{\ell+1}}| \prod_{j=1}^{\ell} |s_{j+1} - s_j|^{1-2H+\delta} \right) \le M \big(1 + \|(\mathcal{J}^z, \mathcal{J}^{z\prime})\|_{X,\gamma}\big)^p,$$

holds uniformly for all $0 \le s_1 < \cdots < s_{\ell+1} \le T$. Furthermore, the exponent p can be chosen to depend only on $|\alpha|$.

REMARK 10. Since the function $t \mapsto t^{2H-1-\delta}$ is square integrable near the origin for δ sufficiently small, the random variable A_t belongs to the stochastic Sobolev space $\mathbf{D}_{\text{loc}}^\infty$. If furthermore $\mathbb{E}\|(\mathcal{J}, \mathcal{J}')\|_{X,\gamma}^p < \infty$ for every p, then A_t belongs to the stochastic Sobolev space $\mathbf{D}^\infty$. See [39], page 49, for the definitions of $\mathbf{D}_{\text{loc}}^\infty$ and $\mathbf{D}^\infty$.

PROOF OF THEOREM 4.1. The proof for $H = 1/2$ follows trivially from Proposition 2. Consider now $s_1 < \cdots < s_\ell$ to be fixed and consider, for $j = 0, \dots, \ell$, the sequence of functions

$$F^{(j)}(r_{j+1}, \dots, r_\ell) = \mathbf{D}_{s_1} \cdots \mathbf{D}_{s_j} \mathrm{D}_{r_{j+1}} \cdots \mathrm{D}_{r_\ell} A_t,$$

so that our aim is to obtain a bound on $F^{(\ell)}$. Note that, by (4.21), the $F^{(j)}$ satisfy the recursive formula

$$\begin{aligned} F^{(j)}(r_{j+1}, \dots, r_\ell) = c \int_{s_j}^{\infty} (r - s_j)^{H-3/2} \big(& F^{(j-1)}(r, r_{j+1}, \dots, r_\ell) \\ & - F^{(j-1)}(s_j, r_{j+1}, \dots, r_\ell)\big)\, dr. \end{aligned} \tag{4.31}$$

We claim now that, for every j, there exists an index set T_j and a family of functions $F_i^{(j,k)}$ such that, for $s_j < r_{j+1} < \cdots < r_\ell < t$, one has the identity

$$F^{(j)}(r_{j+1}, \dots, r_\ell) = \sum_{k \in T_j} \prod_{i=j+1}^{\ell} F_i^{(j,k)}(r_i). \tag{4.32}$$

Furthermore, for every $\beta \in (\frac{1}{2} - H, H)$, there exists a constant M independent of $s_1, \dots, s_\ell$ such that these functions satisfy the bound

$$\prod_{i=j+1}^{\ell} \|F_i^{(j,k)}(r_i)\|_{\beta,j} \le M \prod_{i=1}^{j} |s_{j+1} - s_j|^{H-1/2-\beta} (1 + \|(\mathcal{J}^z, \mathcal{J}^{z'})\|_{X,\gamma})^p \tag{4.33}$$

for some fixed $p > 0$. Here, where we denote by $\|F\|_{\beta,j}$ the $\mathcal{C}^\beta$-norm (not semi-norm!) of F, restricted to the interval $[s_{j+1}, t]$. In the special case $j = \ell$, this is just $|F(t)|$. Once we show that (4.32) and (4.33) hold, the proof is complete since the special case $j = \ell$ and the choice $\beta = \frac{1}{2} - H + \delta$ yields the stated claim for δ sufficiently small. For larger values of δ, the claim can easily be reduced to that for small δ.

Note furthermore that $F^{(j)}(r_{j+1}, \dots, r_\ell) = 0$ if there exists $i > j$ such that $r_i > t$ and that the function $F^{(j)}$ is symmetric under permutations of its arguments. As a consequence, (4.32) is sufficient to determine $F^{(j)}$.

The proof now goes by induction over j. For $j = 0$, we have

$$F^{(0)}(r_1, \dots, r_\ell) = \mathrm{D}_{r_1} \cdots \mathrm{D}_{r_\ell} A_t,$$

which is indeed of the form (4.32) by Proposition 2. In this case, the bound (4.33) reduces to the statement that the $F_i^{(0,k)}$ are β-Hölder continuous, which is also a consequence of Proposition 2. In order to make use of the recursion (4.31), we have to rewrite it in such a way that the arguments of $F^{(j-1)}$ are always ordered. Using the recursion hypothesis, we then have the identity

$$\begin{aligned}
F^{(j)}&(r_{j+1}, \dots, r_\ell) \\
&= c \sum_{k \in T_{j-1}} \int_{s_j}^{r_{j+1}} (r - s_j)^{H-3/2} \big(F_j^{(j-1,k)}(r) - F_j^{(j-1,k)}(s_j)\big)\, dr \\
&\qquad \times \prod_{i>j} F_i^{(j-1,k)}(r_i) \\
&\quad + c \sum_{k \in T_{j-1}} \sum_{i>j} \int_{r_i}^{r_{i+1}} (r - s_j)^{H-3/2} F_i^{(j-1,k)}(r)\, dr \\
&\qquad \times \left(\prod_{q=j}^{i-1} F_q^{(j-1,k)}(r_{q+1})\right)\left(\prod_{q=i+1}^{\ell} F_q^{(j-1,k)}(r_q)\right) \\
&\quad - \frac{2c}{1-2H} (r_{j+1} - s_j)^{H-1/2} F_j^{(j-1,k)}(s_j) \prod_{i>j} F_i^{(j-1,k)}(r_i) \\
&\stackrel{\text{def}}{=} T_1 + T_2 + T_3.
\end{aligned}$$

Rewriting the integral from r_i to r_{i+1} appearing in T_2 as

$$\int_{r_i}^{t}(r-s_j)^{H-3/2}F_i^{(j-1,k)}(r)\,dr-\int_{r_{i+1}}^{t}(r-s_j)^{H-3/2}F_i^{(j-1,k)}(r)\,dr,$$

we see that $F^{(j)}$ is indeed again of the form (4.32). It remains to show that bound (4.33) holds. To show that it holds for T_1, write

$$G_k(s)=\int_{s_j}^{s}(r-s_j)^{H-3/2}\bigl(F_j^{(j-1,k)}(r)-F_j^{(j-1,k)}(s_j)\bigr)\,dr,$$

so that one has, for $s>s_j$, the bound

$$\bigl|\partial_s G_k(s)\bigr|\le(s-s_j)^{H-3/2+\beta}\|F_j^{(j-1,k)}\|_{\beta,j-1}.$$

In particular, one has for $s>s_{j+1}$ the bound

$$\bigl|\partial_s G_k(s)\bigr|\le(s_{j+1}-s_j)^{H-1/2}(s-s_j)^{\beta-1}\|F_j^{(j-1,k)}\|_{\beta,j-1}.$$

Furthermore, we obtain in a similar way the bound

$$\bigl|G_k(s_{j+1})\bigr|\le M(s_{j+1}-s_j)^{H-1/2+\beta}\|F_j^{(j-1,k)}\|_{\beta,j-1}$$

for some constant M, so that a straightforward calculation yields

$$\|G_k\|_{\beta,j}\le M(s_{j+1}-s_j)^{H-1/2}\|F_j^{(j-1,k)}\|_{\beta,j-1}$$

for some constant M. The requested bound on T_1 (actually a bound that it better than requested) then follows at once.

To bound T_2, we proceed similarly by setting

$$G_k(s)=\int_{s}^{t}(r-s_j)^{H-3/2}F_i^{(j-1,k)}(r)\,dr,$$

and noting that $G_k(t)=0$ and

$$\bigl|\partial_s G_k(s)\bigr|\le(s_{j+1}-s_j)^{H-1/2-\beta}(s-s_j)^{\beta-1}\|F_i^{(j-1,k)}\|_{\beta,j-1}.$$

It follows as above that

$$\|G_k\|_{\beta,j}\le M(s_{j+1}-s_j)^{H-1/2-\beta}\|F_i^{(j-1,k)}\|_{\beta,j-1}$$

as requested. Finally, the bound on T_3 follows in the same way. □

5. Regularity of laws. Our aim in this section is to show that if the vector fields V satisfy Hörmander's celebrated Lie bracket condition (see below), then the Malliavin matrix of the process Z_t is almost surely invertible, and to obtain quantitative bounds on its lowest eigenvalue.

In order to state Hörmander's condition, we define recursively the families of vector fields

$$\mathcal{V}_0=\{V_k:k\ge 1\},\qquad \mathcal{V}_{n+1}=\mathcal{V}_n\cup\bigl\{[U,V_k]:U\in\mathcal{V}_n,k\ge 0\bigr\},$$

where $[U, V]$ denotes the Lie bracket between the vector fields U and V. Note that under Assumption 1, the elements in $\mathcal{V}_n$ also have derivatives of all orders that grow at most polynomially. We now formulate Hörmander's bracket condition [24]:

ASSUMPTION 2. For every $z_0 \in \mathbb{R}^n$, there exists $N \in \mathbb{N}$ such that the identity

$$\operatorname{span}\{U(z_0) : U \in \mathcal{V}_N\} = \mathbb{R}^n \tag{5.1}$$

holds.

It is well known from the works of Malliavin, Bismut, Kusuoka, Stroock and others [4, 28–30, 35, 36, 38, 39] that when the driving noise X is Brownian motion, one way of proving the smoothness of the law of Z_T under Hörmander's condition is to first show the invertibility of the "reduced Malliavin matrix"[4]

$$C_T \stackrel{\text{def}}{=} \mathcal{A}_T \mathcal{A}_T^* = \int_0^T J_{0,s}^{-1} V(Z_s) V(Z_s)^* \big(J_{0,s}^{-1}\big)^* \, ds. \tag{5.2}$$

Recall that the matrix norm of a symmetric matrix is equal to its largest eigenvalue. Since C_T is a symmetric matrix, one can write the norm of its inverse as

$$\|C_T^{-1}\|^{-1} = \inf_{\|\varphi\|=1} \langle v, C_T v\rangle, \qquad \varphi \in \mathbb{R}^n. \tag{5.3}$$

5.1. *Deterministic bounds on* $\|C_T^{-1}\|$. In this subsection we only use the fact that $(X, \mathbb{X}) \in \mathcal{D}_g^\gamma$ and the fact that X is θ-rough for some $\theta > H$. Thus the bounds obtained are purely deterministic.

Before we turn to our bound on the inverse of C_T, let us introduce some notation. For any smooth vector field U, define the process $\mathcal{Z}_U(t) = J_{0,t}^{-1} U(Z_t)$, and set

$$\mathcal{R}_z \stackrel{\text{def}}{=} 1 + L_\theta(X)^{-1} + \|(X, \mathbb{X})\|_\gamma + \|(\mathcal{J}^z, \mathcal{J}^{z\prime})\|_{X,\gamma} + |z|, \tag{5.4}$$

where $\mathcal{J}^z$ is as in (4.29). Here, we fix a "roughness exponent" $\theta > H$ which will appear in subsequent statements.

LEMMA 5. *Fix a final time $T > 0$. Under Assumption 1, there exist constants $c, a > 0$ such that the bound*

$$\|\langle \varphi, \mathcal{Z}_U(\cdot)\rangle\|_\infty \le M \mathcal{R}_z^c |\langle \varphi, C_T \varphi\rangle|^a,$$

holds for all $U \in \mathcal{V}_1$, all $\varphi \in \mathbb{R}^n$ such that $\|\varphi\| = 1$, all initial conditions z, all $(X, \mathbb{X}) \in \mathcal{D}_g^\gamma([0, T], \mathbb{R}^d)$ and the constant $M > 0$ is independent of X, φ, z.

[4] This is a slight misnomer since our SDE is driven by fractional Brownian motion, rather than Brownian motion. One can actually rewrite the solution as a function of an underlying Brownian motion by making use of representation (4.2), but the associated Malliavin matrix has a slightly more complicated relation to C_T than usual. Still, it will be useful to first obtain a bound on the inverse of C_T.

PROOF. By definition we have

$$\langle \varphi, C_T\varphi\rangle = \sum_{i=1}^{d} \int_0^T \langle \varphi, J_{0,s}^{-1} V_i(Z_s)\rangle^2\, ds = \sum_{i=1}^{d} \|\langle \varphi, \mathcal{Z}_{V_i}(\cdot)\rangle\|_{L^2[0,T]}^2. \tag{5.5}$$

To obtain an upper bound of order $|\langle \varphi, C_T\varphi\rangle|^a$ on the supremum norm, our main tool is the interpolation inequality

$$\|f\|_\infty \le 2\max\bigl(T^{-1/2}\|f\|_{L^2[0,T]}, \|f\|_{L^2[0,T]}^{2\gamma/(2\gamma+1)}\|f\|_\gamma^{1/(2\gamma+1)}\bigr), \tag{5.6}$$

which holds for every γ-Hölder continuous function $f:[0,T]\mapsto \mathbb{R}$; see, for example, [23], Lemma A.3. Since in our case the final time T is fixed, the L^2 norm is controlled by the γ-Hölder norm, so that

$$\|\langle \varphi, \mathcal{Z}_{V_i}(\cdot)\rangle\|_\infty \le M \|\langle \varphi, \mathcal{Z}_{V_i}(\cdot)\rangle\|_{L^2[0,T]}^{2\gamma/(2\gamma+1)} \|\langle \varphi, \mathcal{Z}_{V_i}(\cdot)\rangle\|_{\mathcal{C}^\gamma}^{1/(2\gamma+1)}.$$

Since the vector fields V_i have derivatives with at most polynomial growth by assumption, we obtain immediately from Lemma 1 the bound

$$\|\langle \varphi, \mathcal{Z}_{V_i}(\cdot)\rangle\|_{\mathcal{C}^\gamma} \le M\mathcal{R}_z^a \tag{5.7}$$

for some exponent a. Combining this with (5.5), the claim follows at once. □

The next lemma involves an iterative argument (similar in spirit to [3, 29, 38]) to show that a similar bound holds with V_j replaced by any vector field obtained by taking finitely many Lie brackets between the V_j's.

LEMMA 6. *Fix a final time $T > 0$. Under Assumption 1, for every $i \ge 1$, there exist constants $c_i, a_i > 0$ such that*

$$\|\langle \varphi, \mathcal{Z}_U(\cdot)\rangle\|_\infty \le M\mathcal{R}_z^{c_i} |\langle \varphi, C_T\varphi\rangle|^{a_i}$$

for every $U \in \mathcal{V}_i$, every $\varphi \in \mathbb{R}^n$ such that $\|\varphi\| = 1$, every initial condition z, every $(X, \mathbb{X}) \in \mathcal{D}_g^\gamma([0,T], \mathbb{R}^d)$ and the constant $M > 0$ is independent of X, φ, z.

PROOF. The proof goes by induction over i. We already know from Lemma 5 that the statement holds for $i = 1$. Assume now that it holds for some $i \ge 1$, and let us show that it holds for $i + 1$. For any $t \le T$ and $U \in \mathcal{V}_i$, a simple application of the chain rule [which holds since $(X, \mathbb{X})$ is assumed to be a geometric rough path] yields

$$\langle \varphi, \mathcal{Z}_U(t)\rangle = \int_0^t \langle \varphi, \mathcal{Z}_{[V_0,U]}(s)\rangle\, ds + \sum_{j=1}^{d} \int_0^t \langle \varphi, \mathcal{Z}_{[V_j,U]}(s)\rangle\, dX_s^j, \tag{5.8}$$

where the second integral is a rough integral as in Theorem 2.1.

First we derive a priori bounds on the two integrands of (5.8) and then apply Theorem 3.1. It follows from Lemma 1 and Assumption 1 that

$$\big\|\langle\varphi, \mathcal{Z}_{[V_j,U]}(\cdot)\rangle\big\|_{X,\gamma} \le M\mathcal{R}_z^a, \qquad j = 0,\dots,d, \tag{5.9}$$

so that a similar bound holds on $\|\langle\varphi, \mathcal{Z}_{[V_j,U]}(\cdot)\rangle\|_\gamma$.

By the induction hypothesis, for every $U \in \mathcal{V}_i$ we have the bound $\|\langle\varphi, \mathcal{Z}_U(\cdot)\rangle\|_\infty \le M\mathcal{R}_z^{c_i}|\langle\varphi, C_T\varphi\rangle|^{a_i}$ for some constants a_i, c_i. Applying Theorem 3.1 to (5.8) and using the a priori bound (5.9), we conclude that there exist constants α_{i+1}, c_{i+1} such that

$$\big\|\langle\varphi, \mathcal{Z}_{[U,V_\ell]}(\cdot)\rangle\big\|_\infty \le M\mathcal{R}_z^{c_{i+1}}\big|\langle\varphi, C_T\varphi\rangle\big|^{\alpha_{i+1}}$$

for $\ell = 0,\dots,d$. Since $\mathcal{V}_{i+1}$ contains precisely the vector fields $[U, V_\ell]$, this concludes the proof. □

Now we combine the above two lemmas and Hörmander's hypothesis, Assumption 2, to obtain lower bounds on the smallest eigenvalue of C_T.

PROPOSITION 3. *Assume that Assumptions 1 and 2 hold. Fix $T > 0$, and let the matrix C_T and the quantity $\mathcal{R}$ be as defined in* (5.2) *and* (5.4), *respectively. Then there exists a constant $c > 0$ such that the bound*

$$\inf_{\|\varphi\|=1}\big|\langle\varphi, C_T\varphi\rangle\big| > M\mathcal{R}_z^{-c} \tag{5.10}$$

holds uniformly over every driving path $(X, \mathbb{X}) \in \mathcal{D}_g^\gamma$ and every initial condition z. The constant $M > 0$ is independent of X, φ, z.

REMARK 11. We emphasize again that (5.10) yields a lower bound on the eigenvalues of C_T that is not probabilistic in nature. All the probabilistic cancellations that take place in the classical probabilistic proofs of Hörmander's theorem are "hidden" in the strict positivity of $L_\theta(X)$ and the boundedness of $\|(X, \mathbb{X})\|_\gamma$.

PROOF OF PROPOSITION 3. Let $N \in \mathbb{N}$ be such that

$$K \stackrel{\text{def}}{=} \inf_{|z|\le\mathcal{R}}\inf_{\|\varphi\|=1}\sum_{U\in\bar{\mathcal{V}}_N}\big|\langle\varphi, U(z)\rangle\big|^2 > 0. \tag{5.11}$$

The existence of such an N follows from Assumption 2 and the smoothness of the vector fields V.

Note now that, considering the right-hand side at time 0, we see that

$$\big|\langle\varphi, U(z_0)\rangle\big|^2 \le \big\|\langle\varphi, \mathcal{Z}_U(\cdot)\rangle\big\|_\infty^2.$$

From Lemmas 5 and 6, there then exist constants c_N, α_N such that

$$K \le \inf_{\|\varphi\|=1}\sup_{U\in\bar{\mathcal{V}}_N}\big\|\langle\varphi, \mathcal{Z}_U(\cdot)\rangle\big\|_\infty^2 \le M\mathcal{R}_z^{c_N}\inf_{\|\varphi\|=1}\big|\langle\varphi, C_T\varphi\rangle\big|^{\alpha_N}, \tag{5.12}$$

which is precisely the required bound. □

Now let $\mathcal{M}_T$ be the Malliavin matrix of the map $W \mapsto Z_T^z$ where W is the underlying Wiener process from representation (4.2). Then we have the following pathwise bound on $\mathcal{M}_T$:

THEOREM 5.1. *Under the assumptions of Proposition* 3, *there exists a constant* $c_1 > 0$ *such that the bound*

$$\inf_{\|\varphi\|=1} \left|\langle \varphi, \mathcal{M}_T \varphi\rangle\right| > M \mathcal{R}_z^{-c_1} \tag{5.13}$$

holds for every driving path $(X, \mathbb{X}) \in \mathcal{D}^\gamma$, *every initial condition* z, *and the constant* $M > 0$ *is independent of* X, φ, z.

PROOF. By virtue of (4.19), we have the identity

$$\left|\langle \varphi, \mathcal{M}_T \varphi\rangle\right| = \left\|(\mathcal{D}^{1/2-H})^* \mathcal{A}_T^* J_{0,T}^* \varphi\right\|_{L_2[0,T]}^2, \tag{5.14}$$

where $(\mathcal{D}^{1/2-H})^*$ is the $L_2[0,T]$ adjoint of the operator $\mathcal{D}^{1/2-H}$ defined in (4.1). Notice that $\mathcal{I}^{1/2-H} : L_2[0,T] \mapsto L_2[0,T]$ is a bounded operator, and since $\mathcal{I}^{1/2-H}$ and $\mathcal{D}^{1/2-H}$ are inverses of each other, we conclude that operator $(\mathcal{D}^{1/2-H})^*$ has a bounded inverse in $L_2[0,T]$. Thus

$$\begin{aligned}\left\|(\mathcal{D}^{1/2-H})^* \mathcal{A}_T^* J_{0,T}^* \varphi\right\|_{L_2[0,T]} &\geq M \left\|\mathcal{A}_T^* J_{0,T}^* \varphi\right\|_{L_2[0,T]}^2 \\ &= M \langle J_{0,T}^* \varphi, C_T J_{0,T}^* \varphi\rangle,\end{aligned}$$

which, from Proposition 3, is bounded from below by

$$M \mathcal{R}_z^{-c} \left\|J_{0,T}^* \varphi\right\|^2 \geq M \mathcal{R}_z^{-c_1} \|\varphi\|^2,$$

where the last bound is a consequence of the fact that $\|J_{0,T}^{-1}\| \leq M \mathcal{R}_z$. □

5.2. *Probabilistic bounds and smoothness of laws*. Recall from (4.2) that the "future" evolution of the fBm conditional on the past w_- may be expressed as

$$w_+ = \mathcal{G} w_- + \alpha_H \mathcal{D}^{1/2-H} W,$$

where $\mathcal{G} w_-$ is the conditional expectation with the operator $\mathcal{G}$ given by (4.3). As in the previous section, we will mostly be interested in the situation when w_- is fixed, and the conditional law of the solution is considered. If $H = 1/2$, then all the statements are simple since in this case $\mathcal{G} = 0$ and $\mathcal{D}^0$ is the identity operator.

One problem is that it is in general quite difficult to obtain moment bounds on the Jacobian (and its inverse) for equations of the type (1.1) when the driving noise is only γ-Hölder for some $\gamma > \frac{1}{3}$ (rather than $\gamma > \frac{1}{2}$). The best bounds obtained in [16] rule out a downright explosion of the Jacobian, but only yield logarithmic moments in general. The very recent article [8] obtains such moment bounds, but under boundedness conditions that are stronger than Assumption 1. See also [14] for a related result. We therefore state the moment bounds on the solution and its Jacobian as an additional assumption. We will use $\tilde{\mathbb{E}}$ and $\tilde{\mathbb{P}}$ to denote the expectation and probability, respectively, conditioned on the past of the driving noise w_-.

ASSUMPTION 3. There exists an exponent $\zeta < 2$ and a seminorm $|\!|\!| \cdot |\!|\!|$ on $\mathcal{C}(\mathbb{R}_-, \mathbb{R}^d)$ such that $|\!|\!| w_- |\!|\!|$ is almost surely finite and such that, for every $R > 0$ and every $p \geq 1$, the bound

$$\tilde{\mathbb{E}} \| (\mathcal{J}^z, \mathcal{J}^{z'}) \|_{X,\gamma}^p \leq M \exp(M |\!|\!| w_- |\!|\!|^\zeta), \tag{5.15}$$

holds for some constant M independent of X, uniformly over all initial conditions with $|z| \leq R$.

REMARK 12. Combining (5.15) with Fernique's theorem immediately yields the unconditioned bound

$$\mathbb{E} \| (\mathcal{J}^z, \mathcal{J}^{z'}) \|_{X,\gamma}^p \leq M \tag{5.16}$$

for any $p \geq 1$.

Now we combine the results above with the results from the previous section to obtain probabilistic bounds on the inverse of the Malliavin matrix, under the additional hypothesis that Assumption 3 holds.

PROPOSITION 4. *Let* (1.1) *be such that Assumptions* 1, 2 *and* 3 *are satisfied. Fix* $T > 0$, *and let* $\mathcal{M}_T$ *be the Malliavin matrix as in* (5.13).

Then, there exists a norm $|\!|\!| \cdot |\!|\!|$ *such that* $|\!|\!| w_- |\!|\!| < \infty$ *almost surely and, for any* $R > 0$ *and any* $p \geq 1$, *there exists a constant* M *such that the bound*

$$\tilde{\mathbb{P}}\Big(\inf_{\|\varphi\|=1} \langle \varphi, \mathcal{M}_T \varphi \rangle \leq \varepsilon \Big) \leq M e^{M |\!|\!| w_- |\!|\!|^\zeta} \varepsilon^p, \tag{5.17}$$

holds for all $\varepsilon \in (0, 1]$ *and all initial conditions* z *with* $|z| \leq R$. *Here, the constant* ζ *is as in* (5.15).

Similarly, the unconditional bound

$$\mathbb{P}\Big(\inf_{\|\varphi\|=1} \langle \varphi, \mathcal{M}_T \varphi \rangle \leq \varepsilon \Big) \leq M \varepsilon^p \tag{5.18}$$

holds.

PROOF. From Theorem 5.1 we deduce that for small enough ε,

$$\tilde{\mathbb{P}}\Big(\inf_{\|\varphi\|=1} |\langle \varphi, \mathcal{M}_T \varphi \rangle| \leq \varepsilon \Big) \leq \tilde{\mathbb{P}}(\mathcal{R} \geq \varepsilon^{-c_1})$$

for some constant $c_1 > 0$. By Markov's inequality, for any $p \geq 1$, this expression is bounded by

$$M\varepsilon^{pc_1}\tilde{\mathbb{E}}\mathcal{R}^p.$$

Now for any $p \geq 1$, from Lemma 3 and Assumption 3 it follows that

$$\tilde{\mathbb{E}}\big(L_\theta(X)^{-p} + \|(\mathcal{J}^z, \mathcal{J}^{z\prime})\|_{X,\gamma}^p\big) \leq Me^{M|||w_-|||^\zeta}.$$

Furthermore, it follows from (4.2) that $\tilde{\mathbb{E}}\|(X, \mathbb{X})\|_\gamma^p \leq Me^{M\|\mathcal{G}w_-\|_\gamma^\zeta}$, thus proving claim (5.17). The second claim then follows from Fernique's theorem. □

As an immediate corollary, we obtain that the Malliavin matrix has all moments:

COROLLARY 2. *Under the assumptions of Proposition 4, the matrix $\mathcal{M}_T$ is almost surely invertible and, for any $R > 0$ and any $p \geq 1$,*

$$\tilde{\mathbb{E}}(\|\mathcal{M}_T^{-p}\|) \leq Me^{M|||w_-|||^\zeta},$$

uniformly over all initial conditions z of (1.1) such that $|z| \leq R$.

As a consequence of Proposition 4, we obtain the smoothness of the laws of Z_t conditioned on an instance of the past w_-:

THEOREM 5.2. *Let (1.1) be such that Assumptions 1, 2 and 3 are satisfied.*

Then, for every realization of the past w_- with $|||w_-||| < \infty$, every initial condition z and every $t > 0$, the conditional distribution of Z_t^z has a smooth density $p(x; z, w_-)$ with respect to Lebesgue measure.

Furthermore, for every multiindex α, the derivative $\partial_x^\alpha p(x; z, w_-)$ has finite moments of all orders, so that the unconditioned distribution $p(x; z) = \mathbb{E}p(x; z, w_-)$ of Z_t^z also has moments of all orders.

REMARK 13. The norm $|||\cdot|||$ appearing in the statement is the same as the one appearing in Assumption 3.

PROOF OF THEOREM 5.2. Combining Theorem 4.1 with Assumption 3, we see that the random variable Z_t^z belongs to the space $\mathbf{D}^\infty$. The claim then immediately follows from the fact that the Malliavin matrix has inverse moments of all orders [39].

The claim about the moments of the density follows from the fact that, by (5.15), $\|\mathcal{M}_T^{-1}\|$ and all Malliavin derivatives of Z_t^z also have unconditional moments of all orders. □

5.3. *A cutoff argument*. While [8] provides a large collection of examples for which Assumption 3 holds, this condition is not always easy to check. In this section, we therefore provide a cutoff argument that allows us to still show the existence of a density for the law of the solutions to (1.1) under Hörmander's condition, without assuming that Assumption 3 holds. Actually, we show slightly more than the mere existence of a density; namely, we show that the density can be approximated from below by a sequence of smooth densities. More precisely, the main result of this section is the following:

THEOREM 5.3. *Assume that Assumptions 1 and 2 hold, and denote by μ_t the conditional law of the solution to (1.1) at time $t > 0$, with fixed initial condition $z \in \mathbb{R}^n$.*

Then, there exists a sequence of increasing positive measures μ_t^n with $\mathcal{C}^\infty$ densities ρ_t^n such that $\lim_{n\to\infty} \mu_t^n(A) = \mu_t(A)$ for every Borel set A. In particular, μ_t has a density ρ_t with respect to Lebesgue measure and $\lim_{n\to\infty} \rho_t^n(x) = \rho_t(x)$ for Lebesgue-almost every x.

REMARK 14. The statement that ρ_t can be approximated from below by smooth functions is strictly stronger than just $\rho_t \in L^1$, which was already obtained in [6]. An example of a density function that cannot be approximated in this way would be the characteristic function of a Cantor set with positive Lebesgue measure.

PROOF OF THEOREM 5.3. The idea is to perform the following cutoff argument. For $\beta > 0$ real, $T \geq t$ and $q \geq 2$ an even integer, we define the function

$$\Lambda_{\beta,q,T}(X, \mathbb{X}) = \int_0^T \int_0^t \frac{|\delta X_{s,t}|^{2q} + \|\tilde{\mathbb{X}}_{st}\|^q}{|t-s|^{2\beta q}} \, ds \, dt, \tag{5.19}$$

where we denote by $\tilde{\mathbb{X}}$ the antisymmetric part of $\mathbb{X}$. This function has the following desirable properties:

(1) From the scaling of the covariance function for fractional Brownian motion and the equivalence of moments for Gaussian measures, we conclude that if $(X, \mathbb{X})$ is fractional Brownian motion with Hurst parameter H, then $\Lambda_{\beta,q,T}(X, \mathbb{X})$ has finite (conditional) moments of all orders, provided that $\beta < H$.

(2) For every $\gamma \in (0, H)$ and every $\beta \in (\gamma, H)$, there exists $q > 0$ and $M > 0$ such that

$$\|X\|_\gamma^2 + \|\mathbb{X}\|_{2\gamma} \leq M \Lambda_{\beta,q,T}^{1/q}(X, \mathbb{X}). \tag{5.20}$$

A proof of this fact can be found in [16], page 149. Note that since we assume $(X, \mathbb{X})$ to be geometric, the symmetric part of $\mathbb{X}_{s,t}$ is given by $\delta X_{s,t} \otimes \delta X_{s,t}$, so that it is indeed sufficient to control the increments of X and the antisymmetric part of $\mathbb{X}$.

(3) For every $(X, \mathbb{X}) \in \mathcal{D}^\gamma$, the map

$$\mathcal{H}_{H,+} \ni h \mapsto \Lambda_{\beta,q,T}\big(\tau_h(X, \mathbb{X})\big),$$

is Fréchet differentiable to all orders [16], where τ_h is the "translation map" as defined in (4.5) below. In particular, the map $w_+ \mapsto \Lambda_{\beta,q,T}(X(w_+), \mathbb{X}(w_+))$ belongs to the space $\mathbf{D}^\infty$ of random variables that are Malliavin differentiable of all orders with all Malliavin derivatives having moments of all orders. The precise statement of this fact is given in Proposition 6 in the Appendix.

The proof is now straightforward. First of all, we let $\gamma < H$ be as in the previous sections, let $\beta \in (\gamma, H)$ and fix q large enough so that (5.20) holds. We also let $\chi : \mathbb{R}_+ \to \mathbb{R}_+$ be a $\mathcal{C}^\infty$ nonincreasing cut-off function so that $\chi(\lambda) = 1$ for $\lambda \le 1$ and $\chi(\lambda) = 0$ for $\lambda \ge 2$. With these definitions at hand, we set

$$\Psi_n(w_+) \stackrel{\text{def}}{=} \chi\big(n^{-1}\Lambda_{\beta,q,T}\big(X(w_+), \mathbb{X}(w_+)\big)\big). \tag{5.21}$$

Fix furthermore $z \in \mathbb{R}^n$, and as before denote by $\Phi_t(z, w_+)$ the Itô map, so that $\mu_t = \Phi_t^*\mathbb{P}$. We then set $\mu_t^n = \Phi_t^*(\Psi_n\mathbb{P})$. In other words, we have the identity

$$\mu_t^n(A) = \int_{\Phi_t^{-1}(A)} \Psi_n(w_+)\mathbb{P}(dw_+),$$

valid for every measurable set $A \subset \mathbb{R}^n$. Since $\Lambda_{\beta,q,T}$ is almost surely finite, we clearly have $\mu_t^n(A) \nearrow \mu_t(A)$ for every measurable set A, so that the claim follows if we can show that every μ_t^n has a smooth density. This in turn follows by Malliavin's lemma [39] if we are able to show that, for every bounded open set $K \subset \mathbb{R}^n$ and every multiindex α, there exists a constant M such that the bound

$$\tilde{\mathbb{E}}\big(D^\alpha G\big(\Phi_t(z, w_+)\big)\Psi_n(w_+)\big) \le M(w_-) \sup_{x \in K}\big|G(x)\big|$$

holds uniformly over all test functions $G : \mathbb{R}^n \to \mathbb{R}$ that are $\mathcal{C}^\infty$ and supported in K.

Let $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$, $\alpha_i \in \{1, 2, \dots, n\}$. Using the chain rule and the integration by parts formula from Malliavin calculus, we have the identity

$$\begin{aligned} &\tilde{\mathbb{E}}\big(D^\alpha G\big(\Phi_t(z, w_+)\big)\Psi_n(w_+)\big) \\ &\quad = \tilde{\mathbb{E}}\big(G\big(\Phi_t(z, w_+)\big)H_\alpha\big(\Phi_t(z, w_+), \Psi_n(w_+)\big)\big), \end{aligned} \tag{5.22}$$

where the random variables H_α are defined as follows. For $\alpha = \varnothing$, the empty multiindex, we set $H_\varnothing = \Psi_n$. Furthermore, given a random variable G and an index α_1, we set

$$\mathcal{H}_{\alpha_1}(G) = \mathbf{D}^*\left(G \sum_{j=1}^n (\mathcal{M}_t^{-1})_{\alpha_1 j} \mathbf{D}_\cdot^j \Phi_t(z, w_+)\right) \tag{5.23}$$

with these definitions at hand, and it is straightforward to see that, for $\alpha = (\alpha_1, \dots, \alpha_k)$, we have

$$H_\alpha = \mathcal{H}_{\alpha_1}(H_{(\alpha_2,\dots,\alpha_k)}).$$

Fortunately, all of these expressions can be controlled in the following way. Define the set

$$\mathcal{S}_n \stackrel{\text{def}}{=} \{w : \Lambda_{\beta,q,T}(X(w_+), \mathbb{X}(w_+)) \le 2n\}. \tag{5.24}$$

It then follows from the local property of the Skorokhod integral [39], Proposition 1.3.15, that $H_\alpha(w_+) = 0$ for $w_+ \notin \mathcal{S}_n$. As a consequence, we also have the identity

$$H_\alpha = \tilde{\mathcal{H}}_{\alpha_1}(H_{(\alpha_2,\dots,\alpha_k)}),$$

where

$$\tilde{\mathcal{H}}_{\alpha_1}(G) = \mathbf{D}^*\left(G \sum_{j=1}^n (\Psi_{2n}(\mathcal{M}_t^{-1})_{\alpha_1 j})\mathbf{D}^j_\cdot(\Psi_{2n}\Phi_t(z, w_+))\right).$$

Note now that, by Corollary 4, Theorems 5.1 and 4.1, both $\Psi_{2n}(\mathcal{M}_t^{-1})_{\alpha_1 j}$ and $\Psi_{2n}\Phi_t(z, w_+)$ belong to the stochastic Sobolev space $\mathbf{D}^{\ell,p}$ for every $\ell, p > 1$, uniformly over every w_- such that $|||w_-|||_\gamma \le R$ for any $R > 0$, where $|||w_-|||_\gamma$ was defined in (4.8).

As a consequence, for $\alpha = (\alpha_1, \dots, \alpha_k)$ and $\ell > 0$, we have the bound

$$\tilde{\mathbb{E}}\|\mathbf{D}^{(\ell)} H_\alpha\|^p \le \mathrm{K}(|||w_-|||) \sum_{m \le \ell+1} (\tilde{\mathbb{E}}\|\mathbf{D}^{(m)} H_{(\alpha_2,\dots,\alpha_k)}\|^{2p})^{1/2},$$

where we denote by $\mathbf{D}^{(k)}$ the kth iterated Malliavin derivative and by $\|\cdot\|$ the L^2-norm. Since $H_\varnothing$ also belongs to $\mathbf{D}^{\ell,p}$ for every $\ell, p > 1$ by Corollary 4, the claim then follows. □

6. Ergodicity of SDEs driven by fBm.

The aim of this section is to use the preceding results in order to obtain ergodicity results for stochastic differential equations driven by fractional Brownian motion. In order to do this, we make use of the abstract framework introduced in [18] and further refined in [19, 22]. This allows us to introduce a notion of a "strong Feller property" for a large class of equations driven by nonwhite noise, together with a corresponding version of the Doob–Khasminskii theorem, stating that the strong Feller property, combined with a form of topological irreducibility and a quasi-Markovian property, is sufficient to deduce the uniqueness of an "invariant measure" in a suitable sense.

In order to use this framework, we view solutions to (1.1) as a *discrete-time* Markov process on a space of the type $\mathcal{W} \times \mathbb{R}^n$, where $\mathcal{W}$ contains all the information about the driving noise X required to solve (1.1) over a (fixed) time interval 1 and to predict the law of its future evolution. In our case, it is natural to choose $\mathcal{W}$ to be of the form

$$\mathcal{W} = \mathcal{W}_- \oplus \mathcal{W}_+,$$

where $\mathcal{W}_-$ contains the "past" of the driving noise up to time 0, and $\mathcal{W}_+$ contains the noise between times 0 and 1. The reason for splitting our space explicitly into two parts is that in order to be able to give a meaning to solutions to (1.1), we consider the driving noise as a rough path; that is, we choose $\mathcal{W}_+ = \mathcal{D}_g^\gamma([0,1], \mathbb{R}^d)$ for some $\gamma \in (\frac{1}{3}, H)$. (Recall that $\mathcal{D}_g^\gamma$ is the closure of the set of lifts of smooth functions in $\mathcal{D}^\gamma$.)

On the other hand, in order to recover the conditional law of fractional Brownian motion given its past, iterated integrals are not needed, and it is sufficient to retain information about the path itself. Therefore, it makes sense to choose $\mathcal{W}_-$ in a way similar to [18]; namely, we choose $\mathcal{W}_- = \mathcal{W}_\gamma$ for some $\gamma < H$, where $\mathcal{W}_\gamma$ was defined in (4.8). Denote as before by $\mathbb{P}_-$ the measure on $\mathcal{W}_\gamma$ such that the canonical process is a fractional Brownian motion with Hurst parameter H under $\mathbb{P}_-$.

For any given $w_- \in \mathcal{W}_-$, we now construct a measure $\hat{\mathcal{P}}(w_-, \cdot)$ on $\mathcal{W}_+$ as the law of a two-sided fractional Brownian motion, conditioned on its past w_-, and enhanced with the corresponding "area process." To this end, let us first denote by $\tilde{\mathbb{P}}_+$ the law of the stochastic process $\{X_t\}_{t\in[0,1]}$, given by

$$X_t = \alpha_H \int_0^t (t-r)^{H-1/2}\, dW_r, \tag{6.1}$$

where W is a standard Wiener process, and α_H is the constant appearing in (4.2). It can be checked that the covariance of $\tilde{\mathbb{P}}_+$ satisfies the assumptions of [9, 15], so that it can be lifted in a canonical way to a measure $\mathbb{P}_+$ on $\mathcal{W}_+$.

With this definition at hand, we define a Markov transition kernel $\hat{\mathcal{P}}$ from $\mathcal{W}_-$ to $\mathcal{W}_+$ by

$$\hat{\mathcal{P}}(w_-, \cdot) = \tau_{\mathcal{G}w_-}^* \mathbb{P}_+,$$

with the shift operator $\tau_{\mathcal{G}w_-}$ as in (4.5). It follows from (4.10), (4.6) and (4.7) that $\hat{\mathcal{P}}$ is Feller. Furthermore, it determines a measure $\mathbb{P}$ on $\mathcal{W} = \mathcal{W}_- \times \mathcal{W}_+$ in a natural way by

$$\mathbb{P}(dw_- \times dw_+) = \mathbb{P}_-(dw_-)\hat{\mathcal{P}}(w_-, dw_+).$$

It follows from our construction that if we denote by $\Pi: \mathcal{W} \to \mathcal{C}((-\infty, 1], \mathbb{R}^d)$, the natural map that concatenates w_- with the "path" component of w_+, then the image of $\mathbb{P}$ under Π is precisely the law of a two-sided fractional Brownian motion with Hurst parameter H. Similarly, we have a natural shift map $\Theta: \mathcal{W} \to \mathcal{W}_-$ that consists of composing Π with the usual time-1 shift map that maps $\mathcal{C}((-\infty, 1], \mathbb{R}^d)$ into $\mathcal{C}((-\infty, 0], \mathbb{R}^d) \supset \mathcal{W}_-$. It follows from the definitions of $\mathcal{W}_-$ and $\mathcal{W}_+$ that the map Θ is actually continuous. This construction allows us to lift $\hat{\mathcal{P}}$ to a Feller Markov transition kernel on $\mathcal{W}$ by

$$\mathcal{P}(w, \cdot) = \delta_{\Theta(w)} \otimes \hat{\mathcal{P}}(\Theta(w), \cdot).$$

It also follows from our construction that $\mathbb{P}$ is invariant (and ergodic) for $\mathcal{P}$. Indeed, the action of $\mathcal{P}(w,\cdot)$ is to shift the "path" component of w backwards by one time unit and to then concatenate it with the canonical lift to $\mathcal{W}_+$ of a piece of two-sided fractional Brownian motion, conditional on its past being given by $\Theta(w)$.

We now combine the noise process $(\mathcal{W},\mathbb{P},\mathcal{P})$ with the solution map for (1.1) in the following way. As before let $\Phi_\cdot(z,w_+)$ denote the map that solves (1.1) for a given initial condition z and a given realization $w=(w_-,w_+)$ of the driving noise. Since the Itô map is continuous on the space of rough paths with fixed Hölder regularity [16], the map Φ is continuous. We can then view the solutions to (1.1) as a Markov process on $\mathbb{R}^n\times\mathcal{W}$ with transition probabilities given by

$$\mathcal{Q}(z,w;\cdot)=\Phi_z^*\mathcal{P}(w,\cdot),$$

where we define $\Phi_z\colon\mathcal{W}\to\mathbb{R}^n\times\mathcal{W}$ by $\Phi_z(w)=(\Phi_1(z,w_+),w)$. In other words, we first shift back the noise by a time interval 1, then draw a sample from the conditional realization of an enhanced fractional Brownian motion on [0, 1], and then use this sample to solve (1.1) between 0 and 1.

The aim of this section is to show that the Markov operator $\mathcal{Q}$ admits a unique invariant measure, modulo a natural equivalence relation described in Section 6.1 below. Note that while $\mathcal{Q}$ is Feller (since Φ is continuous and $\mathcal{P}$ is Feller), it is certainly not strong Feller in the usual sense. We will, however, show in Section 6.1 that there is a natural generalization of the strong Feller property in this context that, in a way, only considers the part of $\mathcal{Q}$ in $\mathbb{R}^n$. In this generalized sense, it turns out that the invertibility of the Malliavin matrix shown in the preceding sections allows us to prove that $\mathcal{Q}$ satisfies the strong Feller property in this generalized sense. Combined with a form of topological irreducibility and a "quasi-Markov" property, this is then sufficient to deduce the uniqueness of the invariant measure for $\mathcal{Q}$ modulo equivalence of the induced laws on the space of trajectories on $\mathbb{R}^n$.

6.1. *General uniqueness criterion for the invariant measure.* From now on, we use the notation $\mathcal{X}=\mathbb{R}^n$ in order to simplify notation and to emphasize the fact that the results do not depend on the linear structure of the space.

The aim of this section is to study the uniqueness of "invariant measures" for (1.1). The question of uniqueness of the invariant measure for the SDE (1.1) should not be interpreted as the question of uniqueness of the invariant measure for the Markov operator $\mathcal{Q}$ constructed in the previous section. This is because one might imagine that the augmented phase space $\mathcal{X}\times\mathcal{W}$ contains some "redundant" randomness that is not necessary to describe the stationary solutions to (1.1). (This would be the case, e.g., if the V_i's are not always linearly independent.) One would like therefore to have a concept of uniqueness for the invariant measure that is independent of the particular description of the driving noise.

To this end, we introduce the Markov transition kernel $\bar{\mathcal{Q}}$ from $\mathcal{X}\times\mathcal{W}$ to $\mathcal{X}^{\mathbf{N}}$ constructed in the following way. Denote by (z_n,w_n) a sample of the Markov chain with transition probabilities $\mathcal{Q}$ starting at (z_0,w_0). We then denote by $\bar{\mathcal{Q}}(z_0,w_0;\cdot)$

the law of $(z_1, z_2, \dots)$. (We do not include the starting point, consistent with the convention that $0 \notin \mathbb{N}$.)

With this notation, we have a natural equivalence relation between measures on $\mathcal{X} \times \mathcal{W}$ given by

$$\mu \sim \nu \quad \Leftrightarrow \quad \bar{\mathcal{Q}}\mu = \bar{\mathcal{Q}}\nu. \tag{6.2}$$

In other terms, two measures on $\mathcal{X} \times \mathcal{W}$ are equivalent if they generate the same dynamics in $\mathcal{X}$. In the particular case when the process in $\mathcal{X}$ is Markov, $\bar{\mathcal{Q}}$ is independent of w, and the equivalence relation simply states that the marginals on $\mathcal{X}$ should agree. Denoting by $\|\cdot\|_{\mathrm{TV}}$ the total variation norm, this suggests that the following is a good generalization of the strong Feller property to our setting:

DEFINITION 4. The solutions to (1.1) are said to be *strong Feller* if there exists a jointly continuous function $\ell: \mathcal{X}^2 \times \mathcal{W} \to \mathbb{R}_+$ such that

$$\|\bar{\mathcal{Q}}(z, w; \cdot) - \bar{\mathcal{Q}}(y, w; \cdot)\|_{\mathrm{TV}} \le \ell(z, y, w), \tag{6.3}$$

and $\ell(z, z, w) = 0$ for every $z \in \mathcal{X}$ and every $w \in \mathcal{W}$.

We stress again that the definition given here has essentially *nothing* to do with the strong Feller property of $\mathcal{Q}$. It rather generalizes the notion of the strong Feller property for the Markov process associated to (1.1) in the case where the driving noise is white in time. See, for example, the review article [19] for more details.

DEFINITION 5. The solutions to (1.1) are said to be *topologically irreducible* if, for every $z \in \mathcal{X}$, $w \in \mathcal{W}$ and every nonempty open set $U \subset \mathcal{X}$, one has $\mathcal{Q}(z, w; U \times \mathcal{W}) > 0$.

REMARK 15. In order to prove topological irreducibility, one usually uses some form of the Stroock–Varadhan support theorem [42]. A version of this theorem was shown in the present context to hold in [16], Theorem 15.63. This shows that, in order to verify that (1.1) is topologically irreducible, it suffices to show that, for every $x_0 \in \mathbb{R}^n$, the set of points that are obtained as the solution at time $t = 1$ to

$$\dot{x}(t) = V_0\big(x(t)\big) + \sum_{i=1}^{d} V_i\big(x(t)\big)u_i(t), \qquad x(0) = x_0,$$

with $u \in \mathcal{C}^\infty([0, 1], \mathbb{R}^d)$ is dense in $\mathbb{R}^n$.

The following result, which is a consequence of [22], Theorem 3.10, is a generalization of the well-known Doeblin–Doob–Khasminskii criterion for the uniqueness of the invariant measure of a general Markov chain:

THEOREM 6.1. *If the solutions to* (1.1) *are strong Feller and and topologically irreducible*, *then* (1.1) *can have at most one invariant measure*, *modulo the equivalence relation* (6.2).

PROOF. The only missing ingredient to be able to apply [22], Theorem 3.10, is the "quasi-Markov" property of the solutions to a stochastic differential equation driven by fractional Brownian motion with $H \in (\frac{1}{3}, \frac{1}{2})$. For the case $H > \frac{1}{2}$, this property was shown to hold in [22], Proposition 5.11, and in the case $H = \frac{1}{2}$, solutions are Markovian anyway. The proof in the case $H < \frac{1}{2}$ is virtually identical, so we only sketch it. It only uses the fact that the set $\mathcal{X} = \{h \in \mathcal{C}^\infty([0,1], \mathbb{R}^d) : h'(0) = 0\}$ has the following properties:

(1) The canonical injection $\mathcal{X} \hookrightarrow \mathcal{W}_+$ has dense image in $\mathcal{W}_+$.

(2) The set $\mathcal{X}$ belongs to the Cameron–Martin space of $\tilde{\mathbb{P}}_+$, viewed as a measure on $\mathcal{C}([0,1], \mathbb{R}^d)$.

(3) The set $\{\hat{\mathcal{G}} h : h \in \mathcal{X}\}$, where $\hat{\mathcal{G}}$ is defined as

$$\hat{\mathcal{G}} h(t) = \gamma_H \int_0^1 \frac{1}{r} g\left(\frac{t}{r}\right)\bigl(h(1-r) - h(0)\bigr)\,dr - \gamma_H h(0) \int_1^\infty \frac{1}{r} g\left(\frac{t}{r}\right)dr,$$

belongs to the Cameron–Martin space of $\tilde{\mathbb{P}}_+$, viewed as a measure on $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$. Indeed, because of the representation (4.2) and the properties of fractional integrals, it suffices to check that $\mathcal{D}^{H+1/2}\hat{\mathcal{G}} h \in L^2(\mathbb{R}_+, \mathbb{R}^d)$ for $h \in \mathcal{X}$. Using (4.9), an explicit calculation shows that $\mathcal{D}^{H+1/2}\hat{\mathcal{G}} h \sim t^{1/2-H}$ for $t \ll 1$ and $\mathcal{D}^{H+1/2}\hat{\mathcal{G}} h \sim t^{-1/2-H}$ for $t \gg 1$ (see also [18], Lemma 4.3), so that this is indeed the case.

Indeed, the first property ensures that, given any two open sets $U, V \in \mathcal{W}_+$, we can find two smaller open sets $\bar{U}, \bar{V} \in \mathcal{W}_+$, such that $\bar{U} \subset U$, $\bar{V} \subset V$ and $\bar{V} = \tau_h \bar{U}$, with $h \in \mathcal{X}$. Furthermore, $\mathbb{P}_+(\bar{U}) > 0$, and so is $\mathbb{P}_+(\bar{V})$ since the topological support of $\mathbb{P}_+$ is all of $\mathcal{W}_+$ by [16], Theorem 15.63.

Since $\mathcal{X}$ belongs to the Cameron–Martin space of $\mathbb{P}_+$, this guarantees that, for every $w_- \in \mathcal{W}_-$, we can construct a subcoupling $\hat{\mathcal{P}}_{U,V}$ on $\mathcal{W}_+^2$ between $\hat{\mathcal{P}}(w_-, \cdot)|_{\bar{U}}$ and $\hat{\mathcal{P}}(w_-, \cdot)|_{\bar{V}}$ such that $\hat{\mathcal{P}}_{U,V}$ charges the set of pairs $(w_+, \bar{w}_+)$ such that $\bar{w}_+ = \tau_h w_+$. In order to check the quasi-Markov property, it now suffices to check that the measures $\bar{\mathcal{Q}}(x, w; \cdot)$ and $\bar{\mathcal{Q}}(x, \bar{w}; \cdot)$ are mutually equivalent if $(w, \bar{w})$ are such that their components in $\mathcal{W}_-$ are identical, and their components in $\mathcal{W}_+$ satisfy $\bar{w}_+ = \tau_h w_+$. This in turn is precisely the content of the third property above. □

The aim of the next section is to show that (1.1) does indeed possess the strong Feller property, provided that the vector fields $\{V_i\}$ satisfy Hörmander's bracket condition.

6.2. *Verification of the strong Feller property.* The main result of this section is that the strong Feller property is a consequence of Hörmander's bracket condition.

THEOREM 6.2. *Under Assumptions* 1 *and* 2, (1.1) *is strong Feller in the sense of Definition* 4.

REMARK 16. The main feature that distinguishes this situation from the usual one is that the process is not Markov. As a consequence, our definition of the strong Feller property implies that, as in [22], we need to construct a coupling between solutions starting from nearby points such that, with high probability, solutions agree not only after some fixed time (say 1), but also for all subsequent times. Furthermore, we will circumvent the fact that we do not assume a priori that the Jacobian of our solution has moments. This will be done by a cutoff procedure similar to [22].

PROOF OF THEOREM 6.2. Fix some arbitrary value $N > 1$ and a Fréchet differentiable map $\psi : \mathcal{X}^N \to \mathbb{R}$, which is bounded with bounded derivative. Denote furthermore by $R_N : \mathcal{X}^{\mathbf{N}} \to \mathcal{X}^N$ the projection onto the first N components, and set as before

$$\bar{\mathcal{Q}}\psi(z, w) \stackrel{\text{def}}{=} \int_{\mathcal{X}^{\mathbf{N}}} \psi(R_N x)\bar{\mathcal{Q}}(z, w; dx),$$

so that $\bar{\mathcal{Q}}\psi : \mathcal{X} \times \mathcal{W} \to \mathbb{R}$.

The strong Feller property will follow if we show the existence of a jointly continuous function $\ell : \mathcal{X}^2 \times \mathcal{W} \mapsto \mathbb{R}_+$ such that

$$\big|\bar{\mathcal{Q}}\psi(z, w) - \bar{\mathcal{Q}}\psi(y, w)\big| \le \ell(z, y, w) \tag{6.4}$$

for all Fréchet differentiable functions ψ with bounded derivatives such that

$$\sup_{x \in \mathcal{X}^N} \big|\psi(x)\big| \le 1,$$

uniformly for all $N > 1$.

To this end, set $z_s = zs + y(1 - s)$ for $s \in [0, 1]$ and $\xi = z - y$. Let $\Phi_{[1,T]}(z, w_+)$ denote the solution to (1.1), restricted to the interval $[1, T]$. Since

$$\mathcal{Q}\psi(z, w) = \tilde{\mathbb{E}}\psi_T\big(\Phi_{[1,T]}(z, w_+)\big),$$

where ψ_T is just ψ, composed with the evaluation map at integer times, we have the identity

$$\bar{\mathcal{Q}}\psi(z, w) - \bar{\mathcal{Q}}\psi(y, w) = \tilde{\mathbb{E}} \int_0^1 D\psi_T\big(\Phi_{[1,T]}(z_s, w_+)\big) J_{0,\cdot}^s \xi \, ds. \tag{6.5}$$

Here, $J_{0,\cdot}^s$ denotes the linearization of (1.1) with the initial condition z_s.

If moment bounds for the Jacobian $J^s_{0,t}$ are available, as in [23], we can proceed via a stochastic control argument using a Bismut–Elworthy–Li type formula [13] to show that that $\mathcal{Q}\psi(z,w)$ is actually differentiable in z. Since we do not assume this, we will combine this with a cutoff argument adapted from [22].

Recall the function $\Lambda_{\beta,q}(X,\mathbb{X}) \stackrel{\text{def}}{=} \Lambda_{\beta,q,1}(X,\mathbb{X})$ from (5.19) with $\beta > \gamma$, and set q to be an even integer such that (5.20) holds. Similar to (5.21), define the cutoff function

$$\Psi_R(w_+) \stackrel{\text{def}}{=} \chi\left(\frac{1}{R}\Lambda_{\beta,q}\big(X(w_+),\mathbb{X}(w_+)\big)\right), \qquad R > 0, \tag{6.6}$$

where $\chi:\mathbb{R}_+ \to \mathbb{R}_+$ is a $\mathcal{C}^\infty$ decreasing function with $\chi(\lambda) = 1$ for $\lambda \le 1$ and $\chi(\lambda) = 0$ for $\lambda \ge 2$.

From (6.5) we obtain that

$$\begin{aligned}\big|\bar{\mathcal{Q}}\psi_T(z,w) - \bar{\mathcal{Q}}\psi_T(y,w)\big| &\le \left|\tilde{\mathbb{E}}\int_0^1 \Psi_R(w_+) D\psi_T\big(\Phi_{[1,T]}(z_s,w_+)\big) J^s_{0,\cdot}\xi\,ds\right| \\ &\quad + \big|\tilde{\mathbb{E}}\big(1-\Psi_R(w_+)\big)\psi_T\big(\Phi_{[1,T]}(z,w_+)\big)\big| \\ &\quad + \big|\tilde{\mathbb{E}}\big(1-\Psi_R(w_+)\big)\psi_T\big(\Phi_{[1,T]}(y,w_+)\big)\big| \\ &\stackrel{\text{def}}{=} T_1 + T_2 + T_3.\end{aligned}$$

Since ψ_T is bounded by 1, we have the bound

$$T_2 + T_3 \le 2\mathcal{P}\big(w, \{w_+ | \Lambda_{\beta,q}\big(X(w_+),\mathbb{X}(w_+)\big) > 2R\}\big), \tag{6.7}$$

which can be made arbitrarily small by choosing R sufficiently large.

For tackling the term T_1, we now outline the stochastic control argument. Recall the operator $\mathcal{A}$ from (4.12). As explained in (4.17), the Fréchet derivative of the flow map with respect to the driving noise w_+ in the direction of $\int_0^\cdot v(s)\,ds$ is given by

$$J^s_{0,T}\mathcal{A}_T v.$$

The key idea underlying Bismut-type formulas is to use relation (4.17) to convert the derivative of $\Phi_T(z_s,w_+)$ with respect to its initial condition z_s into a derivative with respect to the driving noise and to use the integration by parts formula from Malliavin calculus.

To this end, given an initial displacement $\xi \in \mathbb{R}^n$, we seek for a "control" v on the time interval $[0,1]$ that solves the equation $\mathcal{A}_1 v = \xi$. If this can be achieved, then we extend v to all of $\mathbb{R}_+$ by setting $v(s) = 0$ for $s \ge 1$ and define $\tilde{v} = \mathcal{I}^{1/2-H} v$. Note that since $v(s) = 0$ for $s > 1$, it follows from the definition of $\mathcal{A}_T$ that we have $\mathcal{A}_T v = \mathcal{A}_1 v = \xi$ for $T \ge 1$. If v is sufficiently regular in time so that $\tilde{v} \in L^2(\mathbb{R}_+,\mathbb{R}^d)$, we have the identity

$$\mathbf{D}_{\tilde{v}} Z^{z_s}_T = J^s_{0,T}\xi \tag{6.8}$$

for every $T \geq 1$ and therefore, by the chain rule,

$$D\psi_T(\Phi_{[1,T]})J_{0,\cdot}\xi = D\psi_T(\Phi_{[1,T]})\mathbf{D}_{\tilde{v}}Z_{\cdot}^{z_s} = \mathbf{D}_{\tilde{v}}\big(\psi_T(\Phi_{[1,T]})\big). \tag{6.9}$$

It remains to find a control v which solves $\mathcal{A}_1 v = \xi$. From Proposition 3, $C_1 = \mathcal{A}_1\mathcal{A}_1^*$ is invertible, and therefore one possible solution to the equation $\mathcal{A}_1 v = \xi$ is given by the "least squares" formula,

$$v(r) \stackrel{\text{def}}{=} \mathcal{A}_1^*\big((\mathcal{A}_1\mathcal{A}_1^*)^{-1}\xi\big)(r) = V(Z_r)^*\big(J_{0,r}^{-1}\big)^* C_1^{-1}\xi, \qquad r \in (0,1). \tag{6.10}$$

Note that the control v also depends on the initial condition z_s, but since $|z_s| \leq |z| \vee |y|$, all our estimates for the rest of the proof will be uniform in the initial condition by Remark 8.

Inserting identity (6.9) into the definition of T_1, we obtain

$$|T_1| = \left|\int_0^1 \tilde{\mathbb{E}}\big(\Psi_R(w_+)\mathbf{D}_{\tilde{v}}\psi_T\big(\Phi_{[1,T]}(z_s, w_+)\big)\big)\,ds\right|. \tag{6.11}$$

Applying the integration by parts formula from Malliavin calculus, we obtain

$$\begin{aligned}\tilde{\mathbb{E}}\big(\Psi_R(w_+)\mathbf{D}_{\tilde{v}}\psi_T(\Phi_{[1,T]})\big) &= \tilde{\mathbb{E}}\big(\psi_T(\Phi_{[1,T]})\mathbf{D}^*\big(\Psi_R(w_+)\tilde{v}\big)\big)\\ &\leq \big(\tilde{\mathbb{E}}|\mathbf{D}^*\big(\Psi_R(w_+)\tilde{v}\big)|^2\big)^{1/2},\end{aligned} \tag{6.12}$$

where the second inequality follows from the fact that ψ_T is bounded by 1. To conclude the proof, it thus suffices to show that

$$\tilde{\mathbb{E}}\big(\big|\mathbf{D}^*(\Psi_R\tilde{v})\big|^2\big) \leq C(R, w_-, z)|\xi|^2, \tag{6.13}$$

where C is uniformly bounded on $\|w_-\|_\gamma \leq M$ and $|z| \leq M$.

Since the stochastic process $\tilde{v}$ is in general not adapted to the filtration generated by the underlying Wiener process, we use the following extension of Itô's isometry [39]:

$$\begin{aligned}\tilde{\mathbb{E}}\big(\big|\mathbf{D}^*(\Psi_R\tilde{v})\big|^2\big) &= \tilde{\mathbb{E}}\left(\int_0^\infty \big|\Psi_R\tilde{v}(s)\big|^2\,ds\right)\\ &\qquad + \tilde{\mathbb{E}}\int_0^\infty\int_0^\infty \operatorname{tr}\big(\mathbf{D}_t\big(\Psi_R\tilde{v}(s)\big)^T\mathbf{D}_s\big(\Psi_R\tilde{v}(t)\big)\big)\,ds\,dt\\ &\leq \tilde{\mathbb{E}}\|\Psi_R\tilde{v}\|^2 + \tilde{\mathbb{E}}\big\|\mathbf{D}(\Psi_R\tilde{v})\big\|^2 \leq c\big(\tilde{\mathbb{E}}\|\Psi_R v\|^2 + \tilde{\mathbb{E}}\|\mathbf{D}\Psi_R v\|^2\big)\\ &\stackrel{\text{def}}{=} I_1 + I_2.\end{aligned} \tag{6.14}$$

Here, $\|v\|$ denotes the L^2-norm of v and similarly for $\mathbf{D}v$. Since $\tilde{v} = \mathcal{I}^{H-1/2}v$, the second inequality is a consequence of the fact that $\mathcal{I}^{H-1/2}$ is a bounded operator from $L^2([0,1])$ into $L^2(\mathbb{R}_+)$; see Corollary 3 below.

Bound (6.13) on I_1 now follows immediately from Proposition 3 and Remark 8. The bound on I_2 follows similarly by also using Theorem 4.1. The proof of Theorem 6.2 is complete. □

We now show that $\mathcal{I}^{1/2-H}$ is indeed a bounded operator from $L^2([0,1])$ to $L^2(\mathbb{R}_+)$. For this, define the operator $\tilde{\mathcal{I}}_\alpha$ by

$$(\tilde{\mathcal{I}}_\alpha v)(s) = \int_0^1 |s-r|^{\alpha-1} v(r)\,dr.$$

We then have:

LEMMA 7. *For $\alpha \in (0, \frac{1}{2})$, there exists a constant c such that, for positive v,*

$$\|\tilde{\mathcal{I}}_\alpha v\|^2 \le c\|v\|\,\|\tilde{\mathcal{I}}_{2\alpha} v\|.$$

PROOF. We have the bound

$$\begin{aligned}\|\tilde{\mathcal{I}}_\alpha v\|^2 &= \int_0^1\int_0^1\int_0^\infty |s-r|^{\alpha-1}|s-t|^{\alpha-1}\,ds\,v(r)v(t)\,dr\,dt \\ &\le \int_0^1\int_0^1\int_{-\infty}^\infty |s-r|^{\alpha-1}|s-t|^{\alpha-1}\,ds\,v(r)v(t)\,dr\,dt \\ &= c\int_0^1\int_0^1 |r-t|^{2\alpha-1}\,ds\,v(r)v(t)\,dr\,dt,\end{aligned}$$

where the first step follows from the positivity of v, and the second step follows from a simple scaling argument. Since this is nothing but $c\langle v, \tilde{\mathcal{I}}_{2\alpha} v\rangle$, the requested bound follows from the Cauchy–Schwarz inequality. □

COROLLARY 3. *For every $\alpha \in (0,1)$, the operator $\mathcal{I}^\alpha$ is bounded from $L^2([0,1])$ to $L^2(\mathbb{R}_+)$.*

PROOF. Note that

$$|\mathcal{I}^\alpha v(s)| \le \mathcal{I}^\alpha |v|(s) \le \tilde{\mathcal{I}}^\alpha |v|(s).$$

Since $|s-r|^{\alpha-1}$ is square integrable if $\alpha > \frac{1}{2}$, the claim follows for that range of α. For smaller values of α, it is always possible to reduce oneself to the range $(\frac{1}{2}, 1)$ by Lemma 7, noting also that $\|\tilde{\mathcal{I}}_\alpha v\| \le c\|\tilde{\mathcal{I}}_\beta v\|$ for $\alpha > \beta$. □

7. Examples. In this section, we collect a few examples to which our main results apply.

7.1. *Hypoelliptic Ornstein–Uhlenbeck process.* Consider the process x_t given by

$$dx = Ax\,dt + C\,dB_H(t), \tag{7.1}$$

where $x \in \mathbb{R}^n$, B_H is an m-dimensional fractional Brownian motion with Hurst parameter $H > \frac{1}{3}$, A is an $n \times n$ matrix with $A + A^T < 0$ and C is an $n \times m$

matrix. It is well known that (7.1) satisfies Hörmander's condition if and only if there exists $k > 0$ such that the matrix $(C, AC, \dots, A^k C)$ has rank n.

Since the Jacobian is given by $J_{s,t} = \exp(A(t-s))$ and therefore has moments of all orders, we conclude that, for any initial condition x_0 and for any sequence of times $t_1, \dots, t_k$, the joint distribution of $(x_{t_1}, \dots, x_{t_k})$ has a smooth density with respect to Lebesgue measure. Since this distribution is Gaussian, one could have verified directly that its covariance is nondegenerate, but this would have been a rather lengthy calculation.

7.2. *Linear equations/Lévy area.* Let B be a d-dimensional fractional Brownian motion and consider equations in $\mathbb{R}^m$ of the type

$$dX_i = (A_{ijk} X_j + C_{ik}) \circ dB_k(t), \tag{7.2}$$

where we use Einstein's convention of summation over repeated indices. In this case, the derivative of the solution with respect to its initial condition in a direction $\eta \in \mathbb{R}^m$ is nothing but the solution to

$$dJ_i = A_{ijk} J_j \circ dB_k(t), \tag{7.3}$$

with initial condition $J(0) = \eta$. Similar formulas hold for higher order derivatives, so that it follows from the results recently obtained in [14] that Assumption 3 is satisfied and our result on the smoothness of the densities applies, provided that Hörmander's condition holds.

As an immediate consequence, we have the smoothness of the Lévy area, which was recently obtained independently in [10]:

PROPOSITION 5. *Let B be a d-dimensional fractional Brownian motion with Hurst parameter $H > \frac{1}{3}$, and let $W_{ij}(t) = \int_0^t B_i(s) \circ dB_j(s) - \int_0^t B_j(s) \circ dB_i(s)$ for $i < j$. Then, for any fixed $t > 0$, the vector $(B_k(t), W_{ij}(t))$ with $k = 1, \dots, d$ and $i < j$ has a smooth density with respect to Lebesgue measure.*

PROOF. The verification of Hörmander's condition boils down to a simple problem in linear algebra. Writing e_j for the basis vector in the direction B_j and f_{ij} for the basis vector in the direction W_{ij}, we can rewrite $x = B \oplus W$ as the solution to the SDE

$$dx = \sum_j \biggl(e_j + \sum_{i<j} f_{ij} \langle x, e_i \rangle - \sum_{i>j} f_{ji} \langle x, e_i \rangle \biggr) \circ dB_j = \sum_j V_j(x) \circ dB_j.$$

An explicit calculation shows that, for $j < k$, we have

$$[V_j, V_k](x) = 2 f_{jk},$$

so that Hörmander's condition holds after one step. □

REMARK 17. Higher order totally antisymmetric iterated integrals can be treated in exactly the same way with the kth iterated Lie brackets recovering precisely the basis vectors of the elements in the kth antisymmetric tensor.

7.3. *Simplified fractional Langevin equation.* Consider the process (q_t, p_t) on $\mathbb{R}^{2n}$ given by

$$dq = p\,dt, \qquad dp = -\nabla V(q)\,dt - p\,dt + dB_H(t), \tag{7.4}$$

where, for the sake of simplicity, we assume that $V : \mathbb{R}^n \to \mathbb{R}_+$ has bounded second derivative, and there exist $C > 0$ and $\kappa > 0$ such that

$$\langle q, \nabla V(q)\rangle \geq \kappa |q|^2 - C, \qquad V(q) \geq \kappa |q|^2 - C. \tag{7.5}$$

This equation is a simplified version of the fractional Langevin equation. (The equation satisfying the correct physical detailed balance condition would have a more complicated memory kernel instead of the simple friction term $-p\,dt$ appearing above.)

Because we assume V to have a bounded second derivative, the Jacobian of (7.4) is bounded by a deterministic constant over any finite time interval. Furthermore, Hörmander's condition is easy to verify, so that we can apply Theorem 5.2 to infer the existence of smooth densities for the joint distribution of the solution at any time.

Regarding the existence of a unique invariant measure for (7.1), it only remains to obtain a Lyapunov function for the solution to (7.4). For this, similar to [18], we proceed as follows. We consider the process $(\tilde{p}, \tilde{q})$ solution to

$$d\tilde{q} = -\tilde{q}\,dt, \qquad d\tilde{p} = -\tilde{p}\,dt + dB_H(t).$$

It is, of course, trivial to bound solutions to this equation. Then we set $P = p - \tilde{p}$ and $Q = q - \tilde{q}$. The equation for (P, Q) can be written as

$$\dot{Q} = P + R_Q, \qquad \dot{P} = -\nabla V(Q) - P + R_P,$$

where

$$R_Q = \tilde{p} - \tilde{q}, \qquad R_P = \nabla V(Q) - \nabla V(Q + \tilde{q}).$$

Note that since we assumed that V has bounded second derivative, both R_P and R_Q are bounded by a multiple of $|\tilde{p}| + |\tilde{q}|$, independently of P and Q. We now set $\bar{H}(P, Q) = \frac{1}{2}P^2 + V(Q) + \gamma P Q$ for a constant γ to be determined later. An explicit calculation yields the bound

$$\begin{aligned}\frac{d}{dt}\bar{H}(P, Q) = {}& -(1-\gamma)|P|^2 - \gamma\langle Q, \nabla V(Q)\rangle + \langle \nabla V(Q) + \gamma P, R_Q\rangle \\ & + \langle P + \gamma Q, R_P\rangle.\end{aligned}$$

Making use of (7.5) and the bounds on R_Q and R_P, we see that there exists constant $\alpha > 0$ and $C > 0$ such that

$$\frac{d}{dt}\bar{H}(P, Q) \leq -\alpha \bar{H}(P, Q) + C\bigl(1 + |\tilde{p}|^2 + |\tilde{q}|^2\bigr).$$

Since, by (7.5), $\bar{H}$ grows quadratically at infinity for γ small enough, it follows in the same way as in [18], Proposition 3.12, that $|p|^2 + |q|^2$ is a Lyapunov function for (7.4). We therefore have:

THEOREM 7.1. *If V has bounded second derivative and* (7.5) *holds*, *then there exists a unique invariant measure for* (7.4).

PROOF. The existence of an invariant measure follows from the fact that $|p|^2+|q|^2$ is a Lyapunov function. The uniqueness then follows from Theorem 6.1. □

REMARK 18. Our results also apply to more degenerate situations. For example, if we consider the fractional Langevin equation associated to systems of anharmonic oscillators in contact with thermal baths at their boundary, as studied in [11, 12], our results imply the uniqueness of a steady state. Existence of a steady state, however, is a much harder problem in such systems, which is partially unsolved even in the Markovian case.

REMARK 19. The noise in the above example is additive, and thus it might seem that we are not using the rough path nature of the fBm here. This is true in this particular case, but in more complicated situations with additive noise, like, for example, the one in [11], both rough path analysis and our version of Norris's lemma are still needed in order to analyze expressions such as (5.8), when U is given by a higher-order Lie bracket.

APPENDIX: BOUNDS ON THE CUTOFF FUNCTION

In this section, we show that the function $\Lambda_{\beta,q}$ appearing in Sections 5.3 and 6.2 does indeed have the requested smoothness properties. Our main result is the following:

PROPOSITION 6. *Let $\Lambda_{\beta,q}$ be as in* (5.19), *and assume that β and q are such that* (5.20) *holds and such that $2\beta \le \gamma + H$*. [*This is always possible by first setting $\beta = (\gamma + H)/2$ and then choosing q large enough*.]

Then, *for every $k > 0$ and every $R > 0$*, *there exists a constant M such that the bound*

$$\|\mathbf{D}^{(k)}\Lambda_{\beta,q}(X,\mathbb{X})\| \le M,$$

holds for all $(X,\mathbb{X})$ such that $\Lambda_{\beta,q}(X,\mathbb{X}) \le 2R$. Here, $\mathbf{D}^{(k)}$ *denotes the kth iterated Malliavin derivative*, *and $\|\cdot\|$ is the L^2-norm on $[0,T]^k$.*

PROOF. Note first that, by definition,

$$\mathrm{D}_r^i \delta X_{s,t}^j = \delta_{ij}\mathbf{1}_{r\in[s,t]},$$

where δ_{ij} is the Kronecker delta. It thus follows from (4.21) that

$$\mathbf{D}_r^i \delta X_{s,t}^j = c\delta_{ij}\big((t-r)^{H-1/2}\mathbf{1}_{r<t} - (s-r)^{H-1/2}\mathbf{1}_{r<s}\big) \stackrel{\text{def}}{=} \delta_{ij} f_{s,t}(r). \tag{A.1}$$

The L^2-norm of $f_{s,t}$ is given by

$$\|f_{s,t}\|^2 = c^2 \int_s^t (t-r)^{2H-1}\,dr + c^2 \int_0^s \bigl((t-r)^{2H-1} - (s-r)^{2H-1}\bigr)\,dr.$$

Since $H < \frac{1}{2}$ by assumption, one has the inequality

$$\bigl|t^{2H} - s^{2H}\bigr| \le |t-s|^{2H},$$

so that a straightforward calculation yields the bound

$$\|f_{s,t}\| \le \kappa |t-s|^H$$

for some constant $\kappa > 0$.

Concerning $\tilde{\mathbb{X}}^{k\ell}_{s,t} = \mathbb{X}^{k\ell}_{s,t} - \mathbb{X}^{\ell k}_{s,t}$, an explicit calculation yields the identity

$$\mathrm{D}^i_r \tilde{\mathbb{X}}^{k\ell}_{s,t} = \mathbf{1}_{r\in[s,t]}\bigl(\delta_{ik}\bigl(\delta X^\ell_{r,t} - \delta X^\ell_{s,r}\bigr) - \delta_{i\ell}\bigl(\delta X^k_{r,t} - \delta X^k_{s,r}\bigr)\bigr).$$

Applying again (4.21), we obtain

$$\mathbf{D}^j_r \mathbb{X}^{k\ell}_{s,t} = \delta_{ik} G^\ell_{s,t}(r) + \delta_{i\ell} G^k_{s,t}(r),$$

where we set

$$\begin{aligned} G^k_{s,t}(r) = 2\mathbf{1}_{r\in[s,t]} &\int_r^t (u-r)^{H-3/2} \delta X^k_{r,u}\,du \\ &+ \mathbf{1}_{r\in[s,t]}\bigl(\delta X^k_{r,t} - \delta X^k_{s,r}\bigr) \int_t^\infty (u-r)^{H-3/2}\,du \\ &+ \mathbf{1}_{r<s} \int_s^t (u-r)^{H-3/2}\bigl(\delta X^k_{s,u} - \delta X^k_{u,t}\bigr)\,du. \end{aligned}$$

The important fact about $G^k_{s,t}(r)$ is that we can estimate it by

$$\begin{aligned} \bigl|G^k_{s,t}(r)\bigr| \le M\|X\|_\gamma &\mathbf{1}_{r\in[s,t]}(t-r)^{\gamma+H-1/2} \\ &+ M\|X\|_\gamma \mathbf{1}_{r<s}(t-s)^\gamma\bigl((t-r)^{H-1/2} - (s-r)^{H-1/2}\bigr), \end{aligned}$$

so that its L^2-norm is controlled by

$$\|G^k_{s,t}\| \le M\|X\|_\gamma |t-s|^{H+\gamma}. \tag{A.2}$$

We finally compute the second Malliavin derivative of $\mathbb{X}^{k\ell}_{s,t}$. It follows in a rather straightforward way from (4.21) that, for $r_1 \in (s,t)$ and $v \in (r_1, t)$, one has

$$\mathrm{D}^i_v \mathbf{D}^j_{r_1} \mathbb{X}^{k\ell}_{s,t} = (\delta_{ik}\delta_{j\ell} + \delta_{i\ell}\delta_{jk})\biggl(2\int_v^t (u-r_1)^{H-3/2}\,du + \int_t^\infty (u-r_1)^{H-3/2}\,du\biggr)$$

for $r_1 < s$ and $v \in (s,t)$, one has

$$\mathrm{D}^i_v \mathbf{D}^j_{r_1} \mathbb{X}^{k\ell}_{s,t} = (\delta_{ik}\delta_{j\ell} + \delta_{i\ell}\delta_{jk})\biggl(2\int_v^t (u-r_1)^{H-3/2}\,du - \int_s^t (u-r_1)^{H-3/2}\,du\biggr),$$

and for all other combinations with $v > r_1$ one has $\mathrm{D}_v^i \mathbf{D}_{r_1}^j \mathbb{X}_{s,t}^{k\ell} = 0$.

A lengthy but straightforward calculation then yields

$$\mathbf{D}_{r_2}^i \mathbf{D}_{r_1}^j \mathbb{X}_{s,t}^{k\ell} = (\delta_{ik}\delta_{j\ell} + \delta_{i\ell}\delta_{jk}) g_{s,t}(r_1, r_2), \tag{A.3}$$

where, for $s < r_1 < r_2 < t$, the function $g_{s,t}$ is given by

$$\begin{aligned} g_{s,t}(r_1, r_2) &= 2\int_{r_2}^{t} (v - r_2)^{H-3/2} \int_{r_2}^{v} (u - r_1)^{H-3/2}\, du\, dv \\ &\quad + c_g (t - r_2)^{H-1/2} \bigl(2(r_2 - r_1)^{H-1/2} - (t - r_1)^{H-1/2}\bigr) \\ &\stackrel{\text{def}}{=} g_{s,t}^{(1)}(r_1, r_2) + g_{s,t}^{(2)}(r_1, r_2) \end{aligned}$$

for some constant c_g. For $r_1 < s < r_2 < t$ on the other hand, one has

$$\begin{aligned} &g_{s,t}(r_1, r_2) \\ &\quad = 2\int_{r_2}^{t} (v - r_2)^{H-3/2} \int_{r_2}^{v} (u - r_1)^{H-3/2}\, du\, dv \\ &\qquad + c_g (t - r_2)^{H-1/2} \bigl(2(r_2 - r_1)^{H-1/2} - (s - r_1)^{H-1/2} - (t - r_1)^{H-1/2}\bigr) \\ &\quad \stackrel{\text{def}}{=} g_{s,t}^{(3)}(r_1, r_2) + g_{s,t}^{(4)}(r_1, r_2). \end{aligned}$$

Finally, for $r_1 < r_2 < s < t$, one has

$$g_{s,t}(r_1, r_2) = 2\int_{s}^{t} (v - r_2)^{H-3/2} \int_{r_2}^{v} (u - r_1)^{H-3/2}\, du\, dv \stackrel{\text{def}}{=} g_{s,t}^{(5)}(r_1, r_2).$$

It is possible to check that

$$\|g_{s,t}\| \le M|t - s|^{2H}, \tag{A.4}$$

where $\|\cdot\|$ denotes again the L^2-norm. (We postpone the proof of this to Lemma 8 below.)

We now have all the necessary tools to conclude. Write

$$\Lambda_{\beta,q}(X, \mathbb{X}) = \Lambda_{\beta,q}^{(1)}(X) + \Lambda_{\beta,q}^{(2)}(\mathbb{X}),$$

where $\Lambda^{(1)}$ is as (5.19), but keeping only the term proportional to $|\delta X_{s,t}|^{2q}$ in the integral, and similarly for $\Lambda^{(2)}$. It follows from (A.1) that, for $\ell \le 2q$, the multiple Malliavin derivative of $\Lambda_{\beta,q}^{(1)}$ satisfies the bound

$$\bigl|\mathbf{D}_{s_1} \cdots \mathbf{D}_{s_\ell} \Lambda_{\beta,q}^{(1)}(X)\bigr| \le M \int_0^T \int_0^t \frac{|\delta X_{s,t}|^{2q-\ell}}{|t - s|^{\beta(2q-\ell)}} \prod_{j=1}^{\ell} \frac{|f_{s,t}(r_j)|}{|t - s|^{\beta}}\, ds\, dt.$$

Since the L^2-norm of $f_{s,t}$ is bounded by $M|t - s|^H$, it follows immediately that there exits a constant M such that the L^2-norm of $\mathbf{D}_{s_1} \cdots \mathbf{D}_{s_\ell} \Lambda_{\beta,q}^{(1)}$ is bounded by $M\Lambda^{(1)}_{\beta,q}$. Its Malliavin derivative of order $\ell > 2q$ on the other hand vanishes identically.

Similarly, we only need to consider Malliavin derivatives of order $\ell \le 2q$ for $\Lambda^{(2)}_{\beta,q}$. A reasoning similar to the above shows that its Malliavin derivative can be written as

$$\mathbf{D}_{s_1}\cdots\mathbf{D}_{s_\ell}\Lambda^{(2)}_{\beta,q} = \int_0^T\int_0^t \frac{\mathrm{P}(\mathbb{X}_{s,t}, \mathbf{D}_{s.}\mathbb{X}_{s,t}, g_{s,t}(s., s.))}{|t-s|^{2\beta q}}\,ds\,dt,$$

where P is a homogeneous polynomial of degree q and "$s.$" is a generic placeholder for any of the times $s_1,\dots,s_\ell$. It now follows from (A.2), (A.4) and the assumption $2\beta \le \gamma + H$ that the L^2-norm of $\mathbf{D}_{s_1}\cdots\mathbf{D}_{s_\ell}\Lambda^{(2)}_{\beta,q}$ is bounded by $M(\Lambda^{(2)}_{\beta,q} + \|X\|^q_\gamma)$. Since on the other hand, $\|X\|^q_\gamma$ is bounded by $M\Lambda_{\beta,q}$, by assumption, this completes the proof. $\square$

COROLLARY 4. *Let $\Psi_R(w_+)$ be as defined in (6.6) with $\Lambda_{\beta,q}$ as in (5.19), and let β and q be as in Proposition 6.*

Then, for every $R > 0$, $\Psi_R \in \mathbf{D}^\infty$. Furthermore, every multiple Malliavin derivative of Ψ_R vanishes outside of the set $\{\Lambda_{\beta,q}(X,\mathbb{X}) \le 2R\}$.

PROOF. By the chain rule,

$$\mathbf{D}_s\Psi_R(w_+) = \frac{1}{R}\chi'\big(R^{-1}\Lambda_{\beta,q}(X,\mathbb{X})\big)\mathbf{D}_s\Lambda_{\beta,q}(X,\mathbb{X}),$$

and similarly for higher order derivatives. Since all derivatives of χ vanish when the argument is larger than 2, the claim follows from Proposition 6. $\square$

LEMMA 8. *For every $T > 0$, there exists a constant M such that the function $g_{s,t}$ from (A.3) satisfies $\|g_{s,t}\| \le M|t-s|^{2H}$.*

PROOF. We show the bound separately for $g^{(j)}_{s,t}$ with $j = 1,\dots,5$. For $g^{(1)}_{s,t}$, we use the bound

$$(u-r_1)^{H-3/2} \le (u-r_1)^{H-3/2}(r_2-r_1)^{-\beta},$$

in order to conclude that, provided that $1-2H < \beta$, one has the pointwise bound

$$\big|g^{(1)}_{s,t}(r_1,r_2)\big| \le (t-r_2)^{2H-1+\beta}(r_2-r_1)^{-\beta} \le |t-s|^{2H-1+\beta}(r_2-r_1)^{-\beta}.$$

If furthermore $\beta < \frac{1}{2}$ (which is always possible if $H > \frac{1}{4}$), the integral of $(r_2-r_1)^{-2\beta}$ over $s < r_1 < r_2 < t$ is proportional to $|t-s|^{1-2\beta}$, thus yielding the required bound.

Similarly, $g^{(2)}_{s,t}$ satisfies

$$\big|g^{(2)}_{s,t}(r_1,r_2)\big| \le C(t-r_2)^{H-1/2}(r_2-r_1)^{H-1/2},$$

and a straightforward calculation shows that

$$\int_s^t (t-r_2)^{2H-1} \int_s^{r_2} (r_2-r_1)^{2H-1}\,dr_1\,dr_2 = C|t-s|^{4H}$$

for some constant C as required.

For $g^{(3)}_{s,t}$ we have, as for $g^{(1)}_{s,t}$,

$$\big|g^{(3)}_{s,t}(r_1,r_2)\big| \le (t-r_2)^{2H-1+\beta}(r_2-r_1)^{-\beta}.$$

This time, however, we choose $\beta \in (\frac{1}{2}, 1)$, so that

$$\int_s^t \int_0^s \big|g^{(3)}_{s,t}(r_1,r_2)\big|^2\,dr_1\,dr_2 \le M \int_s^t (t-r_2)^{4H-2+2\beta}(r_2-s)^{1-2\beta}\,dr_2,$$

which is indeed proportional to $|t-s|^{4H}$.

To bound $g^{(4)}_{s,t}$ we perform the change of variables $r_1 \mapsto s-r_1$ and $r_2 \mapsto r_2+s$, so that

$$\begin{aligned}\int_s^t \int_0^s &\big|g^{(4)}_{s,t}(r_1,r_2)\big|^2\,dr_1\,dr_2\\ &= M\int_0^{t-s}(t-s-r_2)^{2H-1}\\ &\qquad\times\int_0^s \big(2(r_2+r_1)^{H-1/2} - r_1^{H-1/2} - (t-s+r_1)^{H-1/2}\big)^2\,dr_1\,dr_2.\end{aligned}$$

We then dilate the expression by $t-s$, showing that it is proportional to

$$\begin{aligned}|t-s|^{4H}&\int_0^1\int_0^{s/(t-s)}(1-r_2)^{2H-1}\\ &\qquad\times\big(2(r_2+r_1)^{H-1/2} - r_1^{H-1/2} - (1+r_1)^{H-1/2}\big)^2\,dr_1\,dr_2.\end{aligned}$$

It is straightforward to check that this integral converges for all $H \in (0,1)$, which shows the requested bound on $g^{(4)}_{s,t}$.

Finally, to bound $g^{(5)}_{s,t}$, we perform the change of variables $r_1 \mapsto s-r_1$ and $r_2 \mapsto s-r_2$, followed by a dilatation of $t-s$, so that

$$\int_0^s\int_0^{r_2}\big|g^{(5)}_{s,t}(r_1,r_2)\big|^2\,dr_1\,dr_2 = |t-s|^{4H}\int_0^{s/(t-s)}\int_{r_2}^{s/(t-s)}\big|\tilde g^{(5)}(r_1,r_2)\big|^2\,dr_1\,dr_2,$$

where

$$\tilde g^{(5)}(r_1,r_2) = \int_{-1}^0 (r_2-v)^{H-3/2}\int_v^{r_2}(r_1-u)^{H-3/2}\,du\,dv.$$

Note now that, for every $\beta \in (0, \frac{3}{2}-H)$, there exists a constant M such that, for $r_1 > r_2$, one has the bound

$$\begin{aligned}\big|\tilde g^{(5)}&(r_1,r_2)\big|\\ &\le M(r_1-r_2)^{-\beta}\int_{-1}^0 (r_2-v)^{2H-2+\beta}\,dv \le M\frac{(r_1-r_2)^{-\beta} r_2^{2H-1+\beta}}{1+r_2}\,dv.\end{aligned}$$

Choosing $\beta \approx \frac{1}{2}$ (but slightly larger than $\frac{1}{2}$) for $r_1 > r_2 + 1$ and $\beta = 0$ for $r_1 \le r_2$, we can check that

$$\int_0^\infty \int_{r_2}^\infty \left|\tilde{g}^{(5)}(r_1, r_2)\right|^2 dr_1\, dr_2 < \infty,$$

so that the claim follows. □

Acknowledgments. We are grateful to Tom Cass, Peter Friz, Massimilliano Gubinelli, Terry Lyons and Samy Tindel for many interesting discussions on the subject treated in this work. We also thank an anonymous referee for detailed comments which helped us improve the exposition.

REFERENCES

[1] ANDERSON, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* **6** 170–176. MR0069229

[2] BALL, K. (1992). Ellipsoids of maximal volume in convex bodies. *Geom. Dedicata* **41** 241–250. MR1153987

[3] BAUDOIN, F. and HAIRER, M. (2007). A version of Hörmander's theorem for the fractional Brownian motion. *Probab. Theory Related Fields* **139** 373–395. MR2322701

[4] BISMUT, J.-M. (1981). Martingales, the Malliavin calculus and hypoellipticity under general Hörmander's conditions. *Z. Wahrsch. Verw. Gebiete* **56** 469–505. MR0621660

[5] BUTCHER, J. C. (1972). An algebraic theory of integration methods. *Math. Comp.* **26** 79–106. MR0305608

[6] CASS, T. and FRIZ, P. (2010). Densities for rough differential equations under Hörmander's condition. *Ann. of Math.* (2) **171** 2115–2141. MR2680405

[7] CASS, T., FRIZ, P. and VICTOIR, N. (2009). Non-degeneracy of Wiener functionals arising from rough differential equations. *Trans. Amer. Math. Soc.* **361** 3359–3371. MR2485431

[8] CASS, T., LITTERER, C. and LYONS, T. (2011). Integrability estimates for Gaussian rough differential equations. Available at arXiv:1104.1813.

[9] COUTIN, L. and QIAN, Z. (2002). Stochastic analysis, rough path analysis and fractional Brownian motions. *Probab. Theory Related Fields* **122** 108–140. MR1883719

[10] DRISCOLL, P. (2010). Smoothness of density for the area process of fractional Brownian motion. Available at arXiv:1010.3047.

[11] ECKMANN, J. P. and HAIRER, M. (2000). Non-equilibrium statistical mechanics of strongly anharmonic chains of oscillators. *Comm. Math. Phys.* **212** 105–164. MR1764365

[12] ECKMANN, J. P., PILLET, C. A. and REY-BELLET, L. (1999). Non-equilibrium statistical mechanics of anharmonic chains coupled to two heat baths at different temperatures. *Comm. Math. Phys.* **201** 657–697. MR1685893

[13] ELWORTHY, K. D. and LI, X. M. (1994). Formulae for the derivatives of heat semigroups. *J. Funct. Anal.* **125** 252–286. MR1297021

[14] FRIZ, P. and RIEDEL, S. (2011). Integrability of linear rough differential equations. Available at arXiv:1104.0577.

[15] FRIZ, P. and VICTOIR, N. (2010). Differential equations driven by Gaussian signals. *Ann. Inst. Henri Poincaré Probab. Stat.* **46** 369–413. MR2667703

[16] FRIZ, P. K. and VICTOIR, N. B. (2010). *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications. Cambridge Studies in Advanced Mathematics* **120**. Cambridge Univ. Press, Cambridge. MR2604669

[17] GUBINELLI, M. (2004). Controlling rough paths. *J. Funct. Anal.* **216** 86–140. MR2091358

[18] HAIRER, M. (2005). Ergodicity of stochastic differential equations driven by fractional Brownian motion. *Ann. Probab.* **33** 703–758. MR2123208

[19] HAIRER, M. (2009). Ergodic properties of a class of non-Markovian processes. In *Trends in Stochastic Analysis. London Mathematical Society Lecture Note Series* **353** 65–98. Cambridge Univ. Press, Cambridge. MR2562151

[20] HAIRER, M. (2011). Rough stochastic PDEs. *Comm. Pure Appl. Math.* **64** 1547–1585. MR2832168

[21] HAIRER, M. and MATTINGLY, J. C. (2011). A theory of hypoellipticity and unique ergodicity for semilinear stochastic PDEs. *Electron. J. Probab.* **16** 658–738. MR2786645

[22] HAIRER, M. and OHASHI, A. (2007). Ergodic theory for SDEs with extrinsic memory. *Ann. Probab.* **35** 1950–1977. MR2349580

[23] HAIRER, M. and PILLAI, N. S. (2011). Ergodicity of hypoelliptic SDEs driven by fractional Brownian motion. *Ann. Inst. Henri Poincaré Probab. Stat.* **47** 601–628. MR2814425

[24] HÖRMANDER, L. (1967). Hypoelliptic second order differential equations. *Acta Math.* **119** 147–171. MR0222474

[25] HU, Y. and TINDEL, S. (2011). Smooth density for some nilpotent rough differential equations. Available at arXiv:1104.1972.

[26] JOHN, F. (1948). Extremum problems with inequalities as subsidiary conditions. In *Studies and Essays Presented to R. Courant on His* 60*th Birthday*, *January* 8, 1948 187–204. Interscience, New York, NY. MR0030135

[27] KUELBS, J., LI, W. V. and LINDE, W. (1994). The Gaussian measure of shifted balls. *Probab. Theory Related Fields* **98** 143–162. MR1258983

[28] KUSUOKA, S. and STROOCK, D. (1984). Applications of the Malliavin calculus. I. In *Stochastic Analysis* (*Katata/Kyoto*, 1982). *North-Holland Math. Library* **32** 271–306. North-Holland, Amsterdam. MR0780762

[29] KUSUOKA, S. and STROOCK, D. (1985). Applications of the Malliavin calculus. II. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **32** 1–76. MR0783181

[30] KUSUOKA, S. and STROOCK, D. (1987). Applications of the Malliavin calculus. III. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **34** 391–442. MR0914028

[31] LI, W. V. and SHAO, Q. M. (2001). Gaussian processes: Inequalities, small ball probabilities and applications. In *Stochastic Processes: Theory and Methods. Handbook of Statist.* **19** 533–597. North-Holland, Amsterdam. MR1861734

[32] LYONS, T. and QIAN, Z. (2002). *System Control and Rough Paths*. Oxford Univ. Press, Oxford. MR2036784

[33] LYONS, T. J. (1998). Differential equations driven by rough signals. *Rev. Mat. Iberoam.* **14** 215–310. MR1654527

[34] LYONS, T. J., CARUANA, M. and LÉVY, T. (2007). *Differential Equations Driven by Rough Paths. Lecture Notes in Math.* **1908**. Springer, Berlin. MR2314753

[35] MALLIAVIN, P. (1978). Stochastic calculus of variations and hypoelliptic operators. In *Symp. on Stoch. Diff. Equations, Kyoto* 1976 147–171. Wiley, New York.

[36] MALLIAVIN, P. (1997). *Stochastic Analysis. Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **313**. Springer, Berlin. MR1450093

[37] MANDELBROT, B. B. and VAN NESS, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10** 422–437. MR0242239

[38] NORRIS, J. (1986). Simplified Malliavin calculus. In *Séminaire de Probabilités, XX*, 1984/85. *Lecture Notes in Math.* **1204** 101–130. Springer, Berlin. MR0942019

[39] NUALART, D. (2006). *The Malliavin Calculus and Related Topics*, 2nd ed. Springer, Berlin. MR2200233

[40] NUALART, D. and SAUSSEREAU, B. (2009). Malliavin calculus for stochastic differential equations driven by a fractional Brownian motion. *Stochastic Process*. *Appl*. **119** 391–409. MR2493996

[41] SAMKO, S. G., KILBAS, A. A. and MARICHEV, O. I. (1993). *Fractional Integrals and Derivatives*: *Theory and Applications*. Gordon and Breach, Yverdon. MR1347689

[42] STROOCK, D. W. and VARADHAN, S. R. S. (1972). On the support of diffusion processes with applications to the strong maximum principle. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (*Univ. California*, *Berkeley*, *Calif*., 1970/1971), *Vol*. *III*: *Probability Theory* 333–359. Univ. California Press, Berkeley, CA. MR0400425

MATHEMATICS INSTITUTE
UNIVERSITY OF WARWICK
COVENTRY, CV4 7AL
UNITED KINGDOM
E-MAIL: Martin.Hairer@Warwick.ac.uk

DEPARTMENT OF STATISTICS
HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS 02138
USA
E-MAIL: pillai@fas.harvard.edu