On Malliavin's proof of Hörmander's theorem

March 10, 2011

Martin Hairer

Mathematics Department, University of Warwick
Email: M.Hairer@Warwick.ac.uk

Abstract

The aim of this note is to provide a short and self-contained proof of Hörmander's theorem about the smoothness of transition probabilities for a diffusion under Hörmander's "brackets condition". While both the result and the technique of proof are well-known, the exposition given here is novel in two aspects. First, we introduce Malliavin calculus in an "intuitive" way, without using Wiener's chaos decomposition. While this may make it difficult to prove some of the standard results in Malliavin calculus (boundedness of the derivative operator in L^p spaces for example), we are able to bypass these and to replace them by weaker results that are still sufficient for our purpose. Second, we introduce a notion of "almost implication" and "almost truth" (somewhat similar to what is done in fuzzy logic) which allows, once the foundations of Malliavin calculus are laid out, to give a very short and streamlined proof of Hörmader's theorem that focuses on the main ideas without clouding it by technical details.

Dedicated to the memory of Paul Malliavin.

1 Introduction

One of the main tools in many results on the convergence to equilibrium of Markov processes is the presence of some form of "smoothing" for the semigroup. For example, if a Markov operator $\mathcal{P}$ over a Polish space $\mathcal{X}$ possesses the strong Feller property (namely it maps $\mathcal{B}_b(\mathcal{X})$, the space of bounded measurable functions into $\mathcal{C}_b(\mathcal{X})$, the space of bounded continuous functions), then one can conclude that any two ergodic invariant measures for $\mathcal{P}$ must either coincide or have disjoint topological supports. Since the latter can often been ruled out by some form of controllability argument, we see how the strong Feller property is the basis for many proofs of ergodicity.

It is then desirable to have criteria that are as simple to formulate as possible and that ensure that the Markov semigroup associated to a given Markov process has some smoothing property. One of the most natural classes of Markov processes are given by diffusion processes and this will be the object of study in this note. Our

main object of study is a stochastic differential equation of the form

$$dx = V_0(x)\,dt + \sum_{i=1}^{m} V_i(x) \circ dW_i \;, \tag{1.1}$$

where the V_i's are smooth vector fields on $\mathbf{R}^n$ and the W_i's are independent standard Wiener processes. In order to keep all arguments as straightforward as possible, we will assume throughout this note that these vector fields assume the coercivity assumptions necessary so that the solution flow to (1.1) is smooth with respect to its initial condition and that all of its derivatives have moments of all orders. This is satisfied for example if the V_i's are $\mathcal{C}^\infty$ with bounded derivatives of all orders.

Remark 1.1 We wrote (1.1) as a Stratonowich equation on purpose. This is for two reasons: at a pragmatic level, this is the "correct" formulation which allows to give a clean statement of Hörmander's theorem (see Definition 1.2 below). At the intuitive level, the question of smoothness of transition probabilities is related to that of the extent of their support. The Stroock-Varadhan support theorem [SV72] characterises this as consisting precisely of the closure of the set of points that can be reached if the Wiener processes W_i in (1.1) are replaced by arbitrary smooth control functions. This would not be true in general for the Itô formulation.

It is well-known that if the equation (1.1) is elliptic namely if, for every point $x \in \mathbf{R}^n$, the linear span of $\{V_i(x)\}_{i=1}^m$ is all of $\mathbf{R}^n$, then the law of the solution to (1.1) has a smooth density with respect to Lebesgue measure. Furthermore, the corresponding Markov semigroup $\mathcal{P}_t$ defined by

$$\mathcal{P}_t\varphi(x_0) = \mathbf{E}_{x_0}\varphi(x_t) \;,$$

is so that $\mathcal{P}_t\varphi$ is smooth, even if φ is only bounded measurable. (Think of the solution to the heat equation, which corresponding to the simplest case where $V_0 = 0$ and the V_i form an orthonormal basis of $\mathbf{R}^n$.) In practice however, one would like to obtain a criterion that also applies to some equations where the ellipticity assumption fails. For example, a very well-studied model of equilibrium statistical mechanics is given by the Langevin equation:

$$dq = p\,dt \;, \qquad dp = -\nabla V(q)\,dt - p\,dt + \sqrt{2T}\,dW(t) \;,$$

where $T > 0$ should be interpreted as a temperature, $V\colon \mathbf{R}^n \to \mathbf{R}_+$ is a sufficiently coercive potential function, and W is an n-dimensional Wiener process. Since solutions to this equation take values in $\mathbf{R}^{2n}$ (both p and q are n-dimensional), this is definitely not an elliptic equation. At an intuitive level however, one would expect it to have some smoothing properties: smoothing reflects the spreading of our uncertainty about the position of the solution and the uncertainty on p due to the presence of the noise terms gets instantly transmitted to q via the equation $dq = p\,dt$.

In a seminal paper [Hör67], Hörmander was the first to formulate the "correct" non-degeneracy condition ensuring that solutions to (1.1) have a smoothing effect. To describe this non-degeneracy condition, recall that the Lie bracket $[U, V]$ between two vector fields U and V on $\mathbf{R}^n$ is the vector field defined by

$$[U, V](x) = DV(x)\,U(x) - DU(x)\,V(x)\,,$$

where we denote by DU the derivative matrix given by $(DU)_{ij} = \partial_j U_i$. This notation is consistent with the usual notation for the commutator between two linear operators since, if we denote by A_U the first-order differential operator acting on smooth functions f by $A_U f(x) = \langle U(x), \nabla f(x)\rangle$, then we have the identity $A_{[U,V]} = [A_U, A_V]$.

With this notation at hand, we give the following definition:

Definition 1.2 Given an SDE (1.1), define a collection of vector fields $\mathscr{V}_k$ by

$$\mathscr{V}_0 = \{V_i \,:\, i > 0\}\,, \quad \mathscr{V}_{k+1} = \mathscr{V}_k \cup \{[U, V_j] \,:\, U \in \mathscr{V}_k \;\&\; j \geq 0\}\,.$$

We also define the vector spaces $\mathscr{V}_k(x) = \text{span}\{V(x) \,:\, V \in \mathscr{V}_k\}$. We say that (1.1) satisfies the *parabolic Hörmander condition* if $\bigcup_{k\geq 1} \mathscr{V}_k(x) = \mathbf{R}^n$ for every $x \in \mathbf{R}^n$.

With these notations, Hörmander's theorem can be formulated as

Theorem 1.3 *Consider (1.1) and assume that all vector fields have bounded derivatives of all orders. If it satisfies the parabolic Hörmander condition, then its solutions admit a smooth density with respect to Lebesgue measure and the corresponding Markov semigroup maps bounded functions into smooth functions.*

Hörmander's original proof was formulated in terms of second-order differential operators and was purely analytical in nature. Since the main motivation on the other hand was probabilistic and since, as we will see below, Hörmander's condition can be understood at the level of properties of the trajectories of (1.1), a more stochastic proof involving the original stochastic differential equation was sought for. The breakthrough came with Malliavin's seminal work [Mal78], where he laid the foundations of what is now known as the "Malliavin calculus", a differential calculus in Wiener space and used it to give a probabilistic proof of Hörmander's theorem. This new approach proved to be extremely successful and soon a number of authors studied variants and simplifications of the original proof [Bis81b, Bis81a, KS84, KS85, KS87, Nor86]. Even now, more than three decades after Malliavin's original work, his techniques prove to be sufficiently flexible to obtain related results for a number of extensions of the original problem, including for example SDEs with jumps [Tak02, IK06, Cas09, Tak10], infinite-dimensional systems [Oco88, BT05, MP06, HM06, HM11], and SDEs driven by Gaussian processes other than Brownian motion [BH07, CF10, HP11].

A complete rigorous proof of Theorem 1.3 goes somewhat beyond the scope of these notes. However, we hope to be able to give a convincing argument showing why this result is true and what are the main steps involved in its probabilistic proof. The aim in writing these notes was to be sufficiently self-contained so that a strong PhD student interested in stochastic analysis would be able to fill in the missing gaps without requiring additional ideas. The interested reader can find the technical details required to make the proof rigorous in [Mal78, KS84, KS85, KS87, Nor86, Nua95]. Hörmander's original, completely different, proof using fractional integrations can be found in [Hör67]. A yet completely different functional-analytic proof using the theory of pseudo-differential operators was developed by Kohn in [Koh78] and can also be found in [Hör85] or, in a slightly different context, in the recent book [HN05].

The remainder of these notes is organised as follows. First, in Section 2 below, we will show why it is natural that the iterated Lie brackets appear in Hörmander's condition. Then, in Section 3, we will give an introduction to Malliavin calculus, including in particular its integration by parts formula in Wiener space. Finally, in Section 4, we apply these tools to the particular case of smooth diffusion processes in order to give a probabilistic proof of Hörmander's theorem.

Acknowledgements

These notes were part of a minicourse given at the University of Warwick in July 2010; thanks are due to the organisers for this pleasant event. I would also like to thank Michael Scheutzow and Hendrik Weber who carefully read a previous version of the manuscript and pointed out several misprints. Remaining mistakes were probably added later on! Finally, I would like to acknowledge financial support provided by the EPSRC through grants EP/E002269/1 and EP/D071593/1, the Royal Society through a Wolfson Research Merit Award, and the Leverhulme Trust through a Philip Leverhulme prize.

2 Why is it the correct condition?

At first sight, the condition given in Definition 1.2 might seem a bit strange. Indeed, the vector field V_0 is treated differently from all the others: it appears in the recursive definition of the $\mathscr{V}_k$, but not in $\mathscr{V}_0$. This can be understood in the following way: consider trajectories of (1.1) as curves in space-time. By the Stroock-Varadhan support theorem [SV72], the law of the solution to (1.1) on pathspace is supported by the closure of those smooth curves that, at every point (x,t), are tangent to the hyperplane spanned by $\{\hat{V}_0,\dots,\hat{V}_m\}$, where we set

$$\hat{V}_0(x,t) = \begin{pmatrix} V_0(x) \\ 1 \end{pmatrix}, \qquad \hat{V}_j(x,t) = \begin{pmatrix} V_j(x) \\ 0 \end{pmatrix}, \quad j = 1,\dots,m\,.$$

With this notation at hand, we could define $\hat{\mathscr{V}}_k$ as in Definition 1.2, but with $\hat{\mathscr{V}}_0 = \{\hat{V}_0,\dots,\hat{V}_m\}$. Then, it is easy to check that Hörmander's condition is equivalent to the condition that $\bigcup_{k\geq 1}\hat{\mathscr{V}}_k = \mathbf{R}^{n+1}$ for every $(x,t)\in\mathbf{R}^{n+1}$.

This condition however has a simple geometric interpretation. For a smooth manifold $\mathcal{M}$, recall that $E \subset T\mathcal{M}$ is a smooth subbundle of dimension d if $E_x \subset T_x\mathcal{M}$ is a vector space of dimension d at every $x \in \mathcal{M}$ and if the dependency $x \mapsto E_x$ is smooth. (Locally, E_x is the linear span of finitely many smooth vector fields on $\mathcal{M}$.) A subbundle is called *integrable* if, whenever U, V are vector fields on $\mathcal{M}$ taking values in E, their Lie bracket $[U, V]$ also takes values in E.

With these definitions at hand, recall the well-known Frobenius integrability theorem from differential geometry:

Theorem 2.1 *Let $\mathcal{M}$ be a smooth n-dimensional manifold and let $E \subset T\mathcal{M}$ be a smooth vector bundle of dimension $d < n$. Then E is integrable if and only if there (locally) exists a smooth foliation of $\mathcal{M}$ into leaves of dimension d such that, for every $x \in \mathcal{M}$, the tangent space of the leaf passing through x is given by E_x.*

In view of this result, Hörmander's condition is not surprising. Indeed, if we define $E_{(x,t)} = \bigcup_{k\geq 0} \hat{\mathscr{V}}_k(x,t)$, then this gives us a subbundle of $\mathbf{R}^{n+1}$ which is integrable by construction of the $\hat{\mathscr{V}}_k$. Note that the dimension of $E_{(x,t)}$ could in principle depend on (x,t), but since the dimension is a lower semicontinuous function, it will take its maximal value on an open set. If, on some open set, this maximal value is less than $n+1$, then Theorem 2.1 tells us that, there exists a submanifold (with boundary) $\bar{\mathcal{M}} \subset \mathcal{M}$ of dimension strictly less than n such that $T_{(y,s)}\bar{\mathcal{M}} = E_{(y,s)}$ for every $(y,s) \in \bar{\mathcal{M}}$. In particular, all the curves appearing in the Stroock-Varadhan support theorem and supporting the law of the solution to (1.1) must lie in $\bar{\mathcal{M}}$ until they reach its boundary. As a consequence, since $\bar{\mathcal{M}}$ is always transverse to the sections with constant t, the solutions at time t will, with positive probability, lie in a submanifold of $\mathcal{M}$ of strictly positive codimension. This immediately implies that the transition probabilities cannot be continuous with respect to Lebesgue measure.

To summarise, if Hörmander's condition fails on an open set, then transition probabilities cannot have a density with respect to Lebesgue measure, thus showing that Hörmander's condition is "almost necessary" for the existence of densities. The hard part of course is to show that it is a sufficient condition. Intuitively, the reason is that Hörmander's condition allows the solution to (1.1) to "move in all directions". Why this is so can be seen from the following interpretation of the Lie brackets. Set

$$u_n(t) = \frac{1}{n}\cos(n^2 t)\,, \qquad v_n(t) = \frac{1}{n}\sin(n^2 t)\,,$$

and consider the solution to

$$\dot{x} = U(x)\,\dot{u}_n(t) + V(x)\,\dot{v}_n(t)\,. \tag{2.1}$$

We claim that, as $n \to \infty$, this converges to the solution to

$$\dot{u} = \frac{1}{2}[U,V](x)\,. \tag{2.2}$$

This can be seen as follows. If we integrate (2.1) over a short time interval, we have the first order approximation

$$x(h) \approx x^{(1)}(h) \stackrel{\text{def}}{=} x_0 + U(x_0)u_n(h) + V(x_0)v_n(h) \,,$$

which simply converges to x_0 as $n \to \infty$. To second order, however, we have

$$\begin{aligned} x(h) &\approx x_0 + \int_0^h (U(x^{(1)})\,\dot{u}_n + V(x^{(1)})\,\dot{v}_n)\,dt \\ &\approx x^{(1)}(h) + \int_0^h (DU(x_0)\,\dot{u}_n + DV(x_0)\,\dot{v}_n)(U(x_0)u_n + V(x_0)v_n)\,dt \\ &\approx x_0 + \int_0^h (DU(x_0)V(x_0)v_n\dot{u}_n + DV(x_0)U(x_0)u_n\dot{v}_n)\,dt \,. \end{aligned}$$

Here, we used the fact that the integral of $u_n\dot{u}_n$ (and similarly for $v_n\dot{v}_n$) is given by $\frac{1}{2}u_n^2$ and therefore converges to 0 as $n \to 0$. Note now that over a period, $v_n(t)\dot{u}_n(t)$ averages to $-\frac{1}{2}$ and $u_n(t)\dot{v}_n(t)$ averages to $\frac{1}{2}$, thus showing that one does indeed obtain (2.2) in the limit. This reasoning shows that, by combining motions in the directions U and V, it is possible to approximate, to within arbitrary accuracy, motion in the direction $[U, V]$.

A similar reasoning shows that if we consider

$$\dot{x} = U(x) + V(x)\,\dot{v}_n(t) \,,$$

then, to lowest order in $1/n$, we obtain that as $n \to \infty$, x follows

$$\dot{x} \approx U(x) + \frac{1}{2n}[U, V](x) \,.$$

Combining these interpretations of the meaning of Lie brackets with the Stroock-Varadhan support theorem, it suggests that, if Hörmander's condition holds, then the support of the law of x_t will contain an open set around the solution at time t to the deterministic system

$$\dot{x} = V_0(x) \,, \qquad x(0) = x_0 \,.$$

This should at least render it plausible that under these conditions, the law of x_t has a density with respect to Lebesgue measure. The aim of this note is to demonstrate how to turn this heuristic into a mathematical theorem with, hopefully, a minimum amount of effort.

Remark 2.2 While Hörmander's condition implies that the control system associated to (1.1) reaches an open set around the solution to the deterministic equation $\dot{x} = V_0(x)$, it does *not* imply in general that it can reach an open set around x_0. In particular, it is *not* true that the parabolic Hörmander condition implies that (1.1) can reach every open set. A standard counterexample is given by

$$dx = -\sin(x)\,dt + \cos(x) \circ dW(t) \,, \qquad x_0 = 0 \,,$$

which satisfies Hörmander's condition but can never exit the interval $[-\pi/2, \pi/2]$.

3 An Introduction to Malliavin calculus

In this section, we collect a number of tools that will be needed in the proof. The main tool is the integration by parts formula from Malliavin calculus, as well of course as Malliavin calculus itself.

The main tool in the proof is the Malliavin calculus with its integration by part formula in Wiener space, which was developed precisely in order to provide a probabilistic proof of Theorem 1.3. It essentially relies on the fact that the image of a Gaussian measure under a smooth submersion that is sufficiently integrable possesses a smooth density with respect to Lebesgue measure. This can be shown in the following way. First, one observes the following fact:

Lemma 3.1 *Let μ be a probability measure on $\mathbf{R}^n$ such that the bound*

$$\left| \int_{\mathbf{R}^n} D^{(k)} G(x)\, \mu(dx) \right| \leq C_k \|G\|_\infty \,,$$

holds for every smooth bounded function G and every $k \geq 1$. Then μ has a smooth density with respect to Lebesgue measure. □

Proof. Let $s > n/2$ so that $H^s \subset \mathcal{C}_b$ by Sobolev embedding. By duality, the assumption then implies that every distributional derivative of μ belongs to the Sobolev space H^{-s}, so that μ belongs to H^ℓ for every $\ell \in \mathbf{R}$. The result then follows from the fact that $H^\ell \subset \mathcal{C}^k$ as soon as $\ell > k + \frac{n}{2}$. □

Consider now a sequence of N independent Gaussian random variables δw_k with variances δt_k for $k \in \{1, \dots, N\}$, as well as a smooth map $X : \mathbf{R}^N \to \mathbf{R}^n$. We also denote by w the collection $\{\delta w_k\}_{k \geq 1}$ and we define the $n \times n$ matrix-valued map

$$\mathscr{M}_{ij}(w) = \sum_k \partial_k X_i(w) \partial_k X_j(w)\, \delta t_k \,, \tag{3.1}$$

where we use ∂_k as a shorthand for the partial derivative with respect to the variable δw_k. With this notation, X being a submersion is equivalent to $\mathscr{M}(w)$ being invertible for every w.

Before we proceed, let us introduce additional notation, which hints at the fact that one would really like to interpret the δw_k as the increments of a Wiener process of an interval of length δt_k. When considering a family $\{F_k\}_{k=1}^N$ of maps from $\mathbf{R}^N \to \mathbf{R}^n$, we identify it with a continuous family $\{F_t\}_{t \geq 0}$, where

$$F_t \stackrel{\text{def}}{=} F_k \,, \quad t \in [t_k, t_{k+1}) \,, \qquad t_k \stackrel{\text{def}}{=} \sum_{\ell \leq k} \delta t_\ell \,. \tag{3.2}$$

Note that with this convention, we have $t_0 = 0$, $t_1 = \delta t_1$, etc. This is of course an abuse of notation since F_t is not equal to F_k for $t = k$, but we hope that it will always be clear from the context whether the index is a discrete or a continuous

variable. We also set $F_t = 0$ for $t \geq t_N$. With this notation, we have the natural identity

$$\int F_t \, dt = \sum_{k=1}^{N} F_k \, \delta t_k \; .$$

Furthermore, given a smooth map $G\colon \mathbf{R}^N \to \mathbf{R}$, we will from now on denote by $\mathscr{D}_t G$ the family of maps such that $\mathscr{D}_t G = \partial_k G$ for $t \in [t_k, t_{k+1})$, so that (3.1) can be rewritten as

$$\mathscr{M}_{ij}(w) = \int \mathscr{D}_t X_i(w) \mathscr{D}_t X_j(w) \, dt \; .$$

The quantity $\mathscr{D}_t G$ is called the *Malliavin derivative* of the random variable G.

The main feature of the Malliavin derivative operator $\mathscr{D}_t$ suggesting that one expects it to be well-posed in the limit $N \to \infty$ is that it was set up in such a way that it is invariant under refinement of the mesh $\{\delta t_k\}$ in the following way. For every k, set $\delta w_k = \delta w_k^- + \delta w_k^+$, where $\delta w_k^\pm$ are independent Gaussians with variances $\delta t_k^\pm$ with $\delta t_k^- + \delta t_k^+ = \delta t_k$ and then identify maps $G\colon \mathbf{R}^N \to \mathbf{R}$ with a map $\bar{G}\colon \mathbf{R}^{2N} \to \mathbf{R}$ by

$$\bar{G}(\delta w_1^\pm, \dots, \delta w_N^\pm) = G(\delta w_1^- + \delta w_1^+, \dots, \delta w_N^- + \delta w_N^+) \; .$$

Then, for every $t \geq 0$, $\mathscr{D}_t \bar{G}$ is precisely the map identified with $\mathscr{D}_t G$.

With all of these notations at hand, we then have the following result:

Theorem 3.2 *Let $X\colon \mathbf{R}^N \to \mathbf{R}$ be smooth, assume that $\mathscr{M}(w)$ is invertible for every w and that, for every $p > 1$ and every $m \geq 0$, we have*

$$\mathbf{E}|\partial_{k_1} \cdots \partial_{k_m} X(w)|^p < \infty \; , \qquad \mathbf{E}\|\mathscr{M}(w)^{-1}\|^p < \infty \; . \tag{3.3}$$

Then the law of $X(w)$ has a smooth density with respect to Lebesgue measure. Furthermore, the derivatives of the law of X can be bounded from above by expressions that depend only on the bounds (3.3), but are independent of N, provided that $\sum \delta t_k = T$ remains fixed.

Besides Lemma 3.1, the main ingredient of the proof of Theorem 3.2 is the following integration by parts formula which lies at the heart of the success of Malliavin calculus. If F_k and G are square integrable functions with square integrable derivatives, then we have the identity

$$\begin{aligned} \mathbf{E}\Big(\int \mathscr{D}_t G(w) \, F_t(w) \, dt \Big) &= \mathbf{E} \sum_k \partial_k G(w) F_k(w) \, \delta t_k \\ &= \mathbf{E} G(w) \sum_k F_k(w) \, \delta w_k - \mathbf{E} G(w) \sum_k \partial_k F_k(w) \, \delta t_k \\ &\stackrel{\text{def}}{=} \mathbf{E}\Big(G(w) \int F_t \, dw(t) \Big) \; , \end{aligned} \tag{3.4}$$

where we *defined* the Skorokhod integral $\int F_t\,dw(t)$ by the expression on the second line. Note that in order to obtain (3.4), we only integrated by parts with respect to the variables δw_k.

Remark 3.3 The Skorokhod integral is really an extension of the usual Itô integral, which is the justification for our notation. This is because, if F_t is an adapted process, then F_{t_k} is independent of δw_ℓ for $\ell \geq k$ by definition. As a consequence, the term $\partial_k F_k$ drops and we are reduced to the usual Itô integral.

Remark 3.4 It follows immediately from the definition that one has the identity

$$\mathscr{D}_t \int F_s\,dw(s) = F_t + \int \mathscr{D}_t F_s\,dw(s)\;. \tag{3.5}$$

Formally, one can think of this identity as being derived from the Leibnitz rule, combined with the identity $\mathscr{D}_t(dw(s)) = \delta(t-s)\,ds$, which is a kind of continuous analogue of the trivial discrete identity $\partial_k \delta w_\ell = \delta_{k\ell}$.

This Skorokhod integral satisfies the following extension of Itô's isometry:

Proposition 3.5 *Let F_k be square integrable functions with square integrable derivatives, then*

$$\begin{aligned}\mathbf{E}\Big(\int F_t\,dw(t)\Big)^2 &= \mathbf{E}\int F_t^2(w)\,dt + \mathbf{E}\int\int \mathscr{D}_t F_s(w)\,\mathscr{D}_t F_s(w)\,ds\,dt \\ &\leq \mathbf{E}\int F_t^2(w)\,dt + \mathbf{E}\int\int |\mathscr{D}_t F_s(w)|^2\,ds\,dt\;,\end{aligned}$$

holds.

Proof. It follows from the definition that one has the identity

$$\mathbf{E}\Big(\int F_t\,dw(t)\Big)^2 = \sum_{k,\ell} \mathbf{E}(F_k F_\ell\,\delta w_k \delta w_\ell + \partial_k F_k \partial_\ell F_\ell\,\delta t_k \delta t_\ell - 2F_k \partial_\ell F_\ell\,\delta w_k \delta t_\ell)\;.$$

Applying the identity $\mathbf{E}G\,\delta w_\ell = \mathbf{E}\partial_\ell G\,\delta t_\ell$ to the first term in the above formula (with $G = F_k F_\ell\,\delta w_k$), we thus obtain

$$\ldots = \sum_{k,\ell} \mathbf{E}(F_k F_\ell \delta_{k,\ell} \delta t_\ell + \partial_k F_k\,\partial_\ell F_\ell\,\delta t_k \delta t_\ell + (F_\ell \partial_\ell F_k - F_k \partial_\ell F_\ell)\,\delta w_k \delta t_\ell)\;.$$

Applying the same identity to the last term then finally leads to

$$\ldots = \sum_{k,\ell} \mathbf{E}(F_k F_\ell \delta_{k,\ell} \delta t_\ell + \partial_k F_\ell\,\partial_\ell F_k\,\delta t_k \delta t_\ell)\;,$$

which is precisely the desired result. □

As a consequence, we have the following:

Proposition 3.6 *Assume that $\sum \delta t_k = T < \infty$. Then, for every $p > 0$ there exists $C > 0$ and $k > 0$ such that the bound*

$$\mathbf{E}\Big|\int F_s\,dw(s)\Big|^p \le C\Big(1 + \sum_{0\le\ell\le k}\sup_{t_0,\dots,t_\ell}\mathbf{E}|\mathscr{D}_{t_1}\cdots\mathscr{D}_{t_\ell}F_{t_0}|^{2p}\Big)\,,$$

holds. Here, C may depend on T and p, but k depends only on p.

Proof. Since the case $p \le 2$ follows from Proposition 3.5, we can assume without loss of generality that $p > 2$. Combining (3.4) with (3.5) and then applying Hölder's inequality, we have

$$\begin{aligned}\mathbf{E}\Big|\int F_s\,dw(s)\Big|^p &= (p-1)\,\mathbf{E}\Big|\int F_s\,dw(s)\Big|^{p-2}\int F_t\Big(F_t + \int \mathscr{D}_t F_s\,dw(s)\Big)\,dt\\ &\le \frac{1}{2}\mathbf{E}\Big|\int F_s\,dw(s)\Big|^p + c\mathbf{E}\int\Big|F_t + \int \mathscr{D}_t F_s\,dw(s)\Big|^{\frac{2p}{3}}\,dt + c\mathbf{E}\int |F_t|^{2p}\,dt\\ &\le \frac{1}{2}\mathbf{E}\Big|\int F_s\,dw(s)\Big|^p + c\mathbf{E}\int\Big|\int \mathscr{D}_t F_s\,dw(s)\Big|^{\frac{2p}{3}}\,dt + c\mathbf{E}\int (1+|F_t|)^{2p}\,dt\,.\end{aligned}$$

where c is some constant depending on p and T that changes from line to line. The claim now follows by induction. □

Remark 3.7 The bound in Proposition 3.6 is clearly very far from optimal. Actually, it is known that, for every $p \ge 1$, there exists C such that

$$\mathbf{E}\Big|\int F_s\,dw(s)\Big|^{2p} \le C\mathbf{E}\Big|\int F_s^2\,ds\Big|^p + C\mathbf{E}\Big|\int |\mathscr{D}_t F_s|^2\,ds\,dt\Big|^p\,,$$

even if $T = \infty$. However, this extension of the Burkholder-Davies-Gundy inequality requires highly non-trivial harmonic analysis and, to best of the author's knowledge, cannot be reduced to a short elementary calculation. The reader interested in knowing more can find its proof in [Nua95, Ch. 1.3–1.5].

The proof of Theorem 3.2 is now straightforward:

Proof of Theorem 3.2. We want to show that Lemma 3.1 can be applied. For $\eta \in \mathbf{R}^n$, we then have from the definition of $\mathscr{M}$ the identity

$$(D_j G)(X(w)) = \sum_{k,m}\partial_k(G(X(w)))\partial_k X_m(w)\,\delta t_k\,\mathscr{M}^{-1}_{mj}(w)\,. \tag{3.6}$$

Combining this identity with (3.4), it follows that

$$\mathbf{E}D_j G(X) = \mathbf{E}\Big(G(X(w))\sum_m\int \mathscr{D}_t X_m(w)\,\mathscr{M}^{-1}_{mj}(w)\,dw(t)\Big)\,. \tag{3.7}$$

Note that, by the chain rule, one has the identity

$$\mathscr{D}_t \mathscr{M}^{-1} = -\mathscr{M}^{-1}(\mathscr{D}_t \mathscr{M})\mathscr{M}^{-1} ,$$

and similarly for higher order derivatives, so that the Malliavin derivatives of $\mathscr{M}^{-1}$ can be bounded by terms involving $\mathscr{M}^{-1}$ and the Malliavin derivatives of X.

Combining this with Proposition 3.5 and (3.3) immediately shows that the requested result holds for $k = 1$. Higher values of k can be treated by induction by repeatedly applying (3.6). This will lead to expressions of the type (3.7), with the right hand side consisting of multiple Skorokhod integrals of higher order polynomials in $\mathscr{M}^{-1}$ and derivatives of X.

By Proposition 3.6, the moments of each of the terms appearing in this way can be bounded by finitely many of the expressions appearing in the assumption so that the required statement follows. □

4 Application to Diffusion Processes

We are now almost ready to tackle the proof of Hörmander's theorem. Before we start, we discuss how $\mathscr{D}_s X_t$ can be computed when X_t is the solution to an SDE of the type (1.1) and we use this discussion to formulate precise assumption for our theorem.

4.1 Malliavin Calculus for Diffusion Processes

By taking the limit $N \to \infty$ and $\delta t_k \to 0$ with $\sum \delta t_k = 1$, the results in the previous section show that one can define a "Malliavin derivative" operator $\mathscr{D}$, acting on a suitable class of "smooth" random variables and returning a stochastic process that has all the usual properties of a derivative. Let us see how it acts on the solution to an SDE of the type (1.1).

An important tool for our analysis will be the linearisation of (1.1) with respect to its initial condition. Denote by Φ_t the (random) solution map to (1.1), so that $x_t = \Phi_t(x_0)$. It is then known that, under Assumption 4.2 below, Φ_t is almost surely a smooth map for every t. We actually obtain a flow of smooth maps, namely a two-parameter family of maps $\Phi_{s,t}$ such that $x_t = \Phi_{s,t}(x_s)$ for every $s \le t$ and such that $\Phi_{t,u} \circ \Phi_{s,t} = \Phi_{s,u}$ and $\Phi_t = \Phi_{0,t}$. For a given initial condition x_0, we then denote by $J_{s,t}$ the derivative of $\Phi_{s,t}$ evaluated at x_s. Note that the chain rule immediately implies that one has the composition law $J_{s,u} = J_{t,u} J_{s,t}$, where the product is given by simple matrix multiplication. We also use the notation $J^{(k)}_{s,t}$ for the kth-order derivative of $\Phi_{s,t}$.

It is straightforward to obtain an equation governing $J_{0,t}$ by differentiating both sides of (1.1) with respect to x_0. This yields the non-autonomous linear equation

$$dJ_{0,t} = DV_0(x_t)\, J_{0,t}\, dt + \sum_{i=1}^{m} DV_i(x_t)\, J_{0,t} \circ dW_i(t) , \qquad J_{0,0} = I , \tag{4.1}$$

where I is the $n \times n$ identity matrix. Higher order derivatives $J^{(k)}_{0,t}$ with respect to the initial condition can be defined similarly.

Remark 4.1 For every $s > 0$, the quantity $J_{s,t}$ solves the same equation as (4.1), except for the initial condition which is given by $J_{s,s} = I$.

On the other hand, we can use (3.5) to, at least on a formal level, take the Malliavin derivative of the integral form of (1.1), which then yields for $r \leq t$ the identity

$$\mathscr{D}_r^j X(t) = \int_r^t DV_0(X_s)\, \mathscr{D}_r^j X_s\, ds + \sum_{i=1}^m \int_r^t DV_i(X_s)\, \mathscr{D}_r^j X_s \circ dW_i(s) + V_j(X_r)\,.$$

(Here we denote by $\mathscr{D}^j$ the Malliavin derivative with respect to W_j; the generalisation of the discussion of the previous section to the case of finitely many independent Wiener processes is straightforward.) We see that, save for the initial condition at time $t = r$ given by $V_j(X_r)$, this equation is identical to the integral form of (4.1)!

As a consequence, we have for $s < t$ the identity

$$\mathscr{D}_s^j X_t = J_{s,t} V_j(X_s)\,. \tag{4.2}$$

Furthermore, since X_t is independent of the later increments of W, we have $\mathscr{D}_s^j X_t = 0$ for $s \geq t$.

By the composition property $J_{0,t} = J_{s,t} J_{0,s}$, we can write $J_{s,t} = J_{0,t} J_{0,s}^{-1}$, which will be useful in the sequel. Here, the inverse $J_{0,t}^{-1}$ of the Jacobian can be found by solving the SDE

$$dJ_{0,t}^{-1} = -J_{0,t}^{-1}\, DV_0(x)\, dt - \sum_{i=1}^m J_{0,t}^{-1}\, DV_i(x) \circ dW_i\,. \tag{4.3}$$

This follows from the chain rule by noting that if we denote by $\Psi(A) = A^{-1}$ the map that takes the inverse of a square matrix, then we have $D\Psi(A)H = -A^{-1}HA^{-1}$.

This discussion is the motivation for the following assumption, which we assume to be in force from now on:

Assumption 4.2 *The vector fields V_i are $\mathcal{C}^\infty$ and all of their derivatives grow at most polynomially at infinity. Furthermore, they are such that the solutions to (1.1), (4.1) and (4.3) satisfy*

$$\mathbf{E} \sup_{t \leq T} |x_t|^p < \infty\,, \qquad \mathbf{E} \sup_{t \leq T} |J_{0,t}^{(k)}|^p < \infty\,, \qquad \mathbf{E} \sup_{t \leq T} |J_{0,t}^{-1}|^p < \infty\,,$$

for every initial condition $x_0 \in \mathbf{R}^n$, every terminal time $T > 0$, every $k > 0$, and every $p > 0$.

Remark 4.3 It is well-known that Assumption 4.2 holds if the V_i are bounded with bounded derivatives of all orders. However, this is far from being a necessary assumption.

Remark 4.4 Under Assumption 4.2, standard limiting procedures allow to justify (4.2), as well as all the formal manipulations that we will perform in the sequel.

With these assumptions in place, the version of Hörmander's theorem that we are going to prove in these notes is as follows:

Theorem 4.5 *Let $x_0 \in \mathbf{R}^n$ and let x_t be the solution to (1.1). If the vector fields $\{V_j\}$ satisfy the parabolic Hörmander condition and Assumption 4.2 is satisfied, then the law of X_t has a smooth density with respect to Lebesgue measure.*

Proof. Denote by $\mathscr{A}_{0,t}$ the operator $\mathscr{A}_{0,t}v = \int_0^t J_{s,t}V(X_s)v(s)\,ds$, where v is a square integrable, not necessarily adapted, $\mathbf{R}^m$-valued stochastic process and V is the $n \times m$ matrix-valued function obtained by concatenating the vector fields V_j for $j = 1, \dots, m$. With this notation, it follows from (4.2) that the Malliavin covariance matrix $\mathscr{M}_{0,t}$ of X_t is given by

$$\mathscr{M}_{0,t} = \mathscr{A}_{0,t}\mathscr{A}_{0,t}^* = \int_0^t J_{s,t}V(X_s)V^*(X_s)J_{s,t}^*\,ds\ .$$

It follows from (4.2) that the assumptions of Theorem 3.2 are satisfied for the random variable X_t, provided that we can show that $\|\mathscr{M}_{0,t}^{-1}\|$ has bounded moments of all orders. This in turn follows by combining Lemma 4.7 with Theorem 4.8 below. □

4.2 Proof of Hörmander's Theorem

The remainder of this section is devoted to a proof of the fact that Hörmander's condition is sufficient to guarantee the invertibility of the Malliavin matrix of a diffusion process. For purely technical reasons, it turns out to be advantageous to rewrite the Malliavin matrix as

$$\mathscr{M}_{0,t} = J_{0,t}\mathscr{C}_{0,t}J_{0,t}^*\ , \qquad \mathscr{C}_{0,t} = \int_0^t J_{0,s}^{-1}V(X_s)V^*(X_s)(J_{0,s}^{-1})^*\,ds\ ,$$

where $\mathscr{C}_{0,t}$ is the *reduced Malliavin matrix* of our diffusion process.

Remark 4.6 The reason for considering the reduced Malliavin matrix is that the process appearing under the integral in the definition of $\mathscr{C}_{0,t}$ is adapted to the filtration generated by W_t. This allows us to use some tools from stochastic calculus that would not be available otherwise.

Since we assumed that $J_{0,t}$ has inverse moments of all orders, the invertibility of $\mathscr{M}_{0,t}$ is equivalent to that of $\mathscr{C}_{0,t}$. Note first that since $\mathscr{C}_{0,t}$ is a positive definite symmetric matrix, the norm of its inverse is given by

$$\|\mathscr{C}_{0,t}^{-1}\| = \Big(\inf_{|\eta|=1}\langle \eta, \mathscr{C}_{0,t}\eta\rangle\Big)^{-1}\ .$$

A very useful observation is then the following:

Lemma 4.7 *Let M be a symmetric positive semidefinite $n \times n$ matrix-valued random variable such that $\mathbf{E}\|M\|^p < \infty$ for every $p \geq 1$ and such that, for every $p \geq 1$ there exists C_p such that*

$$\sup_{|\eta|=1} \mathbf{P}(\langle \eta, M\eta \rangle < \varepsilon) \leq C_p \varepsilon^p \,, \tag{4.4}$$

holds for every $\varepsilon \leq 1$. Then, $\mathbf{E}\|M^{-1}\|^p < \infty$ for every $p \geq 1$.

Proof. The non-trivial part of the result is that the supremum over η is taken outside of the probability in (4.4). For $\varepsilon > 0$, let $\{\eta_k\}_{k \leq N}$ be a sequence of vectors with $|\eta_k| = 1$ such that for every η with $|\eta| \leq 1$, there exists k such that $|\eta_k - \eta| \leq \varepsilon^2$. It is clear that one can find such a set with $N \leq C\varepsilon^{2-2n}$ for some $C > 0$ independent of ε. We then have the bound

$$\begin{aligned} \langle \eta, M\eta \rangle &= \langle \eta_k, M\eta_k \rangle + \langle \eta - \eta_k, M\eta \rangle + \langle \eta - \eta_k, M\eta_k \rangle \\ &\geq \langle \eta_k, M\eta_k \rangle - 2\|M\|\varepsilon^2 \,, \end{aligned}$$

so that

$$\begin{aligned} \mathbf{P}\Big(\inf_{|\eta|=1} \langle \eta, M\eta \rangle \leq \varepsilon\Big) &\leq \mathbf{P}\Big(\inf_{k \leq N} \langle \eta_k, M\eta_k \rangle \leq 4\varepsilon\Big) + \mathbf{P}\Big(\|M\| \geq \frac{1}{\varepsilon}\Big) \\ &\leq C\varepsilon^{2-2n} \sup_{|\eta|=1} \mathbf{P}\Big(\langle \eta, M\eta \rangle \leq 4\varepsilon\Big) + \mathbf{P}\Big(\|M\| \geq \frac{1}{\varepsilon}\Big) \,. \end{aligned}$$

It now suffices to use (4.4) for p large enough to bound the first term and Chebychev's inequality combined with the moment bound on $\|M\|$ to bound the second term. □

As a consequence of this, Theorem 4.5 is a corollary of:

Theorem 4.8 *Consider (1.1) and assume that Assumption 4.2 holds. If the corresponding vector fields satisfy the parabolic Hörmander condition then, for every initial condition $x \in \mathbf{R}^n$, we have the bound*

$$\sup_{|\eta|=1} \mathbf{P}(\langle \eta, \mathscr{C}_{0,1}\eta \rangle < \varepsilon) \leq C_p \varepsilon^p \,,$$

for suitable constants C_p and all $p \geq 1$.

Remark 4.9 The choice $t = 1$ as the final time is of course completely arbitrary. Here and in the sequel, we will always consider functions on the time interval $[0, 1]$.

Before we turn to the proof of this result, we introduce a very useful notation which, to the best of the author's knowledge, was first used in [HM11]. Given a family $A = \{A_\varepsilon\}_{\varepsilon \in (0,1]}$ of events depending on some parameter $\varepsilon > 0$, we say that A is "almost true" if, for every $p > 0$ there exists a constant C_p such that

$\mathbf{P}(A_\varepsilon) \geq 1 - C_p\varepsilon^p$ for all $\varepsilon \in (0,1]$. Similarly for "almost false". Given two such families of events A and B, we say that "A almost implies B" and we write $A \Rightarrow_\varepsilon B$ if $A \setminus B$ is almost false. It is straightforward to check that these notions behave as expected (almost implication is transitive, finite unions of almost false events are almost false, etc). Note also that these notions are unchanged under any reparametrisation of the form $\varepsilon \mapsto \varepsilon^\alpha$ for $\alpha > 0$. Given two families X and Y of real-valued random variables, we will similarly write $X \leq_\varepsilon Y$ as a shorthand for the fact that $\{X_\varepsilon \leq Y_\varepsilon\}$ is "almost true".

Before we proceed, we state the following useful result, where $\|\cdot\|_\infty$ denotes the L^∞ norm and $\|\cdot\|_\alpha$ denotes the best possible α-Hölder constant.

Lemma 4.10 *Let $f\colon [0,1] \to \mathbf{R}$ be continuously differentiable and let $\alpha \in (0,1]$. Then, the bound*

$$\|\partial_t f\|_\infty = \|f\|_1 \leq 4\|f\|_\infty \max\Big\{1, \|f\|_\infty^{-\frac{1}{1+\alpha}} \|\partial_t f\|_\alpha^{\frac{1}{1+\alpha}}\Big\}$$

holds, where $\|f\|_\alpha$ denotes the best α-Hölder constant for f.

Proof. Denote by x_0 a point such that $|\partial_t f(x_0)| = \|\partial_t f\|_\infty$. It follows from the definition of the α-Hölder constant $\|\partial_t f\|_{\mathcal{C}^\alpha}$ that $|\partial_t f(x)| \geq \frac{1}{2}\|\partial_t f\|_\infty$ for every x such that $|x - x_0| \leq (\|\partial_t f\|_\infty/2\|\partial_t f\|_{\mathcal{C}^\alpha})^{1/\alpha}$. The claim then follows from the fact that if f is continuously differentiable and $|\partial_t f(x)| \geq A$ over an interval I, then there exists a point x_1 in the interval such that $|f(x_1)| \geq A|I|/2$. □

With these notations at hand, we have the following statement, which is essentially a quantitative version of the Doob-Meyer decomposition theorem. Originally, it appeared in [Nor86], although some form of it was already present in earlier works. The statement and proof given here are slightly different from those in [Nor86], but are very close to them in spirit.

Lemma 4.11 *Let W be an m-dimensional Wiener process and let A and B be $\mathbf{R}$ and $\mathbf{R}^m$-valued adapted processes such that, for $\alpha = \frac{1}{3}$, one has $\mathbf{E}(\|A\|_\alpha + \|B\|_\alpha)^p < \infty$ for every p. Let Z be the process defined by*

$$Z_t = Z_0 + \int_0^t A_s\, ds + \int_0^t B_s\, dW(s)\,. \tag{4.5}$$

Then, there exists a universal constant $r \in (0,1)$ such that one has

$$\{\|Z\|_\infty < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \{\|A\|_\infty < \varepsilon^r\} \,\&\, \{\|B\|_\infty < \varepsilon^r\}\,.$$

Proof. Recall the exponential martingale inequality [RY99, p. 153], stating that if M is any continuous martingale with quadratic variation process $\langle M\rangle(t)$, then

$$\mathbf{P}\Big(\sup_{t\leq T}|M(t)| \geq x \quad \& \quad \langle M\rangle(T) \leq y\Big) \leq 2\exp(-x^2/2y)\,,$$

for every positive T, x, y. With our notations, this implies that for any $q < 1$ and any adapted process F, one has the almost implication

$$\{\|F\|_\infty < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \left\{ \left\| \int_0^\cdot F_t \, dW(t) \right\|_\infty < \varepsilon^q \right\} . \tag{4.6}$$

With this bound in mind, we apply Itô's formula to Z^2, so that

$$Z_t^2 = Z_0^2 + 2\int_0^t Z_s \, A_s \, ds + 2\int_0^t Z_s \, B_s \, dW(s) + \int_0^t B_s^2 \, ds \, . \tag{4.7}$$

Since $\|A\|_\infty \leq_\varepsilon \varepsilon^{-1/4}$ (or any other negative exponent for that matter) by assumption and similarly for B, it follows from this and (4.6) that

$$\{\|Z\|_\infty < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \left\{ \left| \int_0^1 A_s \, Z_s \, ds \right| \leq \varepsilon^{\frac{3}{4}} \right\} \ \& \ \left\{ \left| \int_0^1 B_s \, Z_s \, dW(s) \right| \leq \varepsilon^{\frac{2}{3}} \right\} .$$

Inserting these bounds back into (4.7) and applying Jensen's inequality then yields

$$\{\|Z\|_\infty < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \left\{ \int_0^1 B_s^2 \, ds \leq \varepsilon^{\frac{1}{2}} \right\} \quad \Rightarrow \quad \left\{ \int_0^1 |B_s| \, ds \leq \varepsilon^{\frac{1}{4}} \right\} .$$

We now use the fact that $\|B\|_\alpha \leq_\varepsilon \varepsilon^{-q}$ for every $q > 0$ and we apply Lemma 4.10 with $\partial_t f(t) = |B_t|$ (we actually do it component by component), so that

$$\{\|Z\|_\infty < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \{\|B\|_\infty \leq \varepsilon^{\frac{1}{17}}\} ,$$

say. In order to get the bound on A, note that we can again apply the exponential martingale inequality to obtain that this "almost implies" the martingale part in (4.5) is "almost bounded" in the supremum norm by $\varepsilon^{\frac{1}{18}}$, so that

$$\{\|Z\|_\infty < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \left\{ \left\| \int_0^\cdot A_s \, ds \right\|_\infty \leq \varepsilon^{\frac{1}{18}} \right\} .$$

Finally applying again Lemma 4.10 with $\partial_t f(t) = A_t$, we obtain that

$$\{\|Z\|_\infty < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \{\|A\|_\infty \leq \varepsilon^{1/80}\} ,$$

and the claim follows with $r = 1/80$. □

Remark 4.12 By making α arbitrarily close to $1/2$, keeping track of the different norms appearing in the above argument, and then bootstrapping the argument, it is possible to show that

$$\{\|Z\|_\infty < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \{\|A\|_\infty \leq \varepsilon^p\} \ \& \ \{\|B\|_\infty \leq \varepsilon^q\} ,$$

for p arbitrarily close to $1/5$ and q arbitrarily close to $3/10$. This seems to be a very small improvement over the exponent $1/8$ that was originally obtained in [Nor86], but is certainly not optimal either. The main reason why our result is suboptimal is that we move several times back and forth between L^1, L^2, and L^∞ norms. (Note furthermore that our result is not really comparable to that in [Nor86], since Norris used L^2 norms in the statements and his assumptions were slightly different from ours.)

We now have all the necessary tools to prove Theorem 4.8:

Proof of Theorem 4.8. We fix some initial condition $x_0 \in \mathbf{R}^n$ and some unit vector $\eta \in \mathbf{R}^n$. With the notation introduced earlier, our aim is then to show that

$$\{\langle \eta, \mathscr{C}_{0,1}\eta\rangle < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \phi\,, \tag{4.8}$$

or in other words that the statement $\langle \eta, \mathscr{C}_{0,1}\eta\rangle < \varepsilon$ is "almost false". As a shorthand, we introduce for an arbitrary smooth vector field F on $\mathbf{R}^n$ the process Z_F defined by

$$Z_F(t) = \langle \eta, J_{0,t}^{-1} F(x_t)\rangle\,,$$

so that

$$\langle \eta, \mathscr{C}_{0,1}\eta\rangle = \sum_{k=1}^m \int_0^1 |Z_{V_k}(t)|^2\,dt \geq \sum_{k=1}^m \Big(\int_0^1 |Z_{V_k}(t)|\,dt\Big)^2\,. \tag{4.9}$$

The processes Z_F have the nice property that they solve the stochastic differential equation

$$dZ_F(t) = Z_{[F,V_0]}(t)\,dt + \sum_{i=1}^m Z_{[F,V_k]}(t) \circ dW_k(t)\,, \tag{4.10}$$

which can be rewritten in Itô form as

$$dZ_F(t) = \Big(Z_{[F,V_0]}(t) + \sum_{k=1}^m \frac{1}{2} Z_{[[F,V_k],V_k]}(t)\Big)\,dt + \sum_{i=1}^m Z_{[F,V_k]}(t)\,dW_k(t)\,. \tag{4.11}$$

Since we assumed that all derivatives of the V_j grow at most polynomially, we deduce from the Hölder regularity of Brownian motion that, provided that the derivatives of F grow at most polynomially fast, Z_F does indeed satisfy the assumptions on its Hölder norm required for the application of Norris's lemma. The idea now is to observe that, by (4.9), the left hand side of (4.8) states that Z_F is "small" for every $F \in \mathscr{V}_0$. One then argues that, by Norris's lemma, if Z_F is small for every $F \in \mathscr{V}_k$ then, by considering (4.10), it follows that Z_F is also small for every $F \in \mathscr{V}_{k+1}$. Hörmander's condition then ensures that a contradiction arises at some stage, since $Z_F(0) = \langle F(x_0), \xi\rangle$ and there exists k such that $\mathscr{V}_k(x_0)$ spans all of $\mathbf{R}^n$.

Let us make this rigorous. It follows from Norris's lemma and (4.11) that one has the almost implication

$$\{\|Z_F\|_\infty < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \{\|Z_{[F,V_k]}\|_\infty < \varepsilon^r\}\ \&\ \{\|Z_G\|_\infty < \varepsilon^r\}\,,$$

for $k = 1, \dots, m$ and for $G = [F, V_0] + \frac{1}{2}\sum_{k=1}^m [[F, V_k], V_k]$. Iterating this bound a second time, this time considering the equation for Z_G, we obtain that

$$\{\|Z_F\|_\infty < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \{\|Z_{[[F,V_k],V_\ell]}\|_\infty < \varepsilon^{r^2}\}\,,$$

so that we finally obtain the implication

$$\{\|Z_F\|_\infty < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \{\|Z_{[F,V_k]}\|_\infty < \varepsilon^{r^2}\}\,, \tag{4.12}$$

for $k = 0, \dots, m$.

At this stage, we are basically done. Indeed, combining (4.9) with Lemma 4.10 as above, we see that

$$\{\langle \eta, \mathscr{C}_{0,1}\eta\rangle < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \{\|Z_{V_k}\|_\infty < \varepsilon^{1/5}\}\,.$$

Applying (4.12) iteratively, we see that for every $k > 0$ there exists some $q_k > 0$ such that

$$\{\langle \eta, \mathscr{C}_{0,1}\eta\rangle < \varepsilon\} \quad \Rightarrow_\varepsilon \quad \bigcap_{V \in \mathscr{V}_k} \{\|Z_V\|_\infty < \varepsilon^{q_k}\}\,.$$

Since $Z_V(0) = \langle \eta, V(x_0)\rangle$ and since there exists some $k > 0$ such that $\mathscr{V}_k(x_0) = \mathbf{R}^n$, the right hand side of this expression is empty for some sufficiently large value of k, which is precisely the desired result. □

References

[BH07] F. BAUDOIN and M. HAIRER. A version of Hörmander's theorem for the fractional Brownian motion. *Probab. Theory Related Fields* **139**, no. 3-4, (2007), 373–395.

[Bis81a] J.-M. BISMUT. Martingales, the Malliavin calculus and Hörmander's theorem. In *Stochastic integrals (Proc. Sympos., Univ. Durham, Durham, 1980)*, vol. 851 of *Lecture Notes in Math.*, 85–109. Springer, Berlin, 1981.

[Bis81b] J.-M. BISMUT. Martingales, the Malliavin calculus and hypoellipticity under general Hörmander's conditions. *Z. Wahrsch. Verw. Gebiete* **56**, no. 4, (1981), 469–505.

[BT05] F. BAUDOIN and J. TEICHMANN. Hypoellipticity in infinite dimensions and an application in interest rate theory. *Ann. Appl. Probab.* **15**, no. 3, (2005), 1765–1777.

[Cas09] T. CASS. Smooth densities for solutions to stochastic differential equations with jumps. *Stochastic Process. Appl.* **119**, no. 5, (2009), 1416–1435.

[CF10] T. CASS and P. FRIZ. Densities for rough differential equations under Hörmander's condition. *Ann. of Math. (2)* **171**, no. 3, (2010), 2115–2141.

[HM06] M. HAIRER and J. C. MATTINGLY. Ergodicity of the 2D Navier-Stokes equations with degenerate stochastic forcing. *Ann. of Math. (2)* **164**, no. 3, (2006), 993–1032.

[HM11] M. HAIRER and J. C. MATTINGLY. A theory of hypoellipticity and unique ergodicity for semilinear stochastic PDEs. *Electr. Jour. Probab.* To appear.

[HN05] B. HELFFER and F. NIER. *Hypoelliptic estimates and spectral theory for Fokker-Planck operators and Witten Laplacians*, vol. 1862 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2005.

[Hör67] L. HÖRMANDER. Hypoelliptic second order differential equations. *Acta Math.* **119**, (1967), 147–171.

[Hör85] L. HÖRMANDER. *The Analysis of Linear Partial Differential Operators I–IV*. Springer, New York, 1985.

[HP11] M. HAIRER and N. PILLAI. Ergodicity of hypoelliptic SDEs driven by fractional Brownian motion. *Ann. IHP Ser. B* To appear.

[IK06] Y. ISHIKAWA and H. KUNITA. Malliavin calculus on the Wiener-Poisson space and its application to canonical SDE with jumps. *Stochastic Process. Appl.* **116**, no. 12, (2006), 1743–1769.

[Koh78] J. J. KOHN. Lectures on degenerate elliptic problems. In *Pseudodifferential operator with applications (Bressanone, 1977)*, 89–151. Liguori, Naples, 1978.

[KS84] S. KUSUOKA and D. STROOCK. Applications of the Malliavin calculus. I. In *Stochastic analysis (Katata/Kyoto, 1982)*, vol. 32 of *North-Holland Math. Library*, 271–306. North-Holland, Amsterdam, 1984.

[KS85] S. KUSUOKA and D. STROOCK. Applications of the Malliavin calculus. II. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **32**, no. 1, (1985), 1–76.

[KS87] S. KUSUOKA and D. STROOCK. Applications of the Malliavin calculus. III. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **34**, no. 2, (1987), 391–442.

[Mal78] P. MALLIAVIN. Stochastic calculus of variations and hypoelliptic operators. *Proc. Intern. Symp. SDE* 195–263.

[MP06] J. C. MATTINGLY and É. PARDOUX. Malliavin calculus for the stochastic 2D Navier-Stokes equation. *Comm. Pure Appl. Math.* **59**, no. 12, (2006), 1742–1790.

[Nor86] J. NORRIS. Simplified Malliavin calculus. In *Séminaire de Probabilités, XX, 1984/85*, vol. 1204 of *Lecture Notes in Math.*, 101–130. Springer, Berlin, 1986.

[Nua95] D. NUALART. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, New York, 1995.

[Oco88] D. OCONE. Stochastic calculus of variations for stochastic partial differential equations. *J. Funct. Anal.* **79**, no. 2, (1988), 288–331.

[RY99] D. REVUZ and M. YOR. *Continuous martingales and Brownian motion*, vol. 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third ed., 1999.

[SV72] D. W. STROOCK and S. R. S. VARADHAN. On the support of diffusion processes with applications to the strong maximum principle. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. III: Probability theory*, 333–359. Univ. California Press, Berkeley, Calif., 1972.

[Tak02] A. TAKEUCHI. The Malliavin calculus for SDE with jumps and the partially hypoelliptic problem. *Osaka J. Math.* **39**, no. 3, (2002), 523–559.

[Tak10] A. TAKEUCHI. Bismut–Elworthy–Li-type formulae for stochastic differential equations with jumps. *Journal of Theoretical Probability* **23**, (2010), 576–604.