ANNALES
DE L'INSTITUT
HENRI
POINCARÉ

PROBABILITÉS
ET STATISTIQUES

Ergodicity of hypoelliptic SDEs driven by fractional Brownian motion

M. Hairer[a,b] and N. S. Pillai[c]

[a]Mathematics Institute, The University of Warwick, West Midlands, UK. E-mail: M.Hairer@Warwick.ac.uk
[b]Courant Institute, New York University, New York, NY, USA. E-mail: mhairer@cims.nyu.edu
[c]CRiSM, The University of Warwick, West Midlands, UK. E-mail: N.Pillai@Warwick.ac.uk

Abstract. We demonstrate that stochastic differential equations (SDEs) driven by fractional Brownian motion with Hurst parameter $H > \frac{1}{2}$ have similar ergodic properties as SDEs driven by standard Brownian motion. The focus in this article is on hypoelliptic systems satisfying Hörmander's condition. We show that such systems enjoy a suitable version of the strong Feller property and we conclude that under a standard controllability condition they admit a unique stationary solution that is physical in the sense that it does not "look into the future."

The main technical result required for the analysis is a bound on the moments of the inverse of the Malliavin covariance matrix, conditional on the past of the driving noise.

Résumé. Nous démontrons que les équations différentielles stochastiques (EDS) conduites par des mouvements Browniens fractionnaires à paramètre de Hurst $H > \frac{1}{2}$ ont des propriétés ergodiques similaires aux EDS usuelles conduites par des mouvements Browniens. L'intérêt principal du présent article est de pouvoir traiter également des systèmes hypoelliptiques satisfaisant la condition de Hörmander. Nous montrons qu'une version adéquate de la propriété de Feller forte est vérifiée par de tels systèmes et nous en déduisons que, sous une propriété de contrôlabilité usuelle, ils admettent une unique solution stationnaire qui ait un sens physique.

L'ingrédient principal de notre analyse est une borne supérieure sur les moments inverses de la matrice de Malliavin associée, conditionnée au passé du bruit.

1. Introduction

In this paper we study the ergodic properties of stochastic differential equations (SDEs) of the type

$$\mathrm{d}X_t = V_0(X_t)\,\mathrm{d}t + \sum_{i=1}^{d} V_i(X_t) \circ \mathrm{d}B_t^i, \quad X_0 \in \mathbb{R}^n, \tag{1.1}$$

driven by independent fractional Brownian motions (fBm) B^i with fixed Hurst index $H \in (1/2, 1)$. Recall that fBm is the centred Gaussian process with $B_0^i = 0$ and

$$\mathbb{E}\big|B_t^i - B_s^i\big|^2 = |t - s|^{2H}.$$

Although the basic theory of SDEs driven by fBm (for $H \in (1/4, 1)$) is now well established (see, for example, [6, 10,24,31]), the ergodicity of such SDEs has only been studied recently. The main difficulty is that non-overlapping increments of fBm are not independent, so that we are dealing with processes that do not have the Markov property. As a consequence, the traditional ergodic theory for Markov processes as in [26] does not apply to this situation.

To our knowledge, the first result on the ergodicity of (1.1) was given in [11], where the author considers the case where the vector fields V_i for $i > 0$ are *constant* and *non-degenerate* in the sense that the corresponding matrix is of rank n. In this case, an explicit coupling argument allows to show that any solution converges to the unique stationary solution at speed bounded above by approximately $t^{-1/8}$. The argument however relies in an essential way on the additivity of the noise.

A more general construction was given in [12,16], where the authors built a general framework of stochastic dynamical systems (SDSs) very similar to that of random dynamical systems (RDSs) [1]. In this framework, the notion of an "invariant measure" can be defined for an autonomous equation driven by an ergodic process with stationary but not necessarily independent increments. Besides technical questions of continuity, the main difference between the framework of SDSs and that of RDSs is one of viewpoint. An RDS is considered as a dynamical system on the state space times the space of all 'futures' of the driving process, endowed with a cocycle structure. An SDS on the other hand is considered as a Markov process on the state space times the space of all 'pasts' of the driving process, endowed with a cocycle structure. This shift of viewpoint has two important advantages when studying the ergodic properties of such systems:

- Since the 'future' of the driving noise is not part of the augmented phase space, we do not have to deal with 'unphysical' stationary solutions, where the value of the process at a given time is a function of the future realisation of the driving noise.
- It allows to build our intuition based on the theory of Markov processes, rather than the theory of dynamical systems. There, many criteria for the uniqueness of an invariant measure are known to exist.

Inspired by the celebrated Doob–Khasminskii theorem (see, for example, [8]), [12,16] introduced a notion of a strong Feller property for SDSs that is a natural generalisation of the same notion for Markov semigroups. It turns out that for a large class of 'quasi-Markovian' SDSs (this includes SDEs of the type (1.1), but also many other examples), uniqueness of the invariant measure then follows like in the Markovian case from the strong Feller property, combined with some type of topological irreducibility. To illustrate these results, denote by $A_{t,x}$ the closure of the set of all points that are accessible at time t for solutions to (1.1) starting at x for some smooth control B. Then, one has

Theorem 1.1. *If there exists $t > 0$ such that (1.1) is strong Feller at time t and such that $\bigcap_{x \in \mathbb{R}^n} A_{t,x} \neq \phi$, then (1.1) can have at most one invariant measure.*

Remark 1.2. *Since (1.1) does not determine a Markov process, it is not clear a priori what is the correct notion of an invariant measure. The notion retained here is the one introduced in [12,16]. Essentially, an invariant measure for (1.1) is the law of a stationary solution which does not 'look into the future.' Similarly, it is not clear what is the correct notion of a 'strong Feller property' for such a system. Precise definitions can be found in Section 2.1.*

Furthermore, it was shown in [16] that if (1.1) is 'elliptic' in the sense that the matrix determined by $\{V_i(x)\}_{i>0}$ is of full rank at every $x \in \mathbb{R}^n$, then it does indeed satisfy the strong Feller property. In the case of SDEs driven by white noise, it is known on the other hand that the strong Feller property holds whenever the vector fields V_i satisfy Hörmander's bracket condition, namely that the Lie algebra generated by $\{\partial_t + V_0, V_1, \ldots, V_d\}$ spans $\mathbb{R}^{n+1}$ (the additional coordinate corresponds to the ∂_t direction) at every point. See also Assumption 4.1 below for a precise formulation.

Note at this stage that the strong Feller property of (1.1) is closely related to the existence of densities for the laws of its solutions, but cannot be deduced from it directly. The existence of smooth densities under an ellipticity assumption was shown in [15,29], while it was shown under the assumption of Hörmander's bracket condition in [3]. In this paper, we are going to address the question of showing that the strong Feller property (in the sense given in [12,16], see also Definition 2.1 below) holds for (1.1) under Hörmander's bracket condition.

Our main result is:

Theorem 1.3. *Assume that the vector fields $\{V_i\}$ have bounded derivatives of all orders for $i \geq 0$ and furthermore are bounded for $i \geq 1$. Then, if Hörmander's bracket condition holds at every point, (1.1) is strong Feller. In particular, if there exists $t > 0$ such that $\bigcap_{x \in \mathbb{R}^n} A_{t,x} \neq \phi$ and if there exist constants $M_1, M_2 > 0$ such that*

$$\langle x, V_0(x) \rangle \leq M_1 - M_2|x|^2, \tag{1.2}$$

then (1.1) admits exactly one invariant measure.

Remark 1.4. *Since we would like to satisfy (1.2) in order to guarantee the existence of an invariant measure by [16], Proposition 4.6, we do not impose that V_0 is bounded. This creates some technical difficulties in obtaining a priori bounds on the solutions which are not present in [3].*

Remark 1.5. *The notion of an 'invariant measure' studied in [2] is much more restrictive than the one studied here, since they considered the situation where the initial condition of the equations is* independent *of the driving noise, whereas we consider the case where interdependence with the past of the driving noise is allowed. In particular, even the case $n = d = 1$ with $V_0(x) = -x$ and $V_1(x) = 1$ admits no invariant measure in the sense of [2].*

On the other hand, our definition of an invariant measure is more restrictive than that of a random invariant measure in [1]. Since it generalises the notion of an invariant measure for a Markov process, this is to be expected. Indeed, there are simple examples of elliptic diffusions on the circle whose Markov semigroup admits a unique invariant measure, but that admit more than one random invariant measure when viewed as RDS's.

The remainder of the article is structured as follows. In Section 2, we set up our notations and recall the relevant results from [11] and [16]. In Section 3, we derive the necessary moment estimates for the solutions of the SDE (1.1). In Section 4, we then obtain an invertibility result on the conditioned Malliavin covariance matrix of the process (1.1), provided that Hörmander's condition holds. Similarly to [3], this allows us to show the smoothness of the laws of solutions to (1.1), conditional on the past of the driving noise. In Section 5, we finally show that the strong Feller property is satisfied for (1.1) under Hörmander's condition, thereby concluding the proof of Theorem 1.3.

2. Preliminaries

In this section, we describe the general framework in which we view solutions to (1.1) and we recall some of the basic results from [11,12,16].

2.1. *General framework*

Let $\mathcal{C}_0^\infty(\mathbb{R}_-; \mathbb{R}^d)$ be the set of smooth compactly supported functions on $\mathbb{R}_-$ which take the value 0 at the origin. For $\gamma, \delta \in (0, 1)$ define $\mathcal{W}_{(\gamma,\delta)}$ to be the completion of $\mathcal{C}_0^\infty(\mathbb{R}_-; \mathbb{R}^d)$ with respect to the norm

$$\|\omega\|_{(\gamma,\delta)} \equiv \sup_{\substack{s,t \in \mathbb{R}_- \\ s \neq t}} \frac{|\omega(t) - \omega(s)|}{|t - s|^\gamma (1 + |t| + |s|)^\delta}. \tag{2.1}$$

We write $\widetilde{\mathcal{W}}_{(\gamma,\delta)}$ for the corresponding space containing functions on $\mathbb{R}_+$ instead of $\mathbb{R}_-$. Note that when restricted to a compact time interval, the norm (2.1) is equivalent to the usual Hölder norm with exponent γ. Moreover, $\mathcal{W}_{(\gamma,\delta)}$ is a separable Banach space.

For $H \in (1/2, 1)$, $\gamma \in (1/2, H)$ and $\gamma + \delta \in (H, 1)$, it can be shown that there exists a Borel probability measure $\mathbb{P}_H$ on $\Omega = \mathcal{W}_{(\gamma,\delta)} \times \widetilde{\mathcal{W}}_{(\gamma,\delta)}$ such that the canonical process associated to $\mathbb{P}_H$ is a two-sided fractional Brownian motion with Hurst parameter H [11,16]. We fix such values γ, δ and H once and for all and we drop the subscripts (γ, δ) from the spaces $\mathcal{W}$ and $\widetilde{\mathcal{W}}$ from now on. Note that $\mathbb{P}_H$ is a product measure if and only if $H = 1/2$, but in general we can disintegrate it into a projected measure (denoted again by $\mathbb{P}_H$) on $\mathcal{W}$ and regular conditional probabilities $\mathcal{P}(\omega, \cdot)$ on $\widetilde{\mathcal{W}}$. Since $\mathbb{P}_H$ is Gaussian, there is a unique choice of $\mathcal{P}(\omega, \cdot)$ which is weakly continuous in ω.

We denote by $\varphi : \mathbb{R}^n \times \widetilde{\mathcal{W}} \to \mathcal{C}(\mathbb{R}_+, \mathbb{R}^n)$ the solution map to (1.1). It is well known [15,34] that if the vector fields V_i have bounded derivatives, then this solution is well defined in a pathwise sense (integrals are simply Riemann–Stieltjes integrals) for all times. We also define the shift map $\theta_t : \mathcal{W} \times \widetilde{\mathcal{W}} \to \mathcal{W} \times \widetilde{\mathcal{W}}$ by identifying elements $(\omega, \tilde{\omega}) \in \mathcal{W} \times \widetilde{\mathcal{W}}$ with the corresponding 'concatenated' function $(\omega \sqcup \tilde{\omega}) : \mathbb{R} \to \mathbb{R}^d$ and setting

$$\theta_t(\omega, \tilde{\omega}) \sim (\omega \sqcup \tilde{\omega})(t + \cdot) - (\omega \sqcup \tilde{\omega})(\cdot).$$

Denote $\varphi_t(x, \tilde{\omega}) = \varphi(x, \tilde{\omega})(t)$ and denote by $\Phi_t : \mathbb{R}^n \times \mathcal{W} \times \widetilde{\mathcal{W}} \to \mathbb{R}^n \times \mathcal{W}$ the augmented solution map given by

$$\Phi_t(x, \omega, \tilde{\omega}) = (\varphi_t(x, \tilde{\omega}), \Pi_{\mathcal{W}} \theta_t(\omega, \tilde{\omega})),$$

where $\Pi_{\mathcal{W}}$ is the projection onto the $\mathcal{W}$-component. For a measurable function $f : \mathcal{X} \mapsto \mathcal{Y}$ between two measurable space $\mathcal{X}, \mathcal{Y}$ and a measure μ on $\mathcal{X}$ we define the push forward measure $f^* \mu = \mu \circ f^{-1}$. With this notation, we can view the solutions to (1.1) as a Markov process on $\mathbb{R}^n \times \mathcal{W}$ with transition probabilities given by

$$\mathcal{Q}_t(x, \omega; \cdot) = \Phi_t(x, \omega, \cdot)^* \mathcal{P}(\omega, \cdot).$$

These transition probabilities can actually be shown to be Feller [16], but they are certainly *not* strong Feller in the usual sense, since transition probabilities starting from different instances of ω remain mutually singular for all times. Instead, we will use a notion of strong Feller property that is better suited for the particular structure of the problem at hand, see Definition 2.1 below.

However, the question of uniqueness of the invariant measure for (1.1) should not be interpreted as the question of uniqueness of the invariant measure for $\mathcal{Q}_t$. This is because one might imagine that the augmented phase space $\mathbb{R}^n \times \mathcal{W}$ contains some 'redundant' randomness that is not used to describe the stationary solutions to (1.1). (This would be the case, for example, 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 $\mathbb{R}^n \times \mathcal{W}$ to $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^n)$ given by

$$\bar{\mathcal{Q}}(x, \omega; \cdot) = \varphi(x, \cdot)^* \mathcal{P}(\omega, \cdot).$$

This is the conditional law of the solution to (1.1) given a realisation ω of the past of the driving noise. The action of the Markov transition kernel $\bar{\mathcal{Q}}$ on a measure μ on $\mathbb{R}^n \times \mathcal{W}$ is given by

$$\bar{\mathcal{Q}}\mu(A) = \int_{\mathbb{R}^n \times \mathcal{W}} \bar{\mathcal{Q}}(x, \omega; A)\mu(\mathrm{d}x, \mathrm{d}\omega).$$

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

$$\mu \sim \nu \quad \Longleftrightarrow \quad \bar{\mathcal{Q}}\mu = \bar{\mathcal{Q}}\nu.$$

With these definitions at hand, the statement 'the invariant measure for (1.1) is unique' should be interpreted as 'the Markov semigroup $\mathcal{Q}_t$ has a unique invariant measure, modulo the equivalence relation $\sim$.'

2.2. *Ergodicity of SDEs driven by fBm*

We now summarise some of the relevant results from [11,12,16] giving conditions for the uniqueness of the invariant measure for (1.1). This requires a notion of 'strong Feller' property for (1.1). We stress again that the definition given here has *nothing to do* with the strong Feller property of $\mathcal{Q}_t$. It rather generalises 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.

Let $R_t : \mathcal{C}(\mathbb{R}_+, \mathcal{X}) \mapsto \mathcal{C}([t, \infty), \mathcal{X})$ denote the natural restriction map and let $\| \cdot \|_{\mathrm{TV}}$ denote the total variation norm. We then say that:

Definition 2.1. *The solutions to* (1.1) *are said to be strong Feller at time t if there exists a jointly continuous function* $\ell : (\mathbb{R}^n)^2 \times \mathcal{W} \to \mathbb{R}_+$ *such that*

$$\left\| R_t^* \bar{\mathcal{Q}}(x, \omega; \cdot) - R_t^* \bar{\mathcal{Q}}(y, \omega; \cdot) \right\|_{\mathrm{TV}} \le \ell(x, y, \omega), \tag{2.2}$$

and $\ell(x, x, \omega) = 0$ *for every* $x \in \mathbb{R}^n$ *and every* $\omega \in \mathcal{W}$.

We also introduce the following notion of irreducibility:

Definition 2.2. *The solutions to* (1.1) *are said to be* topologically irreducible *at time t if, for every* $x \in \mathbb{R}^n$, $\omega \in \mathcal{W}$ *and every non-empty open set* $U \subset \mathbb{R}^n$, *one has* $\mathcal{Q}_t(x, \omega; U \times \mathcal{W}) > 0$.

The following result from [16], Theorem 3.10, is the main abstract uniqueness result on which this article is based:

Theorem 2.3. *If there exist times $s > 0$ and $t > 0$ such that the solutions to* (1.1) *are strong Feller at time t and irreducible at time s, then* (1.1) *can have at most one invariant measure.*

Therefore, in order to prove Theorem 1.3, the only missing ingredient that we need to establish is the strong Feller property for (1.1) under Hörmander's bracket condition.

2.3. *Notations and definitions*

For $T > 0$ and measurable $f : [0, T] \mapsto \mathbb{R}^n$, set

$$\|f\|_{0,T,\infty} = \sup_{t \in [0,T]} |f(t)|, \qquad \|f\|_{0,T,\gamma} = \sup_{s,t \in [0,T]} \frac{|f(t) - f(s)|}{|t - s|^\gamma}. \tag{2.3}$$

For $\alpha \in (0, 1)$, we also define the fractional integration operator $\mathcal{I}^\alpha$ and the corresponding fractional differentiation operator $\mathcal{D}^\alpha$ by

$$\begin{aligned}
\mathcal{I}^\alpha f(t) &\equiv \frac{1}{\Gamma(\alpha)} \int_0^t (t - s)^{\alpha - 1} f(s) \, \mathrm{d}s, \\
\mathcal{D}^\alpha f(t) &\equiv \frac{1}{\Gamma(1 - \alpha)} \frac{\mathrm{d}}{\mathrm{d}t} \int_0^t (t - s)^{-\alpha} f(s) \, \mathrm{d}s.
\end{aligned} \tag{2.4}$$

Remark 2.4. *The operators $\mathcal{I}^\alpha$ and $\mathcal{D}^\alpha$ are inverses of each other, see* [32] *for a survey of fractional integral operators.*

The reason why these operators are crucial for our analysis is that the Markov transition kernel $\mathcal{P}$ appearing in Section 2.1 is given by translates of the image of Wiener measure under $\mathcal{I}^{H-1/2}$, see also Lemma 2.7 below.

Indeed, by the celebrated Mandelbrot–Van Ness representation of the fBm [27], we may express the two-sided fBm B with Hurst parameter $H \in (0, 1)$ in terms of a two-sided standard Brownian motion W as

$$B_t = \alpha_H \int_{-\infty}^0 (-r)^{H-1/2} (\mathrm{d}W_{r+t} - \mathrm{d}W_r) \tag{2.5}$$

for some $\alpha_H > 0$. The advantage of this representation is that it is invariant under time-shifts, so that it is natural for the study of ergodic properties, see [33] for more details.

Define now the operator $\mathcal{G} : \mathcal{W} \to \widetilde{\mathcal{W}}$ by

$$\mathcal{G}\omega(t) \equiv \gamma_H \int_0^\infty \frac{1}{r} g\left(\frac{t}{r}\right) \omega(-r) \, \mathrm{d}r, \tag{2.6}$$

where the kernel g is given by

$$g(x) \equiv x^{H-1/2} + (H - 3/2)x \int_0^1 \frac{(u + x)^{H-5/2}}{(1 - u)^{H-1/2}} \, \mathrm{d}u, \tag{2.7}$$

and the constant γ_H is given by $\gamma_H = (H - 1/2)\alpha_H \alpha_{1-H}$. Here, α_H is the constant appearing in (2.5). It was shown in [16], Proposition A.2, that $\mathcal{G}$ is indeed a bounded linear operator from $\mathcal{W}$ into $\widetilde{\mathcal{W}}$. Furthermore, we can quantify its behaviour near $t = 0$ in the following way:

Lemma 2.5. *On any time interval bounded away from* 0, *the map* $t \mapsto \mathcal{G}\omega(t)$ *is* $\mathcal{C}^\infty$. *Furthermore, if we set* $f_\omega(t) = t\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{G}\omega(t)$, *then we have* $f_\omega(0) = 0$ *and for every* $T > 0$, *there exists a constant* M_T *such that* $\|f_\omega\|_{0,T,\gamma} < M_T \|\omega\|_{(\gamma,\delta)}$.

The proof of this result is postponed to the Appendix.

Remark 2.6. *Since* $\mathbb{P}_H$ *is Gaussian, it follows in particular that both* $\|\mathcal{G}\omega\|_{0,T,\gamma}$ *and* $\|f_\omega\|_{0,T,\gamma}$ *have exponential moments under* $\mathbb{P}_H$ *by Fernique's theorem* [5].

Let $\tau_h : w \mapsto w + h$ denote the translation map on $\widetilde{\mathcal{W}}$. We cite the following result from [16], Lemma 4.2:

Lemma 2.7. *The regular conditional probabilities* $\mathcal{P}(\omega, \cdot)$ *of* $\mathbb{P}_H$ *given* $\omega \in \mathcal{W}$ *are*

$$\mathcal{P}(w, \cdot) = \left(\tau_{\mathcal{G}w} \circ \mathcal{I}^{H-1/2}\right)^* \mathbf{W},$$

where $\mathbf{W}$ *is the standard* d-*dimensional Wiener measure over* $\mathbb{R}_+$ *and* $\mathcal{I}^\alpha$ *is as defined in* (2.4).

We will henceforth interpret the above lemma in the following way. The driving fractional Brownian motion B can be written as the sum of two independent processes:

$$B_t \stackrel{d}{=} \tilde{B}_t + (\mathcal{G}\omega)(t) = \tilde{B}_t + m_t, \qquad \tilde{B}_t = \alpha_H \int_0^t (t-s)^{H-1/2}\, \mathrm{d}W(s), \tag{2.8}$$

where W is a standard Wiener process independent of the 'past' $\omega \in \mathcal{W}$. This notation will be used repeatedly in the sequel.

Remark 2.8. *In the notation of Section* 2.1, *we will repeatedly use the notations* $\tilde{\mathbb{E}}$ (*and* $\tilde{\mathbb{P}}$) *for conditional expectations* (*and probabilities*) *over* $\tilde{\omega}$ *with* ω *fixed. In the notation* (2.8), *this is the same as fixing* ω *and taking expectations with respect to* $\tilde{B}$.

3. Estimates on the solutions

In this section, we derive moment estimates for the solutions to equations of the form:

$$\mathrm{d}X_t = V_0(X_t)\, \mathrm{d}t + \sum_{i=1}^d V_i(X_t) \circ \mathrm{d}B_t^i, \quad t \geq 0, \tag{3.1}$$

where $x = X_0 \in \mathbb{R}^n$ and $V_i, i = 1, 2, \ldots, d$, are bounded, $\mathcal{C}^\infty$ vector fields in $\mathbb{R}^n$ with bounded derivatives, and V_0 is a possibly *unbounded* $\mathcal{C}^\infty$ vector field with bounded derivatives. Our bounds will be purely pathwise bounds, so the fact that the B^i's are sample paths of fractional Brownian motion is irrelevant, except to get a bound on their Hölder regularity.

All that we will assume is that the driving process is γ-Hölder continuous for some $\gamma > 1/2$. The main result in this section are a priori bounds not only on (3.1), but also on its Jacobian and its second variation with respect to the initial condition. This will be a slight generalisation of the results in [15], which required the drift term V_0 to be bounded.

Let $\varphi_t(x, \omega) \equiv X_t$ denote the flow map solving (3.1) with initial condition $X_0 = x$ and define its Jacobian by

$$J_{0,t} \equiv \frac{\partial \varphi_t(x, \omega)}{\partial x}.$$

For notational convenience, set $V = (V_1, V_2, \ldots, V_d)$. Then, we can write (3.1) in compact form:

$$\mathrm{d}X_t = V_0(X_t)\, \mathrm{d}t + V(X_t) \circ \mathrm{d}B_t, \tag{3.2a}$$

$$\mathrm{d}J_{0,t} = V_0'(X_t)J_{0,t}\, \mathrm{d}t + V'(X_t)J_{0,t} \circ \mathrm{d}B_t, \tag{3.2b}$$

$$\mathrm{d}J_{0,t}^{-1} = -J_{0,t}^{-1}V_0'(X_t)\, \mathrm{d}t - J_{0,t}^{-1}V'(X_t) \circ \mathrm{d}B_t. \tag{3.2c}$$

Here, both J and J^{-1} are $n \times n$ matrices, and $J_{0,0} = J_{0,0}^{-1} = 1$. One crucial ingredient in order to obtain the bounds on the Malliavin matrix required to show the strong Feller property is control on the moments of both the solution and its Jacobian. These bounds do not quite follow from standard results since most of them require that V_0 is also bounded. However we cannot assume this since we need condition (1.2) for showing ergodicity.

3.1. *Moments estimates*

Let $\mathcal{L}_k(\mathcal{X}_1, \ldots, \mathcal{X}_k; \mathcal{Y})$ denote the set of k-multilinear maps from $\mathcal{X}_1 \times \cdots \times \mathcal{X}_k$ to $\mathcal{Y}$. As usual, $\mathcal{L}_1$ is denoted by $\mathcal{L}$. In this section, we consider a system of differential equations of the form

$$x_t = x_0 + \int_0^t f_0(x_r) \, dr + \int_0^t f(x_r) \, dB_r, \tag{3.3a}$$

$$y_t = y_0 + \int_0^t g(x_r)(y_r, dB_r), \tag{3.3b}$$

where $x_t \in \mathbb{R}^n$, $y_t \in \mathbb{R}^m$, $B : [0, \infty) \mapsto \mathbb{R}^d$ is a Hölder function of order $\gamma > 1/2$ and $f_0 : \mathbb{R}^n \to \mathbb{R}^n$, $f : \mathbb{R}^n \mapsto \mathcal{L}(\mathbb{R}^d; \mathbb{R}^n)$, $g : \mathbb{R}^n \mapsto \mathcal{L}_2(\mathbb{R}^m, \mathbb{R}^d; \mathbb{R}^m)$ are given $\mathcal{C}^1$ functions. Note that (3.2) is indeed of this form with $m = 2n^2$.

Remark 3.1. *The reason why we do not include a 'dr' term in* (3.3b) *is because this is already covered by the present result by setting $B_1(r) = r$, for example. Treating this term separately like in* (3.3a) *might allow to slightly improve our results, but the present formulation is sufficient for our purpose.*

Regularity of solutions to equations of this kind has been well studied, pioneered by the technique of Young [34] and more recently using the theory of rough paths (see, for example, [10,21] and the references therein). Using fractional derivatives, it was shown in [15], Theorem 3, that

$$\sup_{0 \le t \le T} |y_t| \le 2^{1 + M \|B\|_{0,T,\gamma}^{1/\gamma}} |y_0|, \tag{3.4}$$

where $\|B\|_{0,T,\gamma}$ is the Hölder norm defined in (2.3) and M is a constant depending on the supremum norms of f_0, f, g and their first derivatives. As mentioned earlier, these estimates are not sufficient to obtain moment bounds on the Jacobian $J_{0,t}$ and its inverse, since we wish to consider situations where f_0 is unbounded, so that its supremum norm is not finite.

However it turns out that an estimate of the type (3.4) holds even if only the derivative of f_0 is bounded, thanks to the fact that the corresponding "driving noise" (t) is actually a differentiable function. This is the content of the following result:

Lemma 3.2. *For the processes x_t and y_t defined in* (3.3), *we have the pathwise bounds*:

$$\|x\|_{0,T,\gamma} \le M\big(1 + |x_0|\big)\big(1 + \|B\|_{0,T,\gamma}\big)^{1/\gamma}, \tag{3.5a}$$

$$\|y\|_{0,T,\gamma} \le M\big(1 + |x_0|\big)|y_0| e^{M\|B\|_{0,T,\gamma}^{1/\gamma}}, \tag{3.5b}$$

where M is an absolute constant which depends only on $\|f\|_\infty$, $\|f'\|_\infty$, $\|f_0'\|_\infty$, $\|g\|_\infty$, $\|g'\|_\infty$ and T.

Proof. The proof is almost identical to that of [15], so we only try to highlight the main differences. For $s, t \in [0, T]$, we have the a priori bound (see [15], p. 403),

$$\left| \int_s^t f(x_r) \, dB_r \right| \le M \|B\|_{0,T,\gamma}\big((t-s)^\gamma + \|x\|_{s,t,\gamma}(t-s)^{2\gamma}\big),$$

where the constant M is independent of $\|B\|_{0,T,\gamma}$, but depends on f. Since f_0 grows at most linearly, we also have

$$\left| \int_s^t f_0(x_r) \, dr \right| \le M\big((t-s) + |x_s|(t-s) + \|x\|_{s,t,\gamma}(t-s)^{\gamma+1}\big),$$

and therefore

$$\sup_{s \leq m, n \leq t} \left| \int_m^n f_0(x_r) \, dr \right| \leq M\big((t-s) + |x_s|(t-s) + 2\|x\|_{s,t,\gamma}(t-s)^{\gamma+1}\big). \tag{3.6}$$

Thus we have

$$\|x\|_{s,t,\gamma} \leq M_1\big(1 + \|B\|_{0,T,\gamma} + |x_s|(t-s)^{1-\gamma} + \big(\|B\|_{0,T,\gamma} + 1\big)\|x\|_{s,t,\gamma}(t-s)^{\gamma}\big)$$

for a constant M_1 depending on f and the terminal time T. Set $\Delta = (2M_1(\|B\|_{0,T,\gamma} + 1))^{-1/\gamma}$. Then for $|t-s| < \Delta$,

$$\|x\|_{s,t,\gamma} \leq 2M_1\big(1 + \|B\|_{0,T,\gamma} + |x_s|\Delta^{1-\gamma}\big). \tag{3.7}$$

Since by definition we have:

$$\|x\|_{s,t,\infty} \leq |x_s| + \|x\|_{s,t,\gamma}|t-s|^{\gamma}$$

for $|t-s| < \Delta$, from (3.7) it follows that

$$\|x\|_{s,t,\infty} \leq |x_s|(1 + 2M_1\Delta) + 1. \tag{3.8}$$

Iterating the above estimate for $N = T/\Delta$, it follows that

$$\|x\|_{0,T,\infty} \leq |x_0|(1 + 2M_1\Delta)^N + \sum_{k=1}^{N-1} \big(1 + 2M_1\Delta\big)^k \leq \big(1 + |x_0|\big)(1 + 2M_1\Delta)^N N$$

$$= \big(1 + |x_0|\big)(1 + 2M_1\Delta)^{T/\Delta} T/\Delta.$$

Using the fact that $(1 + \frac{x}{\Delta})^{\Delta} \leq e^x$ for $\Delta > 0$, this finally yields

$$\|x\|_{0,T,\infty} \leq \big(1 + |x_0|\big) e^{2M_1 T} T \big(2M_1\big(\|B\|_{0,T,\gamma} + 1\big)\big)^{1/\gamma}. \tag{3.9}$$

Note that in the last step of the above argument we bounded $(1 + 2M_1\Delta)^{T/\Delta}$ from above by $e^{2M_1 T}$ and thus obtained a constant which is exponential in T, but *independent* of $\Delta = \mathcal{O}(\|B\|_{0,T,\gamma}^{-1/\gamma})$. If the noise corresponding to the vector field f_0 were *not* differentiable but only Hölder continuous, then instead of (3.8), the exponent would include a power of the Hölder constant of the driving noise, thus yielding an estimate comparable to those obtained in [15].

Thus substituting the bound (3.9) in (3.7) yields that, for $t - s < \Delta$,

$$\|x\|_{s,t,\gamma} \leq M\big(1 + |x_0|\big)\big(1 + \|B\|_{0,T,\gamma}\big). \tag{3.10}$$

Using the fact the function $f(x) = x^{\gamma}$ is concave for $\gamma < 1$, it is shown in Lemma A.2 (proof given in the Appendix) that for $0 = u_0 < u_1 < u_2 < \cdots < u_{N-1} < u_N = T$, we have the bound

$$\|x\|_{0,T,\gamma} \leq N^{1-\gamma} \max_{0 \leq i \leq N-1} \|x\|_{u_i, u_{i+1}, \gamma}. \tag{3.11}$$

Thus from (3.10) and (3.11) we deduce that

$$\|x\|_{0,T,\gamma} \leq M\big(1 + |x_0|\big)\big(1 + \|B\|_{0,T,\gamma}\big) N^{1-\gamma} = M\big(1 + |x_0|\big)\big(1 + \|B\|_{0,T,\gamma}\big)^{1/\gamma},$$

proving the claim made in (3.5a).

Now we come to the second part of Lemma 3.2. Once again, for $s, t \in [0, T]$ we have the apriori bound (see [15], p. 406)

$$\left| \int_s^t g(x_r) y_r \, dB_r \right| \leq \tilde{M}\big(1 + \|B\|_{0,T,\gamma}\big)\big(\|g\|_\infty \|y\|_{s,t,\infty}(t-s)^{\gamma}$$

$$+ \big(\|g\|_\infty \|y\|_{s,t,\gamma} + \big\|g'\big\|_\infty \|y\|_{s,t,\infty} \|x\|_{s,t,\gamma}\big)(t-s)^{2\gamma}\big) \tag{3.12}$$

for a constant $\tilde{M}$ independent of B. Thus, setting as before $\Delta = (2M_1(\|B\|_{0,T,\gamma} + 1))^{-1/\gamma}$, we deduce from (3.7) and (3.12) that,

$$\|y\|_{s,t,\gamma} \leq M_2\big(1 + \|B\|_{0,T,\gamma}\big)\big(\|y\|_{s,t,\infty} + \big(1 + \|B\|_{0,T,\gamma}\big)\|y\|_{s,t,\infty}\Delta^\gamma + \|y\|_{s,t,\gamma}\Delta^\gamma\big) \tag{3.13}$$

for a large enough constant $M_2 \geq \tilde{M}(1 + |x_0|Te^{M_1 T} + \|g\|_\infty + \|g'\|_\infty)$. From the estimate (3.13), a straightforward calculation (for instance, see [15], p. 407) yields

$$\|y\|_{0,T,\infty} \leq M2^{MT(1+\|B\|_{0,T,\gamma})^{1/\gamma}}|y_0|,$$

$$\|y\|_{0,T,\gamma} \leq M\big(1 + \|B\|_{0,T,\gamma}\big)^{2/\gamma}2^{MT(1+\|B\|_{0,T,\gamma})^{1/\gamma}}|y_0|$$

for a large constant M, thus concluding the proof. $\qquad\square$

In the next sections, we will also require bounds on the *second* derivative of the solution flow with respect to its initial condition. Given an equation of the type (3.3), the second variation $z_t \in \mathbb{R}^p$ of x with respect to its initial condition satisfies an equation of the type

$$z_t = z_0 + \int_0^t h(x_r)(z_r, \mathrm{d}B_r) + \int_0^t \hat{h}(x_r)(y_r, y_r, \mathrm{d}B_r). \tag{3.14}$$

Here, the maps $h\colon \mathbb{R}^n \to \mathcal{L}_2(\mathbb{R}^p, \mathbb{R}^d; \mathbb{R}^p)$ and $\hat{h}\colon \mathbb{R}^n \to \mathcal{L}_3(\mathbb{R}^m, \mathbb{R}^m, \mathbb{R}^d; \mathbb{R}^p)$ are bounded with bounded derivatives. Bounds on z are covered by the following corollary to Lemma 3.2:

Corollary 3.3. *For the process z_t defined in* (3.14), *we have the pathwise bounds*:

$$\|z\|_{0,T,\gamma} \leq M\big(\big(1 + |x_0|^5\big)\big(1 + |y_0|^2\big) + |z_0|\big(1 + |x_0|\big)\big)\exp\big(M\|B\|_{0,T,\gamma}^{1/\gamma}\big),$$

where M is an absolute constant which depends only on $\|f\|_\infty$, $\|f'\|_\infty$, $\|f_0'\|_\infty$, $\|g\|_\infty$, $\|g'\|_\infty$ and T.

Proof. Denote by Z_t the $\mathbb{R}^{p \cdot p}$-valued solution to the homogeneous equation

$$Z_t\xi = \xi + \int_0^t h(x_r)(Z_r\xi, \mathrm{d}B_r), \quad \xi \in \mathbb{R}^p.$$

Note that the inverse matrix Z_t^{-1} then satisfies a similar equation, namely

$$Z_t^{-1}\xi = \xi - \int_0^t Z_r^{-1}h(x_r)(\xi, \mathrm{d}B_r), \quad \xi \in \mathbb{R}^p.$$

It follows from Lemma 3.2 that both Z and Z^{-1} satisfy the bound (3.5b), just like y did. On the other hand, z can be solved by the variation of constants formula:

$$z_t = Z_t z_0 + Z_t \int_0^t Z_s^{-1}\hat{h}(x_s)(y_s, y_s, \mathrm{d}B_s). \tag{3.15}$$

We now use the fact that if f and g are two functions that are γ-Hölder continuous for some $\gamma > \frac{1}{2}$, then there exists a constant M such that

$$\left\|\int_0^\cdot f(t)\,\mathrm{d}g(t)\right\|_{0,T,\gamma} \leq M\big(\|f\|_{0,T,\gamma} + \big|f(0)\big|\big)\|g\|_{0,T,\gamma}. \tag{3.16}$$

(See [34] or, for example, [25], Eq. (2.1), for a formulation that is closer to (3.16).) Furthermore,

$$\|fg\|_{0,T,\gamma} \leq M\big(\big|f(0)\big| + \|f\|_{0,T,\gamma}\big)\big(\big|g(0)\big| + \|g\|_{0,T,\gamma}\big),$$

so that, from (3.15),

$$\|z\|_{0,T,\gamma} \le |z_0| \|Z\|_{0,T,\gamma} + C\big(1 + \|Z\|_{0,T,\gamma}\big)\big(1 + \big\|Z^{-1}\big\|_{0,T,\gamma}\big)$$
$$\times \|\hat{h}'\|_{\infty}\big(|x_0| + \|x\|_{0,T,\gamma}\big)\big(|y_0| + \|y\|_{0,T,\gamma}\big)^2 \|B\|_{0,T,\gamma}.$$

The claim now follows by a simple application of Lemma 3.2. $\qquad\square$

The above two lemmas immediately give us the required *conditional* moment estimates for the Jacobian $J_{0,T}$.

Lemma 3.4. *Let X_t, $J_{0,t}, m_t$ be as defined in (3.2a), (3.2b) and (2.8) respectively. Then for any $T > 0$, $p \ge 1$ and $\gamma \in (1/2, H)$ we have*

$$\tilde{\mathbb{E}}\big(\|J_{0,\cdot}\|_{0,T,\gamma}^p\big) < M\mathrm{e}^{M(1+\|m\|_{0,T,\gamma})^{1/\gamma}},$$
$$\tilde{\mathbb{E}}\big(\big\|J_{0,\cdot}^{-1}\big\|_{0,T,\gamma}^p\big) < M\mathrm{e}^{M(1+\|m\|_{0,T,\gamma})^{1/\gamma}}.$$

(*Recall that $\tilde{\mathbb{E}}$ was defined in Remark 2.8.*)

Proof. We first write $B_t = \tilde{B}_t + m_t$ as in (2.8). Thus (3.2b) can be written as:

$$\mathrm{d}J_{0,t} = V_0'(X_t)J_{0,t}\,\mathrm{d}t + V'(X_t)J_{0,t} \circ (\mathrm{d}\tilde{B}_t + \mathrm{d}m_t).$$

By applying Lemma 3.2 to the SDE in (3.2b) with $y_t = (t, \tilde{B}_t + m_t)$, $f_0 = V_0$, $f = (0, V_1, \ldots, V_d)$, $g = (V_0', \ldots, V_d')$ and using the fact that $1/\gamma > 1$ we obtain:

$$\|J\|_{0,T,\gamma}^p \le M\mathrm{e}^{M(1+\|\tilde{B}\|_{0,T,\gamma}^{1/\gamma}+\|m\|_{0,T,\gamma}^{1/\gamma})}. \tag{3.17}$$

The fact that $\gamma > \frac{1}{2}$ is crucial. Since $\tilde{B}$ is a Gaussian process, it follows by Fernique's theorem [5] that the Hölder norm of $\tilde{B}$ has Gaussian tails, thus proving our first claim. An identical argument for $J_{0,t}^{-1}$ finishes the proof. $\qquad\square$

Remark 3.5. *In an identical way, one can obtain similar bounds on the conditional moments of the solution and of its second variation with respect to its initial condition.*

4. Smoothness of conditional densities

In this section, we consider again our main object of study, namely the SDE

$$\mathrm{d}X_t = V_0(X_t)\,\mathrm{d}t + \sum_{i=1}^d V_i(X_t) \circ \mathrm{d}B_t^i, \quad X_0 = x_0 \in \mathbb{R}^n,$$

where B_t is a *two-sided* fBM. Our aim is to show that, if the vector fields V satisfy Hörmander's celebrated Lie bracket condition (see below), then the *conditional* distribution of X_t given $\{B_s: s \le 0\}$ almost surely has a smooth density with respect to Lebesgue measure on $\mathbb{R}^n$ for every $t > 0$.

In [3], the authors studied the same equation and showed that the law of X_t has a smooth density, but with the key difference that the driving noise B_t was a one-sided fBm and no conditioning was involved. Being able to treat the conditioned process will be crucial for the proof of the strong Feller property later on, this is why we consider it here. As explained in (2.8), we can rewrite B_t as $B_t = \tilde{B}_t + m_t$ and thus obtain

$$\mathrm{d}X_t = V_0(X_t)\,\mathrm{d}t + V(X_t)\,\mathrm{d}m_t + V(X_t) \circ \mathrm{d}\tilde{B}_t. \tag{4.1}$$

While the distribution of the conditioned process $\tilde{B}_t$ is different from that of the one-sided fBm, its covariance satisfies very similar bounds so that the results from [3] would apply. (See Proposition 4.4.)

However, the difficulty comes from the time derivative of the conditional mean m_t, which diverges at $t = 0$ like $\mathcal{O}(t^{H-1})$. To use the results from [3] directly, we would need the boundedness of the driving vector fields, uniformly in time, so that they are not applicable to our case. Our treatment of this singularity is inspired by [14], where a related situation appears due to the infinite-dimensionality of the system considered. Using an iterative argument, we will show the invertibility of a slightly modified reduced Malliavin matrix which will then, by standard arguments, yield the smoothness of laws of the SDE given in (4.1).

Before we state Hörmander's condition, let us introduce a shorthand notation. For any multiindex $I = (i_1, i_2, \ldots, i_k) \in \{0, 1, 2, \ldots, d\}^k$, we define the commutator:

$$V_I \equiv \left[V_{i_1}, \left[V_{i_2}, \ldots, [V_{i_{k-1}}, V_{i_k}] \cdots \right] \right].$$

We also define recursively the families of vector fields

$$\mathcal{V}_n = \left\{ V_I, I \in \{0, \ldots, d\}^{n-1} \times \{1, \ldots, d\} \right\}, \qquad \bar{\mathcal{V}}_n = \bigcup_{k=1}^{n} \mathcal{V}_k. \tag{4.2}$$

With these notations at hand, we now formulate Hörmander's bracket condition [17]:

Assumption 4.1. *For every $x_0 \in \mathbb{R}^n$, there exists $N \in \mathbb{N}$ such that the identity*

$$\mathrm{span}\left\{ U(x_0) : U \in \bar{\mathcal{V}}_N \right\} = \mathbb{R}^n \tag{4.3}$$

holds.

4.1. *Invertibility of the Malliavin matrix*

It is well known from the works of Malliavin, Bismut, Kusuoka, Stroock and others [4,18–20,22,23,28,31] that one way of proving the smoothness of the law of X_T (at least when driven by Brownian motion) is to first show the invertibility of the 'reduced Malliavin matrix':[1]

$$\hat{C}_T \overset{\mathrm{def}}{=} \mathcal{A}_T \mathcal{A}_T^* = \int_0^T J_{0,s}^{-1} V(X_s) V(X_s)^* \left(J_{0,s}^{-1} \right)^* \mathrm{d}s, \tag{4.4}$$

where we defined the (random) operator: $\mathcal{A}_T : L^2([0, T], \mathbb{R}^d) \mapsto \mathbb{R}^n$ by

$$\mathcal{A}_T v = \int_0^T J_{0,s}^{-1} V(X_s) v(s) \, \mathrm{d}s. \tag{4.5}$$

Unfortunately, for the stochastic control argument used in Section 5 below for the proof of the strong Feller property, the invertibility of $\hat{C}_T$ turns out not to be sufficient. We now define a family of smooth functions $h_T : [0, T] \to [0, 1]$:

$$h_T(s) = \begin{cases} 1 & \text{if } s \leq T/2, \\ 0 & \text{if } s \geq 3T/4. \end{cases} \tag{4.6}$$

and consider the (random) matrix C_T given by

$$C_T \overset{\mathrm{def}}{=} \mathcal{A}_T h_T \mathcal{A}_T^*. \tag{4.7}$$

---

[1]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 the representation (2.8), but the associated Malliavin matrix has a slightly more complicated relation to C_T than usual. This will be done in Theorem 4.11.

(See Remark 5.5 below for the reason behind introducing the function h_T and the matrix C_T.) Obviously, $C_T \leq \hat{C}_T$ as positive definite matrices, so that invertibility of C_T will in particular imply inertibility of $\hat{C}_T$, but the converse is of course not true in general. Define the matrix norm:

$$\|C_T\| = \sup_{|\varphi|=1} \langle v, C_T v \rangle, \quad \varphi \in \mathbb{R}^n.$$

Since C_T is a symmetric matrix, by the spectral theorem we have

$$\left\|C_T^{-1}\right\| = \frac{1}{\inf_{|\varphi|=1} \langle v, C_T v \rangle}, \quad \varphi \in \mathbb{R}^n.$$

The aim of this section is to show that C_T is invertible, and to obtain moment estimates for $\|C_T^{-1}\|$. The main result of this section can be formulated in the following way:

Theorem 4.2. *Assume that the vector fields $\{V_i\}$ have bounded derivatives of all orders for $i \geq 0$ and are bounded for $i > 0$. Assume furthermore that Assumption 4.1 holds, fix $T > 0$ and let C_T be as in (4.7). Then, for any $R > 0$ and any $p \geq 1$, there exists a constant M such that the bound*

$$\tilde{\mathbb{P}}\left(\inf_{\|\varphi\|=1} \langle \varphi, C_T \varphi \rangle \leq \varepsilon\right) \leq M \varepsilon^p, \quad \varphi \in \mathbb{R}^n,$$

holds for all $\varepsilon \in (0, 1]$ and for all x_0 and ω such that $|x_0| + \|\omega\|_{(\gamma,\delta)} \leq R$. Here, x_0 is the initial condition of the SDE (4.1).

Corollary 4.3. *The matrix C_T is almost surely invertible and, for any $R > 0$ and any $p \geq 1$, there exists a constant M such that*

$$\tilde{\mathbb{E}}\left\|C_T^{-p}\right\| < M,$$

uniformly over all x_0 and ω such that $|x_0| + \|\omega\|_{(\gamma,\delta)} \leq R$.

Proof. Notice that for $p \geq 1$, we have $\|C_T^{-p}\| = \|C_T^{-1}\|^p$ since C_T is symmetric an positive definite. Thus

$$\begin{aligned}
\tilde{\mathbb{E}}\left\|C_T^{-p}\right\| &= p \int_0^\infty x^{p-1} \mathbb{P}\left(\left\|C_T^{-1}\right\| \geq x\right) dx \\
&= p \int_0^\infty x^{p-1} \tilde{\mathbb{P}}\left(\inf_{\|\varphi\|=1} \langle \varphi, C_T^{-1} \varphi \rangle \leq \frac{1}{x}\right) dx < \infty,
\end{aligned}$$

where the last inequality follows from Theorem 4.2, since for $x > 1$ there exists a constant M (depending only on R) such that $\tilde{\mathbb{P}}(\inf_{\|\varphi\|=1} \langle \varphi, C_T^{-1} \varphi \rangle \leq \frac{1}{x}) \leq M x^{-1-p}$ and we are done. □

Before proceeding further, we need the following version of Norris' lemma [19,28] adapted from [3] which will be the key technical estimate needed for showing the invertibility of C_T. In the case of white driving noise, Norris' lemma essentially states that if a semimartingale is small and if one has some a priori bounds on its components, then both the bounded variation part and the martingale part have to be small. In this sense, it can be viewed as a quantitative version of the Doob–Meyer semimartingale decomposition theorem. One version of this result for fractional driving noise is given by:

Proposition 4.4. *Let $H \in (\frac{1}{2}, 1)$ and $\gamma \in (\frac{1}{2}, H)$ and let $\tilde{B}$ be as in (2.8). Then there exist exponents $q > 0$ and $\eta > 0$ such that the following statement holds.*

Consider any two families a^ε, b^ε of processes taking values in $\mathbb{R}$ and $\mathbb{R}^d$ respectively such that $\mathbb{E}(\|a^\varepsilon\|^p_{0,T,\gamma} + \|b^\varepsilon\|^p_{0,T,\gamma}) < M_p \varepsilon^{-\eta p}$, uniformly in $\varepsilon \in (0, 1]$, for some $M_p > 0$ and for every $p \geq 1$. Let y^ε be the process defined by

$$y^\varepsilon_t = \int_0^t a^\varepsilon_s \, \mathrm{d}s + \int_0^t b^\varepsilon_s \circ \mathrm{d}\tilde{B}_s, \quad t \in [0, T].$$

Then, the estimate

$$\mathbb{P}\big(\|y\|_{0,T,\infty} < \varepsilon \text{ and } \|a\|_{0,T,\infty} + \|b\|_{0,T,\infty} > \varepsilon^q\big) < C_p \varepsilon^p$$

holds for every $p > 0$, uniformly for $\varepsilon \in (0, 1]$.

Proof. Setting $f(s, t) = \mathbb{E}|\tilde{B}(t) - \tilde{B}(s)|^2$ as in [3], we have

$$f(s, t) = \frac{t^{2H}}{2H - 1} + \frac{s^{2H}}{2H - 1} - 2 \int_0^s (t - r)^{H-1/2} (s - r)^{H-1/2} \, \mathrm{d}r.$$

(We use the convention $s < t$.) A somewhat lengthy but straightforward calculation shows that this can be rewritten as

$$f(s, t) = |t - s|^{2H} \left(\frac{1}{2H - 1} + \int_0^{s/|t-s|} \big((1 + x)^{H-1/2} - x^{H-1/2}\big)^2 \, \mathrm{d}x \right).$$

Similarly, one has

$$\partial_s \partial_t f(s, t) = -2 \left(H - \frac{1}{2} \right)^2 |t - s|^{2H-2} \int_0^{s/|t-s|} (1 + x)^{H-3/2} x^{H-3/2} \, \mathrm{d}x.$$

Since the integrands of both of the integrals appearing in these expressions decay like x^{2H-3}, they are integrable so that $\tilde{B}$ is indeed a Gaussian process of 'type H', see [3], Eq. (3.5).

The proof is now a slight modification of that of [3], Proposition 3.4. It follows from [3], Eq. (3.18), and the last displayed equation in the proof of [3], Proposition 3.4, that there exist exponents q and β such that the bound

$$\mathbb{P}\big(\|y\|_{0,T,\infty} < \varepsilon \text{ and } \|a\|_{0,T,\infty} + \|b\|_{0,T,\infty} > \varepsilon^q\big) \leq C_p \varepsilon^p + \mathbb{P}\big(\|b\|^2_\gamma + \|a\|^2_\gamma > \varepsilon^{-\beta}\big)$$

holds. The claim now follows from a simple application of Chebyshev's inequality, provided that we ensure that η is such that $2\eta < \beta$. $\qquad\square$

For making further calculations easier, we now introduce some definitions that are inspired by the proofs in [14].

Definition 4.5. *A family of events $A \equiv \{A_\varepsilon\}_{\varepsilon \leq 1}$ of a probability space is called a "family of negligible events" if, for every $p \geq 1$ there exists a constant C_p such that $\mathbb{P}(A_\varepsilon) \leq C_p \varepsilon^p$ for every $\varepsilon \leq 1$.*

Definition 4.6. *Given a statement Φ_ε, we will say that "Φ_ε holds modulo negligible events," if there exists a family $\{A_\varepsilon\}$ of negligible events such that, for every $\varepsilon \leq 1$, the statement Φ_ε holds on the complement of A_ε.*

With the above definitions, Theorem 4.2 can be restated as:

Theorem 4.7. *Fix $R > 0$. Then, modulo negligible events, the bound*

$$\inf_{\|\varphi\|=1} \langle \varphi, C_T \varphi \rangle \geq \varepsilon$$

holds for all x_0 and ω such that $|x_0| + \|\omega\|_{(\gamma, \delta)} \leq R$.

Remark 4.8. *The family of events in question can (and will in general) depend on R, but once R is fixed the bounds are uniform over* $|x_0| + \|\omega\|_{(\gamma,\delta)} \leq R$. *While this uniformity is not really required in this section, it will be crucial in Section* 5.

Proof of Theorem 4.7. It suffices to show that for any fixed φ with $\|\varphi\| = 1$, the bound $\langle \varphi, C_T \varphi \rangle \geq \varepsilon$ holds modulo a family of negligible events. The conclusion then follows by a standard covering argument, see, for example, [28], p. 127. Furthermore, throughout all of this proof, we will denote by M a constant that can change from one expression to the next and only depends on R and on the vector fields V defining the problem.

Fix now such a φ and assume "by contradiction" that $\langle \varphi, C_T \varphi \rangle < \varepsilon$. We will show that this assumption leads to the conclusion that

$$\sup_{U \in \bar{\mathcal{V}}_N} \left| \langle \varphi, U(x_0) \rangle \right| < \varepsilon^{q'} \quad \text{for some } q' > 0,$$

modulo negligible events. However, if Hörmander's condition is satisfied, the conclusion above cannot hold since $\{U(x_0), U \in \bar{\mathcal{V}}_N\}$ span $\mathbb{R}^n$ so that there exists a constant c such that $\sup_{U \in \bar{\mathcal{V}}_N} |\langle \varphi, U(x_0) \rangle| \geq c$ almost surely, thus leading to a contradiction and concluding the proof. For any smooth vector field U, let us introduce the process $Z_U(t) = J_{0,t}^{-1} U(X_t)$. With this notation, we have

Lemma 4.9. *There exists* $r > 0$ *such that, modulo negligible events, the implication*

$$\langle \varphi, C_T \varphi \rangle < \varepsilon \quad \Longrightarrow \quad \left\| \langle \varphi, Z_U(\cdot) \rangle \right\|_{0,T/2,\infty} < \varepsilon^r$$

holds for all $U \in \mathcal{V}_1$.

Proof. By the definition of h, we have the bound

$$\langle \varphi, C_T \varphi \rangle = \sum_{i=1}^{d} \int_0^T \langle \varphi, J_{0,s}^{-1} V_i(X_s) \rangle^2 h_T(s) \, ds \geq \sum_{i=1}^{d} \int_0^{T/2} \langle \varphi, J_{0,s}^{-1} V_i(X_s) \rangle^2 \, ds. \tag{4.8}$$

We now wish to turn this into a lower bound of order ε^r (for some $r > 0$) on the supremum of the integrand. Our main tool for this is the bound

$$\|f\|_{0,T,\infty} \leq 2 \max \left(T^{-1/2} \|f\|_{L^2[0,T]}, \|f\|_{L^2[0,T]}^{2\gamma/(2\gamma+1)} \|f\|_{0,T,\gamma}^{1/(2\gamma+1)} \right),$$

which holds for every γ-Hölder continuous function $f : [0, T] \mapsto \mathbb{R}$. The proof of this is given in the Appendix as Lemma A.3.

By Lemma 3.4 and Chebyshev's inequality it follows that $\|\langle \varphi, J_{0,\cdot}^{-1} V(X_\cdot) \rangle\|_{0,T,\gamma} < \varepsilon^{-1/2}$ modulo negligible events. Thus if $\langle \varphi, C_T \varphi \rangle < \varepsilon$, from Lemma A.3 and (4.8) we deduce that, for each $1 \leq j \leq d$

$$\left\| \langle \varphi, J_{0,\cdot}^{-1} V_j(X_\cdot) \rangle \right\|_{0,T/2,\infty} < M \varepsilon^{(4\gamma-1)/(2(2\gamma+1))}, \tag{4.9}$$

modulo negligible events. The claim follows by choosing $r < \frac{4\gamma-1}{2(2\gamma+1)}$.   □

We will now use an induction argument 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. Whilst the gist of the argument is standard, see [3,19, 28], a technical difficulty arises from the divergence of the derivative of $m(t)$ at $t = 0$. To surmount this difficulty, we take inspiration from [14] and introduce a (small) parameter $\alpha > 0$ to be specified at the end of the iteration. Then, by Lemma 2.5, we have

$$\left\| \frac{dm_\cdot}{dt} \right\|_{\varepsilon^\alpha, T/2, \infty} \leq M \varepsilon^{\alpha(H-1)}. \tag{4.10}$$

Instead of considering supremums over the time interval $[0, T/2]$, we will now consider supremums over $[\varepsilon^\alpha, T/2]$ instead, thus avoiding the singularity at zero.

In order to formalise this, for $r > 0$ and $i \geq 1$, define the events

$$\mathscr{K}_i^\varepsilon(\alpha, r) = \big\{\big\|\langle\varphi, Z_U(\cdot)\rangle\big\|_{\varepsilon^\alpha, T/2, \infty} < \varepsilon^r, \ \forall U \in \mathcal{V}_i\big\}.$$

With this notation at hand, the recursion step in our argument can be formulated as follows:

Lemma 4.10. *There exists $\bar{q} > 0$ such that, for every $\alpha > 0$ and every $i \geq 1$, the implication*

$$\mathscr{K}_i^\varepsilon(\alpha, r) \quad \Longrightarrow \quad \mathscr{K}_{i+1}^\varepsilon\big(\alpha, \bar{q}r + \alpha(H - 1)\big)$$

holds modulo negligible events.

Proof. Assume that $\mathscr{K}_i^\varepsilon(\alpha, r)$ holds. For any $t \leq T/2$, it then follows from the chain rule that

$$\langle\varphi, Z_U(t) - Z_U(\varepsilon^\alpha)\rangle = \int_{\varepsilon^\alpha}^t \langle\varphi, J_{0,s}^{-1}[V_0, U](X_s)\rangle \, \mathrm{d}s$$
$$+ \sum_{i=1}^d \int_{\varepsilon^\alpha}^t \langle\varphi, J_{0,s}^{-1}[V_i, U](X_s)\rangle \, \mathrm{d}m_s^i + \sum_{i=1}^d \int_{\varepsilon^\alpha}^t \langle\varphi, J_{0,s}^{-1}[V_i, U](X_s)\rangle \circ \mathrm{d}\tilde{B}_s^i, \tag{4.11}$$

where $\tilde{B}_\cdot$ is as in (2.8).

Now by the hypothesis, Proposition 4.4 and (4.11), it follows that for any $U \in \mathcal{V}_i$,

$$\left\|\left\langle\varphi, \sum_{i=1}^d J_{0,\cdot}^{-1}[V_i, U](X_\cdot)\right\rangle\right\|_{\varepsilon^\alpha, T/2, \infty} < M\varepsilon^{rq}, \tag{4.12a}$$

$$\left\|\left\langle\varphi, J_{0,\cdot}^{-1}[V_0, U](X_\cdot) + \sum_{i=1}^d \frac{\mathrm{d}m_\cdot^i}{\mathrm{d}s} J_{0,\cdot}^{-1}[V_i, U](X_\cdot)\right\rangle\right\|_{\varepsilon^\alpha, T/2, \infty} < M\varepsilon^{rq} \tag{4.12b}$$

modulo negligible events, with q as given by Proposition 4.4. Plugging (4.10) and (4.12a) into (4.12b), it follows that for every $U \in \mathcal{V}_i$,

$$\big\|\langle\varphi, J_{0,\cdot}^{-1}[V_0, U](X_\cdot)\rangle\big\|_{\varepsilon^\alpha, T/2, \infty} < M\varepsilon^{rq + \alpha(H-1)}$$

modulo negligible events. Choosing any $\bar{q} < q$, the claim then follows from the fact that for every exponent $\delta > 0$, one has $M\varepsilon^\delta < 1$ modulo negligible events. (Recall that $H < 1$, so that (4.12a) is actually slightly better than the claimed bound.) $\qquad\square$

Now we iterate the previous argument. Let $N \in \mathbb{N}$ be such that

$$\inf_{|x| \leq R} \inf_{|\xi|=1} \sum_{U \in \bar{\mathcal{V}}_N} \big|\langle\xi, U(x)\rangle\big|^2 > 0. \tag{4.13}$$

It follows from the Hörmander condition and the fact that the vector fields are smooth that such an N exists. From Lemmas 4.9 and 4.10, we obtain that for every $U \in \bar{\mathcal{V}}_N$, modulo negligible events,

$$\big\|\langle\varphi, J_{0,\cdot}^{-1}U(X_\cdot)\rangle\big\|_{\varepsilon^\alpha, T/2, \infty} < \varepsilon^{q_1}, \tag{4.14}$$

where $q_1 = r\bar{q}^N + \alpha(H-1)\frac{(1-(r\bar{q})^N)}{1-r\bar{q}}$. Since $\bar{q}$ and r are fixed, we can choose α small enough such that $q_1 > 0$.

Now we make use of the Hölder continuity of $J_{0,\cdot}^{-1}$ and X. Indeed, it follows from Lemma 3.4 that

$$\left\| J_{0,\cdot}^{-1} U(X_\cdot) - U(x_0) \right\|_{0,\varepsilon^\alpha,\infty} < \varepsilon^{q_2}, \tag{4.15}$$

modulo negligible events, for any $q_2 < \alpha\gamma$. Combining (4.14) and (4.15), we conclude that

$$\sup_{U \in \tilde{\mathcal{V}}_N} \left| \langle \varphi, U(x_0) \rangle \right| < \varepsilon^{q_1} + \varepsilon^{q_2},$$

modulo negligible events. This is in direct contradiction with (4.13), thus concluding the proof of Theorem 4.7. □

We conclude that the matrix $\tilde{C}_T$ defined in (4.4) is almost surely invertible and $\tilde{\mathbb{E}}(\|\tilde{C}_T\|^{-p}) < M$ for $p \geq 1$. As a consequence of Theorem 4.2, we obtain the smoothness of the laws of X_t conditional on a given past $\omega \in \mathcal{W}$:

Theorem 4.11. *Let $\varphi_t(x, \tilde{\omega})$ be the solution to* (1.1) *with notations as in Section* 2.1, *and assume that Assumption* 4.1 *holds. Then, for every $\omega \in \mathcal{W}$, the distribution of $\varphi_t(x, \tilde{\omega})$ under $\mathcal{P}(\omega, \cdot)$ has a smooth density with respect to Lebesgue measure.*

Proof. In order to obtain the smoothness of the densities, it suffices by [3,23,30,31] to show that the Malliavin matrix $\mathcal{M}_T$ of the map $W \mapsto X_T^x$, where W is the standard Wiener process describing the solution as in (2.8), is invertible with negative moments of all orders. As we will see in Section 5 below (see Eq. (5.3)), $\mathcal{M}_T$ is given by

$$\langle \xi, \mathcal{M}_T \xi \rangle = \left\| \left(\mathcal{I}^{H-1/2} \right)^* \mathcal{A}^* J_{0,T}^* \xi \right\|_{L^2}^2,$$

where the L^2-adjoint of the fractional integral operator $\mathcal{I}^{H-1/2}$ is given by

$$\left(\left(\mathcal{I}^{H-1/2} \right)^* g \right)(t) = \int_t^T (s-t)^{H-3/2} g(s) \, ds.$$

We already know from Theorem 4.2 that $\|\mathcal{A}^* J_{0,T}^* \xi\|_{L^2}^2 \geq \varepsilon$ modulo negligible events and we would like to make use of this information in order to obtain a lower bound on $\mathcal{M}_T$.

Since $\mathcal{I}^{H-1/2}$ and $\mathcal{D}^{H-1/2}$ are inverses of each other, for any measurable f we have,

$$\|f\|_{L^2}^2 = \left\langle f, \mathcal{I}^{H-1/2} \mathcal{D}^{H-1/2} f \right\rangle_{L^2} \leq \left\| \left(\mathcal{I}^{H-1/2} \right)^* f \right\|_{L^2} \left\| \mathcal{D}^{H-1/2} f \right\|_{L^2}$$

$$\leq M \left\| \left(\mathcal{I}^{H-1/2} \right)^* f \right\|_{L^2} \|f\|_{0,T,\gamma}.$$

Thus we obtain,

$$\left\| \left(\mathcal{I}^{H-1/2} \right)^* \mathcal{A}^* J_{0,T}^* \xi \right\|_{L^2} \geq M \left\| \mathcal{A}^* J_{0,T}^* \xi \right\|_{L^2}^2 \left\| \mathcal{A}^* J_{0,T}^* \xi \right\|_{0,T,\gamma}^{-1}.$$

Since we know from the results in Section 3 that $\|\mathcal{A}^* J_{0,T} \xi\|_{0,T,\gamma} \leq \varepsilon^{-1}$ modulo negligible events, we conclude that

$$\left\| \left(\mathcal{I}^{H-1/2} \right)^* \mathcal{A}^* J_{0,T}^* \xi \right\|_{L^2} \geq \varepsilon \left\| \mathcal{A}^* J_{0,T}^* \xi \right\|_{L^2}^2,$$

holds modulo negligible events. The proof now follows from Theorem 4.2. □

5. Strong Feller property

Recall from the arguments in Section 2.2 that the strong Feller property (2.2) is the only ingredient required for the proof of Theorem 1.3. We will actually show something slightly stronger than just (2.2). Fix some arbitrary

final time $T > 1$, a Fréchet differentiable bounded map $\psi : \mathcal{C}([1, T], \mathbb{R}^n) \to \mathbb{R}$ with bounded derivative,[2] denote by $R_{1,T} : \mathcal{C}(\mathbb{R}_+, \mathbb{R}^n) \to \mathcal{C}([1, T], \mathbb{R}^n)$ the restriction operator, and set as before

$$\bar{\mathcal{Q}}\psi(x, \omega) \stackrel{\text{def}}{=} \int_{\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)} \psi(R_{1,T}z)\bar{\mathcal{Q}}(x, \omega; dz).$$

With these notations, the main result of this article is:

Theorem 5.1. *Assume that the vector fields $\{V_i\}_{i \geq 0}$ have bounded derivatives of all orders and that furthermore the vector fields $\{V_i\}_{i \geq 1}$ are bounded. Then, if Hörmander's bracket condition (Assumption 4.1) holds, there exists a continuous function $K : \mathbb{R}^n \to \mathbb{R}_+$ and a constant $M > 0$ such that for $\omega \in \mathcal{W}_{\gamma,\delta}$ and $x \in \mathbb{R}^n$, the bound*

$$\left| D_x \bar{\mathcal{Q}}\psi(x, \omega) \right| \leq K(x)e^{M\|\mathcal{G}\omega\|_{0,T,\gamma}^{1/\gamma}} \tag{5.1}$$

holds uniformly over all test functions ψ as above with $\|\psi\|_\infty \leq 1$. Here, $\mathcal{G}$ was defined in (2.6) and both K and M are independent of T and of ψ. In particular, (1.1) is strong Feller.

The remainder of this section is devoted to the proof of this result. For conciseness, from now onwards we will use the shorthand notation

$$\left| \Psi(x, \omega) \right| < \mathscr{E}(x, \omega),$$

to indicate that there exist a continuous function K on $\mathbb{R}^n$ and a constant M such that the expression $\Psi(x, \omega)$ is bounded by the right-hand side of (5.1).

To obtain the bound for $D\bar{\mathcal{Q}}\psi(\cdot, \omega)$, we use a 'stochastic control' argument which is inspired by [13,14] and is similar in spirit to the arguments based on the Bismut–Elworthy–Li formula [7,9]. We will explain this argument in more detail in Section 5.1 below, but before that let us introduce the elements of Malliavin calculus required for the construction.

Let X_t^x and $J_{0,t}$ be the solution and its Jacobian, as defined in equations (3.2a) and (3.2b) respectively with initial condition $X_0^x = x$. Let $\varphi_t(x, \tilde{\omega}) = X_t^x$ denote the flow map as in Section 2.1. Henceforth we will use $\varphi_t(x, \tilde{\omega})$ and X_t^x interchangeably, with a preference for the first notation when we want to emphasise the dependence of the solution either on the realisation of the noise or on its initial condition.

By the variation of constants formula, the Fréchet derivative of the flow map *with respect to the driving noise $\tilde{\omega}$ in* the direction of $\int_0^\cdot v(s) \, ds$ is then given by

$$D_v X_t^x = D_v \varphi_t(x, \tilde{\omega}) = J_{0,t}\mathcal{A}_t v, \tag{5.2}$$

where $\mathcal{A}_t$ is as defined in (4.5).

Remark 5.2. *Notice that the operator D defined above is a standard Fréchet derivative unlike in the case $H = \frac{1}{2}$ where it would have to be interpreted as a Malliavin derivative. This is because the integration with respect to the fBM for $H > \frac{1}{2}$ is just the Reimann–Stieltjes integral, whereas for $H = \frac{1}{2}$ the stochastic integral is not even a continuous (let alone differentiable) map of the noise in the case of multiplicative noise.*

As already mentioned in Remark 2.4, the operator $\mathcal{D}^{H-1/2}$ and its inverse $\mathcal{I}^{H-1/2}$ provide an isometry between the Cameron–Martin space of the Gaussian measure $\mathcal{P}(\omega, \cdot)$ and that of Wiener measure.

As a consequence, if v is such that $\tilde{v} = \mathcal{D}^{H-1/2}v \in L^2(\mathbb{R}_+, \mathbb{R}^d)$, then we have the identity

$$\mathbf{D}_{\tilde{v}} X_t^x = D_v \varphi_t(x, \tilde{\omega}) = J_{0,t}\mathcal{A}_t v = J_{0,t}\mathcal{A}_t \mathcal{I}^{H-1/2}\tilde{v}, \tag{5.3}$$

where $\mathbf{D}$ is the Malliavin derivative of X_t^x when interpreted as a function on Wiener space via the representation (2.8).

---

[2]The space $\mathcal{C}([1, T], \mathbb{R}^n)$ is endowed with the usual supremum norm.

Before we continue, let us make a short digression on the definition of the Malliavin derivative that appears in the above expression. In general, if $Z : \widetilde{\mathcal{W}} \to \mathbb{R}$ is any Fréchet differentiable function, then its Malliavin derivative with respect to the underlying Wiener process is given by

$$\mathbf{D}_{\tilde{v}} Z \stackrel{\text{def}}{=} D_v Z, \tag{5.4}$$

where $D_v Z$ denotes the Fréchet derivative of Z in the direction $\int_0^{\cdot} v(t) \, dt$ and where $\tilde{v}$ and v are related by $\tilde{v} = \mathcal{D}^{H-1/2} v$ as before.

Since, for every $T > 0$, the map $\mathcal{D}^{H-1/2}$ restricted to functions on $[0, T]$ is bounded from $\widetilde{\mathcal{W}}$ to $\mathcal{C}([0, T], \mathbb{R}^d)$ and since the dual of the latter space consists of finite signed measures, we conclude from (5.4) that if Z is Fréchet differentiable, then there exists a function $s \mapsto \mathbf{D}_s Z$ which is locally of bounded variation and such that the Malliavin derivative of Z can be represented as

$$\mathbf{D}_{\tilde{v}} Z = \int_0^{\infty} \langle \mathbf{D}_s Z, \tilde{v}(s) \rangle \, ds.$$

(The scalar product is that of $\mathbb{R}^d$ here.) The quantity $\mathbf{D}_s Z$ should be interpreted as the variation of Z with respect to the infinitesimal increment of the underlying Wiener process W at time s.

With this convention, it follows from (5.3) and the definition of $\mathcal{I}^{H-1/2}$ that for some constant c_H one has the identity

$$\mathbf{D}_s X_t^x = c_H \int_s^t J_{r,s} V(X_r^x)(r - t)^{H-3/2} \, dr \tag{5.5}$$

for $s \leq t$. Here, V denotes the matrix $(V_1, \ldots, V_d)$ as before. Note also that $\mathbf{D}_s X_t^x = 0$ for $s > t$ since X is an adapted process. Of course, the whole point of Malliavin calculus is to also be able to deal with functions that are not Fréchet differentiable, but for our purposes the framework discussed here will suffice.

5.1. *Stochastic control argument*

At the intuitive level, the central idea of the stochastic control argument is as follows. In a nutshell, the strong Feller property states that for two nearby initial conditions x_0 and y_0 and an arbitrary realisation ω of the *past* of the driving noise, it is possible to construct a coupling between the solutions x_t and y_t such that with high probability (tending to 1 as $y \to x$), one has $x_t = y_t$ for $t \geq 1$. One way of achieving such a coupling is to perform a change of measure on the driving process for one of the two solutions (say y) in such a way, that the noise $\tilde{\omega}_y$ driving y is related to the noise $\tilde{\omega}_x$ driving x by

$$d\tilde{\omega}_y(t) = d\tilde{\omega}_x(t) + v_{x,y}(t) \, dt,$$

where v is a 'control' process that aims at steering the solution y towards the solution x. If one takes $y = x + \varepsilon \xi$ and looks for controls of the form $v_{x,y} = \varepsilon v$, then in the limit $\varepsilon \to 0$, the effect of the control v onto the solution x after time t is given by (5.3), while the effect of shifting the initial condition by ξ is given by $J_{0,t}\xi$. It is therefore not surprising that in order to prove the strong Feller property, our main task will be to find a control v such that $J_{0,1}\mathcal{A}_1 v = J_{0,1}\xi$, or equivalently $\mathcal{A}_1 v = \xi$. This is the content of the following proposition:

Proposition 5.3. *Assume that, for every $x \in \mathbb{R}^n$ and $\omega \in \mathcal{W}$, there exists a stochastic processes $v \in \mathcal{C}^\gamma(\mathbb{R}_+, \mathbb{R}^d)$ such that the identity*

$$\mathcal{A} v = \xi$$

holds almost surely, where $\mathcal{A} \stackrel{\text{def}}{=} \mathcal{A}_1$ defined in (4.5). Assume furthermore that the map $\tilde{w} \mapsto v$ is Fréchet differentiable from $\widetilde{\mathcal{W}}$ to $\mathcal{C}^\gamma(\mathbb{R}_+, \mathbb{R}^d)$. Finally, we assume that the Malliavin derivative of the process $\tilde{v} \stackrel{\text{def}}{=} \mathcal{D}^{H-1/2} v$ satisfies the bounds

$$M_1 = \tilde{\mathbb{E}} \int_0^{\infty} |\tilde{v}(s)|^2 \, ds < \infty, \qquad M_2 = \tilde{\mathbb{E}} \int_0^{\infty} \int_0^{\infty} |\mathbf{D}_t \tilde{v}(s)|^2 \, ds \, dt < \infty. \tag{5.6}$$

Then, the bound

$$\left|\langle D_x \bar{\mathcal{Q}} \psi(x, \omega), \xi \rangle\right| \leq \|\psi\|_\infty \sqrt{M_1 + M_2}$$

holds uniformly over all test functions ψ as in Theorem 5.1, all $T > 1$, and all $\xi \in \mathbb{R}^n$.

In particular, if v is such that $M_1 + M_2 < \mathcal{E}(x, \omega)$, then the conclusions of Theorem 5.1 hold.

Remark 5.4. *Since $\mathcal{D}^{H-1/2}$ is bounded from $\mathcal{C}^\gamma(\mathbb{R}_+, \mathbb{R}^d)$ into $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$, the Malliavin derivative of $\tilde{v}$ exists, so that the expression in (5.6) makes sense.*

Proof of Proposition 5.3. 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 $\mathcal{A}v = \xi$, where we use the shorthand notation $\mathcal{A} = \mathcal{A}_1$. If this can be achieved, then we extend v to all of $\mathbb{R}_+$ by setting $v(s) = 0$ for $s \geq 1$ and we define $\tilde{v} = \mathcal{D}^{H-1/2}v$. Note that since $v(s) = 0$ for $s > 1$, it follows from the definition of $\mathcal{A}_t$ that we have $\mathcal{A}_t v = \mathcal{A}v = \xi$ for $t \geq 1$. If v is sufficiently regular in time so that this definition makes sense, we thus have the identity

$$\mathbf{D}_{\tilde{v}} X_T^x = J_{0,T} \xi \tag{5.7}$$

for every $T \geq 1$. In particular, for a $\mathcal{C}^1$ test function ψ as above and for $\xi \in \mathbb{R}^n$,

$$\langle D_x \bar{\mathcal{Q}} \psi(x, \omega), \xi \rangle = \tilde{\mathbb{E}}\big((D\psi)\big(\varphi(x, \tilde{\omega})\big) J_{0,.} \xi\big)$$

$$= \tilde{\mathbb{E}}\big((D\psi)\big(\varphi(x, \tilde{\omega})\big) \mathbf{D}_{\tilde{v}} X_.\big) = \tilde{\mathbb{E}}\big(\mathbf{D}_{\tilde{v}}\big(\psi\big(\varphi(x, \tilde{\omega})\big)\big)\big).$$

It then follows from the integration by parts formula of Malliavin calculus [31] that this quantity is equal to

$$\cdots = \tilde{\mathbb{E}}\bigg(\psi\big(\varphi_T(x, \tilde{\omega})\big) \int_0^T \langle \tilde{v}(s), \, \mathrm{d}W_s \rangle\bigg) \leq \|\psi\|_\infty \tilde{\mathbb{E}}\bigg|\int_0^T \langle \tilde{v}(s), \, \mathrm{d}W_s \rangle\bigg|.$$

The stochastic integral appearing in this expression is in general not an Itô integral, but a Skorohod integral, since its integrand is in general not adapted to the filtration generated by W.

For Skorohod integrals one has the following extension of Itô's isometry [31]:

$$\tilde{\mathbb{E}}\bigg(\int_0^\infty \langle \tilde{v}(s), \, \mathrm{d}W_s \rangle\bigg)^2 = \tilde{\mathbb{E}}\bigg(\int_0^\infty |\tilde{v}(s)|^2 \, \mathrm{d}s\bigg) + \tilde{\mathbb{E}} \int_0^\infty \int_0^\infty \mathrm{tr}\big(\mathbf{D}_t \tilde{v}(s)^T \mathbf{D}_s \tilde{v}(t)\big) \, \mathrm{d}s \, \mathrm{d}t$$

$$\leq M_1 + M_2, \tag{5.8}$$

where as before $\mathbf{D}$ denotes Malliavin derivative with respect to the underlying Wiener process. The claim then follows at once. $\qquad \square$

In order to have the identity (5.1), it therefore remains to find a control v satisfying $\mathcal{A}v = \xi$ and such that

$$M_1 + M_2 < \mathcal{E}(x, \omega), \tag{5.9}$$

where the quantities M_1 and M_2 are defined by (5.6) in terms of $\tilde{v} = \mathcal{D}^{H-1/2}v$. A natural candidate for such a control v can be identified by a simple least squares formula. Indeed, note that the adjoint $\mathcal{A}^* : \mathbb{R}^n \mapsto L^2(\mathbb{R}_+, \mathbb{R}^d)$ of $\mathcal{A}$ is given by

$$\big(\mathcal{A}^* \xi\big)(s) = V(X_s)^* \big(J_{0,s}^{-1}\big)^* \xi, \quad \xi \in \mathbb{R}^n. \tag{5.10}$$

Let $h = h_1$ where $h_1 : [0, 1] \to \mathbb{R}_+$ is as defined in (4.6). Assuming that $\mathcal{A}h\mathcal{A}^*$ is invertible, one possible solution to the equation $\mathcal{A}v = \xi$ is given by

$$v(s) \equiv h(s)\big(\mathcal{A}^*\big(\mathcal{A}h\mathcal{A}^*\big)^{-1}\xi\big)(s). \tag{5.11}$$

Note that $\mathcal{A}h\mathcal{A}^*$ is precisely equal to the 'modified Malliavin matrix' C_1 defined in (4.7). By the results from the previous section the matrix C_1 is almost surely invertible and therefore the expression (5.11) does make sense.

The aim of the next section is to show that the choice (5.11) does indeed satisfy the assumptions of Proposition 5.3.

Remark 5.5. *For the identity* (5.3) *to hold, the stochastic control* $v(\cdot)$ *needs to belong to the Cameron–Martin space of* $\mathcal{P}(0, \cdot)$, *and the function* $h(\cdot)$ *is introduced just for this purpose. Indeed, since we extend the control* $v(\cdot)$ *to vanish outside the interval* $[0, 1]$, *one needs to be careful about the regularity of* $v(s)$ *at* $s = 1$. *It can be shown using the fractional integrals that, in order for* $v(\cdot)$ *to have the required regularity, the function* $h(s)$ *should be* $\mathcal{O}((1 - s)^H)$ *for* $s \approx 1$. *The specific* $h_1(\cdot)$ *we chose in* (4.6), *of course, satisfies this.*

5.2. *Proof of the main result, Theorem* 5.1

In view of Proposition 5.3, the proof of Theorem 5.1 is complete if we can show that the stochastic process v defined by (5.11) satisfies the bounds (5.9). This will be the content of Propositions 5.8 and 5.11 below.

We start with an estimate on the Hölder norm of v which will be used repeatedly. Here and in the rest of this section, we will make use of the notation

$$\|v\|_\gamma \stackrel{\text{def}}{=} \|v\|_{0,1,\gamma} + \|v\|_{0,1,\infty}$$

for the Hölder norm of v. This has the advantage of being a norm rather than just a seminorm and, for any two Hölder continuous functions u and v, we have the bound

$$\|uv\|_\gamma \leq M\|u\|_\gamma\|v\|_\gamma$$

for some fixed constant $M > 0$. With this notation, our bound is:

Lemma 5.6. *The stochastic control* v *defined in* (5.11) *is continuous and satisfies the bound*

$$\tilde{\mathbb{E}}\|v\|_\gamma^2 < \mathscr{E}(x, \omega).$$

Proof. Setting $v(s) = h(s)u(s)$, we obtain from (5.11) the bound

$$\|u\|_\gamma^2 \leq \|\mathcal{A}^*\eta\|_\gamma^2, \tag{5.12}$$

where $\eta = (\mathcal{A}h\mathcal{A}^*)^{-1}\xi$. From (5.10) and the fact that V has bounded derivatives it follows that

$$\left\|\mathcal{A}^*\eta\right\|_\gamma \leq M\left\|V(X_\cdot)\right\|_\gamma\left\|J_{0,\cdot}^{-1}\right\|_\gamma|\eta| \leq M\|X_\cdot\|_\gamma\left\|J_{0,\cdot}^{-1}\right\|_\gamma|\eta|,$$

and therefore by Cauchy–Schwarz

$$\left(\tilde{\mathbb{E}}\left\|\mathcal{A}^*\eta\right\|_\gamma^2\right)^2 \leq M\tilde{\mathbb{E}}\left(\|X_\cdot\|_\gamma^4\left\|J_{0,\cdot}^{-1}\right\|_\gamma^4\right)\tilde{\mathbb{E}}|\eta|^4.$$

From Lemma 3.4 we deduce that $\tilde{\mathbb{E}}(\|X_\cdot\|_\gamma^4\|J_{0,\cdot}^{-1}\|_\gamma^4) < \mathscr{E}(x, \omega)$. Also from Theorem 4.2, we have $\tilde{\mathbb{E}}|(\mathcal{A}h\mathcal{A}^*)^{-1}\xi|^p < \mathscr{E}(x, \omega)$ for any $p \geq 1$. Thus we obtain

$$\tilde{\mathbb{E}}\|u\|_\gamma^2 \leq \mathscr{E}(x, \omega).$$

The claim then follows immediately from $v = hu$. $\qquad\square$

Remark 5.7. *Had we not introduced the control* h, *we could still obtain continuity of* v *on* $[0, 1]$, *but we would have a discontinuity at* $t = 1$.

As a consequence, we show that $\tilde{v} = \mathcal{D}^{H-1/2}v$ is square integrable:

Proposition 5.8. *For the stochastic control $v(\cdot)$ defined in* (5.11) *and the operator $\mathcal{D}$ defined in* (2.4), *the following estimate holds*:

$$\tilde{\mathbb{E}}\left(\int_0^\infty |\tilde{v}(s)|^2 \, ds\right) < \mathscr{E}(x, \omega).$$

Before we turn to the proof of Proposition 5.8, we state the following useful result:

Lemma 5.9. *For any $s \geq 0, \alpha > 0$ and $f \in \mathcal{C}^\gamma$ with $\gamma > \alpha$,*

$$\Gamma(1 - \alpha)\mathcal{D}^\alpha f(s) = \frac{d}{ds} \int_0^s (s - r)^{-\alpha} f(r) \, dr$$

$$= s^{-\alpha} f(s) - \alpha \int_0^s (s - r)^{-\alpha - 1}\big(f(r) - f(s)\big) \, dr. \tag{5.13}$$

Proof. See [32], Theorem 13.1, as well as Lemma 13.1 in the same monograph. $\square$

Remark 5.10. *The right-hand side of Eq.* (5.13) *is called the Marchaud fractional derivative. Lemma* 5.9 *actually holds true for a larger class of functions f, namely those such that $f = \mathcal{I}^\alpha g$ for some $g \in L^1([0, 1], \mathbb{R})$.*

Proof of Proposition 5.8. The main idea behind this proof, and a recurrent theme in this section is the following. Although the stochastic control $v(\cdot)$ vanishes outside $[0, 1]$, the integrand $\tilde{v}(s)$ in non-zero for all $s \in [0, \infty)$. Thus the square integrability of $\tilde{v}(s)$ on the interval $[0, \infty)$ is shown by dealing with the singularities of the integrand differently at $s = 0$, and $s = \infty$. At $s = 0$, we use the fact that v is Hölder continuous whereas at $s = \infty$, we use the fact that v is bounded.

By definition,

$$\tilde{v}(s) = \frac{d}{ds} \int_0^s (s - r)^{1/2 - H} v(r) \, dr.$$

Applying Lemma 5.9 with $\alpha = H - \frac{1}{2}$, we obtain the bound

$$\big|\tilde{v}(s)\big| \leq M s^{1/2 - H} \big|v(s)\big| + M \int_0^s (s - r)^{-1/2 - H} \big|v(r) - v(s)\big| \, dr$$

for some constant $M > 0$. Noting that the control $v(\cdot)$ vanishes outside the interval $[0, 1]$, we have

$$\tilde{\mathbb{E}}\left(\int_0^\infty |\tilde{v}(s)|^2 \, ds\right) \leq M \tilde{\mathbb{E}}\|v\|_\gamma^2 \int_0^1 s^{1 - 2H} \, ds$$

$$+ M \tilde{\mathbb{E}} \int_0^\infty \left(\int_0^s (s - r)^{-1/2 - H} \big|v(r) - v(s)\big| \, dr\right)^2 \, ds. \tag{5.14}$$

The first integral $\int_0^1 s^{1 - 2H} \, ds$ in Eq. (5.14) is clearly finite since $H < 1$, so that this term is bounded by $\mathscr{E}(x, \omega)$ by Lemma 5.6. Now we show that the second term has a similar bound. To this end, set

$$\zeta(s) \equiv \int_0^s (s - r)^{-1/2 - H} \big|v(r) - v(s)\big| \, dr.$$

For $s < 1$, we obtain

$$\big|\zeta(s)\big| = \int_0^s (s - r)^{-1/2 - H} \big|v(r) - v(s)\big| \, dr \leq M \|v\|_\gamma \int_0^s s^{-1/2 - H + \gamma} \, ds$$

$$\leq M \|v\|_\gamma s^{1/2 - H + \gamma},$$

so that the required bound follows again from Lemma 5.6. For $s \geq 1$, we observe that $\zeta(s)$ has a square integrable singularity, and thus the boundedness of the control $v(\cdot)$ will be sufficient. Indeed it follows from the definition of h (4.6) that

$$\left|\zeta(s)\right| \leq \int_0^{3/4} (s-3/4)^{-1/2-H} \left|v(r) - v(s)\right| dr \leq M\|v\|_\gamma (s-3/4)^{-1/2-H}.$$

Since this expression is square integrable, it follows that $\tilde{\mathbb{E}} \int_0^\infty |\zeta^2(s)| \, ds < \mathscr{E}(x, \omega)$, and the claim follows from (5.14). $\qquad\square$

We conclude with the corresponding bound on the Malliavin derivatives of the control v:

Proposition 5.11. *Let v be as defined in* (5.11), *the operator $\mathcal{D}$ as defined in* (2.4) *and $\mathbf{D}$ be the Malliavin derivative. Then following estimate holds*:

$$\tilde{\mathbb{E}}\left(\int_0^\infty \int_0^\infty \left\|\mathbf{D}_t \tilde{v}(s)\right\|^2 ds \, dt\right) < \mathscr{E}(x, \omega). \tag{5.15}$$

Proof. The main difficulty here is to obtain control over the modulus of continuity of the quantity $\mathbf{D}_s v(t)$ for all $s, t \in [0, 1]$. On route to establishing the Hölder continuity of the Malliavin derivative applied to the stochastic control, we need to perform several careful estimates using the fractional derivatives.

First notice that, since $\tilde{v}(s) \in \mathcal{F}_1$ for every $s > 0$ ($\mathcal{F}_t, t \geq 0$ denotes the filtration generated by the increments of W), we have $\mathbf{D}_t \tilde{v}(s) = 0$ for $t > 1$, so that the integrand in (5.15) vanishes outside the set $[0, 1] \times [0, 1]$. Furthermore, it is straightforward to check that $\mathbf{D}_t$ and $\mathcal{D}^{H-1/2}$ commute, so that

$$\tilde{\mathbb{E}}\left(\int_0^\infty \int_0^\infty \left\|\mathbf{D}_t \tilde{v}(s)\right\|^2 ds \, dt\right) \leq \tilde{\mathbb{E}}\left(\int_0^1 \int_0^1 \left\|\left(\mathcal{D}^{H-1/2}\mathbf{D}_t v\right)(s)\right\|^2 dt \, ds\right).$$

By Lemma 5.9, the claim follows if we can show that

$$\tilde{\mathbb{E}}\left(\int_0^1 \int_0^1 \left|\int_0^s (s-r)^{-1/2-H}\left(\mathbf{D}_t v(s) - \mathbf{D}_t v(r)\right) dr\right|^2 dt \, ds\right) < \mathscr{E}(x, \omega). \tag{5.16}$$

Expanding the integrand, we obtain

$$\left|\mathbf{D}_t\left(\int_0^s (s-r)^{-1/2-H}\left(v(s) - v(r)\right) dr\right)\right|$$

$$\leq \int_0^t (s-r)^{-1/2-H}\left|\mathbf{D}_t v(s)\right| dr + \left|\mathbf{D}_t\left(\int_t^s (s-r)^{-1/2-H}\left(v(s) - v(r)\right) dr\right)\right|$$

$$\leq \left|s^{1/2-H} - (s-t)^{1/2-H}\right|\|\mathbf{D}_t v\|_{0,1,\infty}$$

$$+ \left|\int_t^s (s-r)^{-1/2-H}\mathbf{D}_t\left(v(s) - v(r)\right) dr\right|. \tag{5.17}$$

Since the double integral of $|s^{1/2-H} - (s-t)^{1/2-H}|^2$ is finite, the first term will be bounded by $\mathscr{E}(x, \omega)$ if we can show that $\tilde{\mathbb{E}} \sup_{t \in [0,1]} \|\mathbf{D}_t v\|_{0,1,\infty}^2 < \mathscr{E}(x, \omega)$. Since this bound will be a byproduct of the analysis of the second term, we postpone it to (5.23) below.

The second term of equation (5.17) needs more delicate calculations. We will show that there exists a random variable K and an exponent $\alpha < 1 + \gamma - H$ such that the following bounds hold:

$$\left|\mathbf{D}_t v(s) - \mathbf{D}_t v(r)\right| \leq K|s-r|^\gamma |r-t|^{-\alpha}, \quad \text{with } \tilde{\mathbb{E}}K^2 < \mathscr{E}(x, \omega). \tag{5.18}$$

Assuming these bounds, we proceed similarly to the proof of Proposition 5.8:

$$\tilde{\mathbb{E}}\left(\int_0^1\int_0^1\left|\int_t^s (s-r)^{-1/2-H}\mathbf{D}_t\big(v(s)-v(r)\big)\,\mathrm{d}r\right|^2\mathrm{d}t\,\mathrm{d}s\right)$$

$$\leq \tilde{\mathbb{E}}\left(K^2\int_0^1\int_0^1\left|\int_t^s (s-r)^{-1/2-H+\gamma}(r-t)^{-\alpha}\,\mathrm{d}r\right|^2\mathrm{d}t\,\mathrm{d}s\right)$$

$$\leq \tilde{\mathbb{E}}\left(K^2\int_0^1\int_0^1 (s-t)^{1+2\gamma-2H-2\alpha}\,\mathrm{d}t\,\mathrm{d}s\right) < \mathscr{E}(x,\omega),$$

where we used the bound on α to obtain the last inequality. It therefore remains to show (5.18).

The Malliavin derivative of v is given by

$$\mathbf{D}_t v(s) = h(s)\mathbf{D}_t\big(V(X_s)^*\big(J_{0,s}^{-1}\big)^*\big(\mathcal{A}h\mathcal{A}^*\big)^{-1}\xi\big).$$

Applying the product rule, we break this into a sum of three terms:

$$\mathbf{D}_t v(s) = h(s)\big[\mathbf{D}_t\big(V(X_s)^*\big)\big(J_{0,s}^{-1}\big)^*\big(\mathcal{A}h\mathcal{A}^*\big)^{-1}\xi + V(X_s)^*\mathbf{D}_t\big(\big(J_{0,s}^{-1}\big)^*\big)\big(\mathcal{A}h\mathcal{A}^*\big)^{-1}\xi$$

$$+ V(X_s)^*\big(J_{0,s}^{-1}\big)^*\mathbf{D}_t\big(\big(\mathcal{A}h\mathcal{A}^*\big)^{-1}\xi\big)\big]$$

$$\stackrel{\text{def}}{=} \mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3. \tag{5.19}$$

To make the proof compact, we introduce the stochastic process Y_s given by collecting the various objects appearing in this expression:

$$Y_s = \big(X_s, V(X_s), \big(J_{0,s}^{-1}\big)^*\big). \tag{5.20}$$

It the follows from (1.1) and the expression (3.2b) for the Jacobian $J_{0,t}$ that the components of Y_s solve equations of the form (3.5). One can check that both the Jacobian $\mathcal{J}_{0,s}$ for Y and its inverse $\mathcal{J}_{0,s}^{-1}$ then satisfy equations of the type (3.14). So by Corollary 3.3, we obtain

$$\|Y\|_\gamma + \|\mathcal{J}_{0,\cdot}\|_\gamma + \big\|\mathcal{J}_{0,\cdot}^{-1}\big\|_\gamma \leq K(x_0)\exp\big(M\|\tilde{B}+\mathcal{G}\omega\|_\gamma^{1/\gamma}\big) \tag{5.21}$$

for some constant M and some continuous function K.

Furthermore, just as in (5.5), the Malliavin derivative of Y is given by

$$\mathbf{D}_t Y_s = c_H\int_t^s \mathcal{J}_{a,s}U(Y_a)(a-t)^{H-3/2}\,\mathrm{d}a\,1_{t\leq s} \tag{5.22}$$

for some smooth function U which grows linearly at infinity.

We now tackle the bounds on the two very similar terms $\mathcal{I}_1$, $\mathcal{I}_2$. Since $t^{H-3/2}$ is integrable at 0 and since U grows linearly at infinity, it follows immediately that

$$\sup_{s,t\in[0,1]}|\mathbf{D}_t Y_s| \leq M\|\mathcal{J}_{0,\cdot}\|_\gamma\big\|\mathcal{J}_{0,\cdot}^{-1}\big\|_\gamma\|Y_\cdot\|_\gamma. \tag{5.23}$$

Thus by (5.21) we have $\tilde{\mathbb{E}}\sup_{t\in[0,1]}\|\mathbf{D}_t v\|_{0,1,\infty}^2 < \mathscr{E}(x,\omega)$. Regarding its modulus of continuity, note that

$$\mathbf{D}_t Y_s - \mathbf{D}_t Y_r = \int_r^s \mathcal{J}_{a,s}U(Y_a)(a-t)^{H-3/2}\,\mathrm{d}a + \int_t^r (\mathcal{J}_{a,s}-\mathcal{J}_{a,r})U(Y_a)(a-t)^{H-3/2}\,\mathrm{d}a.$$

The first term is bounded by

$$M|r-t|^{H-3/2}|r-s|\|\mathcal{J}_{0,\cdot}\|_\gamma\big\|\mathcal{J}_{0,\cdot}^{-1}\big\|_\gamma\|Y_\cdot\|_\gamma,$$

and the second term is bounded by

$$M|r - s|^\gamma \|\mathcal{J}_{0,\cdot}\|_\gamma \|\mathcal{J}_{0,\cdot}^{-1}\|_\gamma \|Y.\|_\gamma,$$

so that there does exist a constant M such that

$$|\mathbf{D}_t Y_s - \mathbf{D}_t Y_r| \le M|r - t|^{H-3/2}|r - s|\|\mathcal{J}_{0,\cdot}\|_\gamma \|\mathcal{J}_{0,\cdot}^{-1}\|_\gamma \|Y.\|_\gamma,$$

uniformly over $t \le r \le s \le 1$. Since $\gamma > \frac{1}{2}$, we do indeed have $\frac{3}{2} - H < 1 + \gamma - H$, so that this bound is of the type (5.18) with $\tilde{\mathbb{E}} K^p < \mathscr{E}(x, \omega)$ for every $p > 0$.

Notice furthermore that by Theorem 4.2, the matrix $C_1 = \mathcal{A}h\mathcal{A}^* \in \mathbb{R}^{n^2}$ is almost surely invertible and $\tilde{\mathbb{E}}|\mathcal{A}h\mathcal{A}^*|^{-p} < \mathscr{E}(x, \omega)$ for all $p \ge 1$. Combining this with (5.21) and the expressions for $\mathcal{I}_1$ and $\mathcal{I}_2$, we conclude that (5.18) does indeed hold for these terms.

Let us now turn to the term $\mathcal{I}_3$. Recall that the Fréchet derivative of the inverse an $n \times n$ matrix A in the direction H is given by $D^H A^{-1} = -A^{-1}HA^{-1}$. It thus follows from the chain rule that

$$\mathbf{D}_t\big((\mathcal{A}h\mathcal{A}^*)^{-1}\xi\big) = -(\mathcal{A}h\mathcal{A}^*)^{-1}\big(\mathbf{D}_t(\mathcal{A}h\mathcal{A}^*)\big)(\mathcal{A}h\mathcal{A}^*)^{-1}\xi.$$

Thus by Cauchy–Schwartz we have

$$\big(\tilde{\mathbb{E}}\|\mathbf{D}_t\big((\mathcal{A}h\mathcal{A}^*)^{-1}\xi\big)\|_{0,1,\infty}^p\big)^2 \le \tilde{\mathbb{E}}\big|(\mathcal{A}h\mathcal{A}^*)^{-1}\big|^{4p}\tilde{\mathbb{E}}\|\mathbf{D}_t(\mathcal{A}h\mathcal{A}^*)\|_{0,1,\infty}^{2p}|\xi|^2.$$

By the product rule and the definition of $\mathcal{A}$,

$$\mathbf{D}_t(\mathcal{A}h\mathcal{A}^*) = \int_0^1 \big(\mathbf{D}_t J_{0,s}^{-1}\big)V(X_s)h(s)V^*(X_s)\big(J_{0,s}^{-1}\big)^* \,\mathrm{d}s + \cdots, \tag{5.24}$$

where "$\cdots$" stands for the remaining three terms that are very similar. Thus from (5.23) it follows that

$$\tilde{\mathbb{E}}\left|\int_0^1 \big(\mathbf{D}_t J_{0,s}^{-1}\big)V(X_s)h(s)V^*(X_s)\big(J_{0,s}^{-1}\big)^* \,\mathrm{d}s\right|^p \le M\tilde{\mathbb{E}}\|\mathcal{J}_{0,\cdot}\|_\gamma^p \|\mathcal{J}_{0,\cdot}^{-1}\|_\gamma^p \|Y.\|_\gamma^{2p} \|V(X.)\|_\gamma^{2p} < \mathscr{E}(x, \omega).$$

Since J^{-1} and $V(X)$ play indistinguishable roles in (5.20) and (5.21), the bounds on the remaining three terms follow in an identical fashion, so that

$$\mathbb{E}\sup_{t \in [0,1]} \big|\mathbf{D}_t\big((\mathcal{A}h\mathcal{A}^*)^{-1}\xi\big)\big|^p < \mathscr{E}(x, \omega).$$

Combining this with (5.21) and the definition of $\mathcal{I}_3$, we finally conclude that (5.18) also holds for the term $\mathcal{I}_3$ (this time with $\alpha = 0$), thus concluding the proof. $\qquad\square$

Appendix: Some technical lemmas

We give the proofs of some technical lemmas in this appendix. First, we show that Lemma 2.5 holds, which we reformulate here:

Lemma A.1. *Let $\omega \in \mathcal{W}$. Then, on any time interval bounded away from 0, the map $t \mapsto \mathcal{G}\omega(t)$ is $\mathcal{C}^\infty$. Furthermore, if we set $f_\omega(t) = t\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{G}\omega(t)$, then we have $f_\omega(0) = 0$ and for every $T > 0$, there exists a constant M_T such that $\|f_\omega\|_{0,T,\gamma} < M_T \|\omega\|_{(\gamma,\delta)}$.*

Proof. The proof follows from the calculations done in [16]. Since the function $g(\cdot)$ defined in (2.7) is smooth everywhere except at 0, it follows from (2.6) that m_t is smooth on any interval bounded away from 0. Now a simple change of variables in (2.6) yields

$$t\frac{\mathrm{d}m_t}{\mathrm{d}t} = \gamma_H \int_0^\infty \frac{1}{r}tg'(t)\omega\left(-\frac{t}{r}\right)\mathrm{d}r. \tag{A.1}$$

Since $g'(t)$ is smooth everywhere except at 0, for the map $t \mapsto t\frac{dm}{dt}(t)$ to have finite γ-Hölder norm on $[0, T]$, by (A.1) we need to only check the regularity of $tg'(t)$ at 0 and at infinity. Indeed, if we show that the function $tg'(t)$ behaves like $\mathcal{O}(t)$ for $t \ll 1$, and $\mathcal{O}(t^{H-1/2})$ for $t \gg 1$, the conclusion follows from Lemma A.1 and Proposition A.2 of [16].

An explicit calculation shows that we have the identity

$$tg'(t) = g(t) + \left(H - \frac{3}{2}\right)\left(\left(H - \frac{5}{2}\right)t^2 \int_0^1 \frac{(u+t)^{H-7/2}}{(1-u)^{H-1/2}}\,du + t^{H-1/2}\right). \tag{A.2}$$

Notice that by Lemma A.1 of [16] we have that $g(t)$ behaves like $\mathcal{O}(t)$ for $t \ll 1$ and $\mathcal{O}(t^{H-1/2})$ for $t \gg 1$. It can be seen easily that for $t \gg 1$, the term $t^2 \int_0^1 \frac{(u+t)^{H-7/2}}{(1-u)^{H-1/2}}\,du$ behaves like $\mathcal{O}(t^{H-3/2})$. For $t \ll 1$, we have

$$t^2 \int_0^1 \frac{(u+t)^{H-7/2}}{(1-u)^{H-1/2}}\,du = t^2 \int_0^1 (u+t)^{H-7/2}\,du + M(t)$$

$$= -\frac{1}{(H-5/2)}t^{H-1/2} + M(t) + \mathcal{O}(t^2), \tag{A.3}$$

where

$$M(t) = t^2 \int_0^1 (u+t)^{H-7/2}\left((1-u)^{1/2-H} - 1\right)du.$$

Notice that $|M(t)| < t|\int_0^1 u^{H-5/2}((1-u)^{1/2-H} - 1)\,du|$. Since the factor $((1-u)^{1/2-H} - 1)$ is $\mathcal{O}(u)$ for $u \approx 0$, and diverges like $\mathcal{O}((1-u)^{1/2-H})$ for $u \approx 1$, $\int_0^1 u^{H-5/2}((1-u)^{1/2-H} - 1)\,du < \infty$ and thus $M(t) = \mathcal{O}(t)$ for $t \ll 1$. Thus from (A.2) it follows that $tg'(t)$ is $\mathcal{O}(t)$ for $t \ll 1$ and the claim follows. $\qquad\square$

Our next technical results yields a bound on the Hölder constant of a function defined on a large interval as a function of the Hölder constants of its restriction to smaller subintervals:

Lemma A.2. *For any $N \geq 1$ and any sequence of times $0 = u_0 < u_1 < u_2 < \cdots < u_{N-1} < u_N = T$, the bound*

$$\|x\|_{0,T,\gamma} \leq N^{1-\gamma} \max_{0 \leq i \leq N-1} \|x\|_{u_i, u_{i+1}, \gamma} \tag{A.4}$$

holds.

Proof. We proceed by induction. For $N = 1$, the bound (A.4) holds trivially. Now assume (A.4) holds for some $N \in \mathbb{N}$ and consider the sequence $0 = u_0 < u_1 < u_2 < \cdots < u_{N-1} < u_N < u_{N+1} = T$. If the supremum $S \stackrel{\text{def}}{=} \sup_{s \neq t \in [0,T]} \frac{|x(t)-x(s)|}{|t-s|^\gamma}$ is attained either for $0 \leq s < t < u_N$ or for $u_1 \leq s < t < u_{N+1}$ then we can use the induction hypothesis to conclude that

$$\|x\|_{0,T,\gamma} = \|x\|_{0,u_N,\gamma} \vee \|x\|_{u_1,u_{N+1},\gamma}$$

$$\leq N^{1-\gamma} \max_{0 \leq i \leq N} \|x\|_{u_i,u_{i+1},\gamma} \leq (N+1)^{1-\gamma} \max_{0 \leq i \leq N} \|x\|_{u_i,u_{i+1},\gamma},$$

and we are done. It remains to consider the case where the supremum S is attained for some $s \in [0, u_1]$ and $t \in [u_N, T]$. We then have

$$S \leq \frac{|x(u_1) - x(s)| + |x(u_2) - x(u_1)| + \cdots + |x(t) - x(u_N)|}{|t-s|^\gamma}$$

$$\leq \frac{|u_1 - s|^\gamma \|x\|_{0,u_1,\gamma}}{|t-s|^\gamma} + \frac{|u_2 - u_1|^\gamma \|x\|_{u_1,u_2,\gamma}}{|t-s|^\gamma} + \cdots + \frac{|t - u_N|^\gamma \|x\|_{u_N,u_{N+1},\gamma}}{|t-s|^\gamma}$$

$$\leq \left(\frac{|u_1 - s|^\gamma}{|t-s|^\gamma} + \frac{|u_2 - u_1|^\gamma}{|t-s|^\gamma} + \cdots + \frac{|t - u_N|^\gamma}{|t-s|^\gamma}\right) \max_{0 \leq i \leq N} \|x\|_{u_i,u_{i+1},\gamma}. \tag{A.5}$$

Since the function $x \mapsto x^\gamma$ is concave for $\gamma < 1$ and since $|u_1 - s| + \cdots + |t - u_N| = |t - s|$, we have

$$\frac{1}{N+1}\left(\frac{|u_1 - s|^\gamma}{|t - s|^\gamma} + \frac{|u_2 - u_1|^\gamma}{|t - s|^\gamma} + \cdots + \frac{|t - u_N|^\gamma}{|t - s|^\gamma}\right)$$
$$\leq \left(\frac{1}{N+1}\frac{|u_1 - s|}{|t - s|} + \frac{1}{N+1}\frac{|u_2 - u_1|}{|t - s|} + \cdots + \frac{1}{N+1}\frac{|t - u_N|}{|t - s|}\right)^\gamma = \frac{1}{(N+1)^\gamma}.$$

Combining this with (A.5), we obtain

$$S \leq (N+1)^{1-\gamma} \max_{0 \leq i \leq N} \|x\|_{u_i, u_{i+1}, \gamma},$$

thus concluding the proof.   □

Finally, we provide the following interpolation result:

Lemma A.3. *The bound*

$$\|f\|_{0,T,\infty} \leq 2\max\left(T^{-1/2}\|f\|_{L^2[0,T]}, \|f\|_{L^2[0,T]}^{2\gamma/(2\gamma+1)}\|f\|_{0,T,\gamma}^{1/(2\gamma+1)}\right)$$

holds for every $\gamma \in (0, 1]$ and every γ-Hölder continuous function $f : [0, T] \mapsto \mathbb{R}$.

Proof. Let $f_{\min} = \inf_{t\in[0,T]}|f(t)|$, so that $f_{\min} \leq \frac{\|f\|_{L^2[0,T]}}{\sqrt{T}}$. Therefore it follows from the definition of Hölder continuity that

$$\|f\|_{0,T,\infty} \leq f_{\min} + \|f\|_{0,T,\gamma}T^\gamma \leq \frac{\|f\|_{L^2[0,T]}}{\sqrt{T}} + \|f\|_{0,T,\gamma}T^\gamma. \tag{A.6}$$

For $a > 0$ define the set

$$L_f^a = \left\{x \in [0, T]; \left|f(x)\right| > a\right\}.$$

By Chebyshev's inequality we have, for any $a > 0$, $\lambda(L_f^a) \leq \frac{\|f\|_{L^2[0,T]}^2}{a^2}$ where λ denotes the Lebesgue measure. Thus for any $a > f_{\min}$ and $x \in L_f^a$, we have

$$\left|f(x)\right| \leq a + \|f\|_{0,T,\gamma}\left|\lambda\left(L_f^a\right)\right|^\gamma,$$

so that

$$\|f\|_{0,T,\infty} \leq a + a^{-2\gamma}\|f\|_{0,T,\gamma}\|f\|_{L^2[0,T]}^{2\gamma}. \tag{A.7}$$

Now we combine the estimates obtained in (A.6) and (A.7). Making the choice $a = \|f\|_{0,T,\gamma}^{1/(2\gamma+1)}\|f\|_{L^2[0,T]}^{2\gamma/(2\gamma+1)}$ in (A.7) yields the bound

$$\|f\|_{0,T,\infty} \leq 2\|f\|_{0,T,\gamma}^{1/(2\gamma+1)}\|f\|_{L^2[0,T]}^{2\gamma/(2\gamma+1)}, \tag{A.8}$$

provided that we have $\|f\|_{0,T,\gamma}^{1/(2\gamma+1)}\|f\|_{L^2[0,T]}^{2\gamma/(2\gamma+1)} \geq \frac{\|f\|_{L^2[0,T]}}{\sqrt{T}} \geq f_{\min}$. This is the case when $\|f\|_{0,T,\gamma} \geq \|f\|_{L^2[0,T]}T^{-\gamma-1/2}$. When $\|f\|_{0,T,\gamma} \leq \|f\|_{L^2[0,T]}T^{-\gamma-1/2}$, from (A.6) we obtain

$$\|f\|_{0,T,\infty} \leq 2\|f\|_{L^2[0,T]}T^{-1/2}, \tag{A.9}$$

thus completing the proof.   □

Acknowledgements

References

[1] L. Arnold. *Random Dynamical Systems*. Springer, Berlin, 1998. MR1723992

[2] F. Baudoin and L. Coutin. Operators associated with a stochastic differential equation driven by fractional Brownian motions. *Stochastic Process. Appl.* **117** (2007) 550–574. MR2320949

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

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

[5] V. I. Bogachev. *Gaussian Measures. Mathematical Surveys and Monographs* **62**. Amer. Math. Soc., Providence, RI, 1998. MR1642391

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

[7] G. Da Prato, K. D. Elworthy and J. Zabczyk. Strong Feller property for stochastic semilinear equations. *Stochastic Anal. Appl.* **13** (1995) 35–45. MR1313205

[8] G. Da Prato and J. Zabczyk. *Ergodicity for Infinite-Dimensional Systems. London Mathematical Society Lecture Note Series* **229**. Cambridge Univ. Press, Cambridge, 1996. MR1417491

[9] K. D. Elworthy and X.-M. Li. Formulae for the derivatives of heat semigroups. *J. Funct. Anal.* **125** (1994) 252–286. MR1297021

[10] P. K. Friz. *Multidimensional Stochastic Processes as Rough Paths. Theory and Applications. Cambridge Studies in Advanced Mathematics* **120**. Cambridge Univ. Press, Cambridge, 2010.

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

[12] M. Hairer. Ergodic properties of a class of non-Markovian processes. In *Trends in Stochastic Analysis. LMS Lecture Note Series* **353**. Cambridge Univ. Press, Cambridge, 2009. MR2562151

[13] M. Hairer and J. C. Mattingly. Ergodicity of the 2D Navier–Stokes equations with degenerate stochastic forcing. *Ann. of Math. (2)* **164** (2006) 993–1032. MR2259251

[14] M. Hairer and J. C. Mattingly. A theory of hypoellipticity and unique ergodicity for semilinear stochastic PDEs. Preprint, 2009.

[15] Y. Hu and D. Nualart. Differential equations driven by Hölder continuous functions of order greater than 1/2. In *Stochastic Analysis and Applications* 399–413. *Abel Symp.* **2**. Springer, Berlin, 2007. MR2397797

[16] M. Hairer and A. Ohashi. Ergodic theory for SDEs with extrinsic memory. *Ann. Probab.* **35** (2007) 1950–1977. MR2349580

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

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

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

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

[21] T. Lyons and Z. Qian. *System Control and Rough Paths*. Oxford Univ. Press, Oxford, 2002. MR2036784

[22] P. Malliavin. Stochastic calculus of variations and hypoelliptic operators. In *Symp. on Stoch. Diff. Equations, Kyoto 1976* 147–171. Wiley, New York, 1978. MR0536013

[23] P. Malliavin. *Stochastic Analysis. Grundlehren der Mathematischen Wissenschaften* **313**. Springer, Berlin, 1997. MR1450093

[24] Y. S. Mishura. *Stochastic Calculus for Fractional Brownian Motion and Related Processes. Lecture Notes in Mathematics* **1929**. Springer, Berlin, 2008. MR2378138

[25] L. Mazliak and I. Nourdin. Optimal control for rough differential equations. *Stoch. Dyn.* **8** (2008) 23–33. MR2399923

[26] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Springer, London, 1993. MR1287609

[27] B. B. Mandelbrot and J. W. van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10** (1968) 422–437. MR0242239

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

[29] I. Nourdin and T. Simon. On the absolute continuity of one-dimensional SDEs driven by a fractional Brownian motion. *Statist. Probab. Lett.* **76** (2006) 907–912. MR2268434

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

[31] D. Nualart. *The Malliavin Calculus and Related Topics*, 2nd edition. Springer, Berlin, 2006. MR2200233

[32] S. G. Samko, A. A. Kilbas and O. I. Marichev. *Fractional Integrals and Derivatives*. Gordon and Breach Science Publishers, Yverdon, 1993. MR1347689

[33] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes*. Chapman & Hall, New York, 1994. MR1280932

[34] L. C. Young. An inequality of the Hölder type, connected with Stieltjes integration. *Acta Math.* **67** (1936) 251–282. MR1555421