STOCHASTIC DIFFERENTIAL EQUATIONS AND HYPOELLIPTIC OPERATORS

Denis R. Bell
Department of Mathematics
University of North Florida
Jacksonville, FL 32224
U. S. A.

1. Introduction
The first half of the twentieth century saw some remarkable developments in analytic probability theory. Wiener constructed a rigorous mathematical model of Brownian motion. Kolmogorov discovered that the transition probabilities of a diffusion process define a fundamental solution to the associated heat equation. Itô developed a stochastic calculus which made it possible to represent a diffusion with a given (infinitesimal) generator as the solution of a stochastic differential equation. These developments created a link between the fields of partial differential equations and stochastic analysis whereby results in the former area could be used to prove results in the latter.

More specifically, let $X_0, \ldots, X_n$ denote a collection of smooth vector fields on $\mathbb{R}^d$, regarded also as first-order differential operators, and define the second-order differential operator

$$L \equiv \frac{1}{2} \sum_{i=1}^n X_i^2 + X_0.$$  \hspace{1cm} (1.1)

Consider also the Stratonovich stochastic differential equation (sde)

$$d\xi_t = \sum_{i=1}^n X_i(\xi_t) \circ dw_i + X_0(\xi_t)dt$$  \hspace{1cm} (1.2)

where $w = (w_1, \ldots, w_n)$ is an $n$–dimensional standard Wiener process. Then the solution $\xi_t$ is a time-homogeneous Markov process with generator $L$, whose transition probabilities $p(t, x, dy)$ satisfy the following PDE (known as the Kolmogorov forward equation) in the weak sense

$$\frac{\partial p}{\partial t} = L^*_p.$$  

A differential operator $G$ is said to be hypoelliptic if, whenever $Gu$ is smooth for a distribution $u$ defined on some open subset of the domain of $G$, then $u$ is smooth.

In 1967, Hörmander proved that an operator of the form $L$ in (1.1) is hypoelliptic if the Lie algebra generated by $X_0, \ldots, X_n$ has dimension $d$ at each point*. It follows immediately that the transition probabilities $p(t, x, dy)$ have smooth densities at all positive times provided the parabolic operator $\partial/\partial t - L^*_p$ satisfies Hörmander’s condition at $\xi_0$. Of course, this is a rather circuitous route for proving a result of a decidedly probabilistic nature, a fact that had hardly escaped the notice of PDE theorists and stochastic analysts alike.

In the mid-seventies Malliavin ([Ma1], [Ma2]) outlined a new and innovative method for directly proving the smoothness of the transition probabilities $p$. The idea is as follows: the measure $p(t, x, dy)$ is the image of Wiener measure under the Itô

* This hypothesis will be referred to as Hörmander’s (Lie bracket) condition.
map $g : w \mapsto \xi_t$. Now the Wiener measure, which can be considered as an infinite-dimensional analogue of standard Gaussian measure on Euclidean space, has a well-understood analytic structure. If the map $g$ were smooth, then regularity properties of $p$ could be studied by a process of integration by parts (cf. Section 2). In fact, this is not the case; $g$ is only defined up to a set of full Wiener measure and is most pathological from the standpoint of classical calculus. Malliavin solved this problem by constructing an extended calculus applicable to solution maps of sde’s. His method (which has since been termed Malliavin calculus) was based on the infinite-dimensional Ornstein-Uhlenbeck semigroup and was rather elaborate. It has since been simplified and extended by many authors and has become a powerful tool in stochastic analysis. Following further pioneering work by Kusuoka and Stroock, it has led to a complete probabilistic proof of Hörmander’s theorem, as Malliavin originally intended, and has inspired a host of other results. Examples include filtering theorems by Michel [Mi], a deeper understanding of the Skorohod integral and the development of an anticipating stochastic calculus by Nualart and Pardoux [NP], an extension of Clark’s formula by Ocone [O], and Bismut’s probabilistic analysis of the small-time asymptotics of the heat kernel of the Dirac operator on a Riemannian manifold [Bi2]. In this article, I describe my own contributions to the subject. The interested reader is encouraged to also study the aforementioned works.

The contents of the article are as follows. In Section 2, we present an elementary derivation of Malliavin’s integration by parts formula by which one establishes smoothness of the transition probabilities discussed above. We also include a result of Bismut that illustrates how this formula is related to Hörmander’s condition. Much of the notation, together with the basic tools that will be used later, are introduced in this section of the article.

The material presented in Sections 3 and 4 is joint work with S. Mohammed. In Section 3, we prove a sharp generalization of Hörmander’s theorem. This result asserts that the operator $L$ in (1.1) is hypoelliptic under hypotheses that allow Hörmander’s Lie bracket condition to fail on hypersurfaces in the domain of $L$. The theorem establishes the hypoellipticity of a large class of operators with points of infinite-type, and is sharp in the category of Hörmander operators with smooth coefficients.

In Section 4, we establish sufficient conditions for the existence of smooth densities for a class of stochastic functional equations of the form

$$dx_t = g(x_{t-})dw + B(t, x)dt$$

where $g$ denotes a matrix-valued function, $r$ is a positive time-delay, and $B$ is a non-anticipating functional defined on the space of paths. Our hypotheses allow degeneracy of $g$ on a hypersurface in the ambient space of the equation. Probabilistically, there is an important difference between the solution $\xi$ to equation (1.2) and the solution $x$
to (1.3). Namely, while ξ is a Markov process, x is non-Markov. As a consequence, the aforementioned result cannot be proved by appealing to existing results in PDE theory.

Finally, in Section 5 we describe some open problems, which we hope will stimulate further study into this interesting subject.

2. Integration by parts and the regularity of induced measures

As in Section 1, let ξ denote the solution to the sde (1.2). In this section we will give an elementary proof of a key result in Malliavin’s paper [Ma1], which gives a criterion under which the random variables ξ_t have absolutely continuous distributions at all positive times t. The material in this section is taken from the author’s doctoral dissertation [Be1]. Following Malliavin, we will make use the following result from harmonic analysis (see [Ma1] for a proof).

**Lemma 2.1.** Let μ denote a finite Borel measure on \( \mathbb{R}^d \). Suppose that for \( k \in \mathbb{N} \) and each set of non-negative integers \( d_1, \ldots, d_k \), there exists a constant \( C \) such that for all test (\( C^\infty \) compact support) functions \( \phi \) on \( \mathbb{R}^d \)

\[
\left| \int_{\mathbb{R}^d} \frac{\partial d_1 + \ldots + d_k}{\partial x_1^{d_1} \ldots \partial x_d^{d_k}} \phi \, d\mu \right| \leq C ||\phi||_\infty.
\]

(2.1)

If this condition holds for \( k = 1 \), then the measure \( \nu \) is absolutely continuous with respect to the Lebesgue measure on \( \mathbb{R}^d \). If the condition holds for every \( k \geq 1 \), then \( \nu \) has a smooth density.

The following result is included in order to motivate the approach that follows. It treats a finite-dimensional analogue of the infinite-dimensional problem that will be studied later.

**Theorem 2.2.** Suppose \( T : \mathbb{R}^m \mapsto \mathbb{R}^d \) is a \( C^2 \) map. Let \( d\gamma(x) = (2\pi)^{-m/2} \times \exp(-|x|^2/2)dx \) denote the standard Gaussian measure on \( \mathbb{R}^m \). Let \( \nu \equiv T(\gamma) \) denote the induced measure on \( \mathbb{R}^d \), \( \nu(B) \equiv \gamma(T^{-1}(B)) \) for Borel subsets \( B \subseteq \mathbb{R}^d \). Define

\[
N \equiv \{ x \in \mathbb{R}^d / DT(x) \in L(\mathbb{R}^m, \mathbb{R}^d) \text{ is non surjective} \}.
\]

Then the condition

\[
\gamma(N) = 0
\]

(2.2)

is necessary and sufficient for absolute continuity of the measure \( \nu \).
Proof. The necessity of condition (2.2) follows immediately from Sard’s theorem which asserts that the set $T(N)$ has zero Lebesgue measure.

To prove sufficiency, we argue as follows. Define a $d \times d$ matrix

$$\sigma(x) = DT(x)DT(x)^*.$$  \hspace{1cm} (2.3)

Let $\{\psi_k\}$ denote a sequence of smooth bump functions from $[0, \infty) \rightarrow [0, 1]$ such that
(i) $\psi_k(t) = 1$ if $0 \leq t \leq k$
(ii) $\psi_k(t) = 0$ if $t \geq k + 1$
(iii) For each $p \geq 1$, $\sup_k |D^p \psi_k(t)| < \infty$.

For each $k \in \mathbb{N}$, define $R_k : \mathbb{R}^d \otimes \mathbb{R}^d \rightarrow [0, 1]$ by

$$R_k(\alpha) = \begin{cases} \psi_k(||\alpha^{-1}||), & \alpha \in GL(d) \\ 0, & \alpha \notin GL(d). \end{cases}$$

Note that $R_k$ is $C^\infty$, since $GL(d)$ is an open subset of $\mathbb{R}^d \otimes \mathbb{R}^d$. We also define sequences of measures $\{\gamma_k\}$ on $C_0$ and $\{\nu_k\}$ on $\mathbb{R}^d$ by

$$\frac{d\gamma_k}{d\gamma}(x) = R_k(\sigma(x)), \quad \nu_k = T(\gamma_k).$$

Assume now that (2.2) holds. Since $N$ is precisely the set on which $\sigma$ is degenerate, this is equivalent to the condition $\sigma(x) \in GL(d)$ a.s., which implies $\nu_k \rightarrow \nu$ in variation norm. Hence it suffices to prove absolute continuity of $\nu_k$ for each $k$. We do this as follows: let $e_1, \ldots, e_d$ denote the standard basis of $\mathbb{R}^d$ and define, for any $1 \leq i \leq d$, a function $h_i : \mathbb{R}^m \rightarrow \mathbb{R}^m$ by

$$h_i(x) = \begin{cases} DT(x)^*\sigma^{-1}(x)e_i, & \sigma \in GL(d) \\ 0, & \sigma \notin GL(d). \end{cases}$$

Let $\phi$ be a test function on $\mathbb{R}^d$. By definition of $h_i$, we have

$$\int_{\mathbb{R}^d} \frac{\partial \phi}{\partial x_i} \; d\nu_k = \int_{\mathbb{R}^m} D\phi(T(x)e_i\psi_k(||\sigma^{-1}(x)||))d\gamma(x)$$

$$= \int_{\mathbb{R}^m} D(\phi \circ T)(x)h_i(x)\psi_k(||\sigma^{-1}(x)||)d\gamma(x).$$ \hspace{1cm} (2.4)

Define the (Gaussian) divergence operator $\text{Div}$ acting on smooth functions $G : \mathbb{R}^n \rightarrow \mathbb{R}^n$ by

$$\text{Div}(G)(x) = <G(x), x> - \text{Trace } DG(x).$$

Integrating by parts in (2.4) yields

$$\int_{\mathbb{R}^d} \frac{\partial \phi}{\partial x_i} \; d\nu_k = \int_{\mathbb{R}^m} \phi \circ T(x)X_i(x)d\gamma(x)$$
where
\[ X_i = \text{Div}[\psi_k(||\sigma^{-1}||h_i)] \]  
\[ (2.5) \]

Since \( X_i \) is a continuous function with compact support, it follows that \( X_i \in L^1(\gamma) \). Thus (2.5) yields
\[ \int_{\mathbb{R}^d} \frac{\partial \phi}{\partial x_i} d\nu_k \leq ||\phi||_{\infty} ||X_i||_{L^1(\gamma)} \]
and the absolute continuity of \( \nu_k \) follows from Lemma 2.1.

With the above finite-dimensional case as motivation, we now move to the heart of the matter. Let \( \gamma \) denote the Wiener measure, defined on the space \( C_0 \) of continuous paths \( \{ w : [0, 1] \mapsto \mathbb{R}^n/ w(0) = 0 \} \). Recall that we wish to study the law \( \nu \) of a random variable \( \xi_t \), for \( t > 0 \), where \( \xi \) is the solution to a stochastic differential equation. Let \( A \) and \( B \) denote bounded smooth maps from \( \mathbb{R}^d \) into \( \mathbb{R}^d \otimes \mathbb{R}^n \) and \( \mathbb{R}^d \) respectively with bounded derivatives of all orders, and let \( x \in \mathbb{R}^d \). As before, let \( w = (w_1, \ldots, w_n) \) denote a standard Wiener process and consider the Itô sde
\[ \xi_t = x + \int_0^t A(\xi_s)dw_s + \int_0^t B(\xi_s)ds, \ t \geq 0. \]  
\[ (2.6) \]
Denoting by \( g \) the map \( w \mapsto \xi \) and by \( g_t \) the composition of \( g \) with evaluation at time \( t > 0 \), we have \( \nu = g_t(\gamma) \). We seek conditions under which the measure \( \nu \) is absolutely continuous with respect to the Lebesgue measure on \( \mathbb{R}^d \). The setting looks similar to that of Theorem 2.2. There are, however, two important differences:

(i) the measure \( \gamma \) is defined on an infinite-dimensional vector space
(ii) the map \( g \) (and hence \( g_t \)) is non-differentiable in the classical sense.

Point (i) can be handled without too much difficulty. Although there is nothing like a Lebesgue measure on \( C_0 \), the usual backdrop for an integration by parts calculation, the Wiener measure has a well-developed analytic structure. In particular, there are formulae for integration by parts on Wiener space*. The second point is more serious. Malliavin developed his stochastic calculus of variations in [Ma1] in order to address this.

We shall adopt here a more elementary approach, based on the following observation. There is a Hilbert subspace \( H \subset C_0 \) of paths, known as the Cameron-Martin space, canonically associated with the Wiener measure. The space \( H \) is the set of Lebesgue a.e. differentiable paths \( h \) such that
\[ \int_0^1 |h'(t)|^2 dt < \infty. \]
equipped with the inner product
\[ < h, k >_H \equiv \int_0^1 < h'(t), k'(t) > dt, \ h, k \in H. \]

* See, for example, Gross’ divergence theorem for abstract Wiener spaces [G].
The point is that, when restricted to $H$, the map $g$ becomes entirely regular. In fact, it is intuitively clear that the restriction* of $g$ to $H$, which we denote by $\tilde{g}$, is the map $h \in H \mapsto k$ defined by the ordinary integral equation

$$k_t = x + \int_0^t A(k_s)h'_s ds + \int_0^t B(k_s)ds, \quad t \geq 0$$

and it is easily seen that the map $\tilde{g}$ possesses the same degree of smoothness as the coefficient functions $A$ and $B$. For example, an equation for the derivative $\eta \equiv D_r \tilde{g}(h)$, $r \in H$, is obtained by formally differentiating in (2.7) with respect to $h$. Thus $\eta$ satisfies

$$\eta_t = \int_0^t \{DA(k_s)(\eta_s, h'_s) + A(k_s)r'_s + DB(k_s)\eta_s\} ds.$$  

Integral equations for higher order derivatives of $\tilde{g}$ can be similarly obtained.

The foregoing remarks provide an alternative elementary approach to Malliavin’s work. As before, let $\phi$ denote a test function on $\mathbb{R}^d$ and let $\tilde{g}_t$ denote the composition of $\tilde{g}$ with evaluation at time $t$. Suppose that for each standard basis vector $e_i$, one can construct a sequence of paths $h^m_i$ such that

$$D \tilde{g}_t(P_m w)h^m_i = e_i.$$  

Applying the dominated convergence theorem and integrating by parts with respect to the measure $\gamma$, we will then have

$$\int_{\mathbb{R}^d} \frac{\partial \phi}{\partial x_i} d\nu = \lim_m \int_{C_0} D\phi(\tilde{g}_t(P_m w))e_i d\gamma$$

$$= \lim_m \int_{C_0} D(\phi \circ \tilde{g}^m_t)(P_m w)h^m_i d\gamma$$

$$= \lim_m \int_{C_0} \phi(\tilde{g}_t(P_m w))Div[h^m_i] d\gamma$$

If we can show that the sequence $\{Div[h^m_i]\}_m$ is bounded in $L^1(\gamma)$, then we obtain

$$\left| \int_{\mathbb{R}^d} \frac{\partial \phi}{\partial x_i} d\nu \right| \leq ||\phi||_\infty \sup_m ||Div[h^m_i]||_{L^1(\gamma)}$$

* The term “restriction” requires clarification because $g$ is only defined up to a set of $\gamma$-measure 1, and $\gamma(H) = 0$. It is more correct to say that $g$ stochastically extends $\tilde{g}$ in the following sense. For each $m \in \mathbb{N}$, define $P_m : C_0 \mapsto H$ to be the operator that piecewise linearizes on the uniform partition $[0, 1/m, 2/m \ldots, 1]$. It can be shown that, as $m \to \infty$, $P_m$ converges strongly to the identity map on $C_0$, and $\tilde{g}(P_m w) \to g(w)$ a.s.
and the absolute continuity of $\nu$ follows from Lemma 2.1 as before.

We will actually carry out a modified version of the above procedure, using a sequence of piecewise linear approximations to $g$. Being finite-rank operators, these approximations finite-dimensionalize the differential analysis in the problem. Thus, at each level of approximation, we need only perform an elementary integration by parts in Euclidean space, as in the proof of Theorem 2.2.

Our approximation scheme is as follows. For each $m \in \mathbb{N}$ and $w \in C_0$, let $\Delta_j w (= \Delta_j^m w)$ denote $w(j + 1) t/m) - w(tj/m)$. Define $v_0, v_{t/m}, \ldots, v_t \in \mathbb{R}^d$ inductively by

$$v_{kt/m} = x + \sum_{j=0}^{k-1} A(v_{jt/m}) \Delta_j w + \frac{t}{m} \sum_{j=0}^{k-1} B(v_{jt/m}), \quad k = 1, \ldots, m.$$  \hspace{1cm} (2.8)

Let $v^m : [0, 1] \to \mathbb{R}^d$ denote the path piecewise linear between the points $(kt/m, v_{kt/m})$, $k = 0, \ldots, m$ and constant on $[t, 1]$.

It is easy to see that for each $m$, the map $g^m : C_0 \to H$ is $C^\infty$. We prove in [Be2, pp 35 - 37]

**Theorem 2.3.** For every $p \in \mathbb{N}$,

$$\lim_{m \to \infty} \sup_{s \in [0, t]} E[|\xi_s - v^m_s|^p] = 0$$

where $\xi$ is the solution to the sde (2.6). We now define an analogue of the matrix $\sigma$ appearing in the proof of Theorem 2.2. Let $\sigma^m(w) \equiv Dg^m(w)Dg^m(w)^* \in \mathbb{R}^d \otimes \mathbb{R}^d$.

We prove in [Be2, pp 37 - 39]

**Theorem 2.4.** As $m \to \infty$, the matrix sequence $\sigma^m$ converges in probability to a limit $\sigma$ in $\mathbb{R}^d \otimes \mathbb{R}^d$. Let $I$ denote the $d \times d$ identity matrix and consider the $d \times d$ matrix-valued equations

$$Y_s = I + \int_0^s DA(\xi_u)(Y_u, dw_u) + \int_0^s DB(\xi_u)Y_u du$$

and

$$Z_s = I - \int_0^s Z_u DA(\xi_u)(, dw_u) - \int_0^s Z_u DB(\xi_u) du.$$  \hspace{1cm} (2.9)

Then

$$\sigma = Y_t \left[ \int_0^t Z_s A(\xi_s)A(\xi_s)^* Z_s^* ds \right] Y_t^*.$$  \hspace{1cm} (2.9)

where $^*$ denotes matrix transpose.
Remarks

The matrix processes $Y_t$ and $Z_t$ have a natural interpretation in terms of the stochastic flow of the SDE (2.6), i.e. the random map on $\mathbb{R}^d$, $\phi_t: x \mapsto \xi_t$. It can be shown that $\phi_t$ is a.s. a $C^\infty$ map and an easy computation shows that $Y_t = D\phi_t$. Furthermore, $Y_t \in GL(d)$ for all $t \geq 0$ and $Z_t = Y_t^{-1}$.

The matrix $\sigma$ defined in (2.9) is called the Malliavin covariance matrix associated to the random variable $\xi_t$. The next theorem shows that establishing non-degeneracy of this matrix is the key to proving regularity of the distribution of $\xi_t$, as might be suspected from Theorem 2.2 and its proof.

**Theorem 2.5 (Malliavin).** Suppose $\sigma \in GL(d)$ a.s. Then the random variable $\xi_t$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^d$.

Our proof of Theorem 2.5 will make use of the following technical result, which is tailor-made for the estimations we will need to do later (see [Be2, pp34-35] for its proof).

**Lemma 2.6.** Suppose that $X, U_0, \ldots, U_{m-1}$ are $\mathbb{R}^d$-valued random variables, $V_0, \ldots, V_{m-1}$ and $Y_0, \ldots, Y_{m-1}$ are random linear maps from $\mathbb{R}^n$ to $\mathbb{R}^d$, and $Z_0, \ldots, Z_{m-1}$ are random bilinear maps from $\mathbb{R}^d \times \mathbb{R}^n$ to $\mathbb{R}^d$, satisfying the following conditions

(i) For all $0 \leq i \leq m-1$, $U_i, V_i, Y_i$, and $Z_i$ are measurable with respect to $F_{i+m}$, where $\{F_t\}$ is the filtration generated by $\{w_t\}$.

(ii) $\max \{||X||_p, ||U_i||_p, ||V_i||_p, ||Y_i||_p, ||Z_i||_p, 0 \leq i \leq m-1\} \leq M$, where $||.||_p$ denotes the $L^p$ norms of the (norms of) the various quantities in their respective spaces.

Let $\eta_{kt/m}, 0 \leq k \leq m$, be random variables satisfying the equations

$$
\eta_{kt/m} = X + \frac{1}{m} \sum_{j=0}^{k-1} U_j + \frac{1}{m} \sum_{j=0}^{k-1} V_j(\Delta_j w) + \frac{1}{m} \sum_{j=0}^{k-1} Y_j(\eta_{jt/m}) + \sum_{j=0}^{k-1} Z_j(\eta_{jt/m}, \Delta_j w).
$$

Then there exists a constant $N$, depending only on $M$ and $p$, such that

$$
||\eta_{kt/m}||_p \leq N, \quad \forall \ 0 \leq k \leq m.
$$

**Proof of Theorem 2.5** Let $V_m$ denote the finite-dimensional subspace of $C_0$ consisting of paths that are piecewise linear between the times $0, t/m, \ldots, t$, and constant on $[t, 1]$. Following the method used to prove Theorem 2.2, we define

$$
\frac{d\gamma_k}{d\gamma}(w) = R_k(\sigma(w)), \quad \nu_k = g_t(\gamma_k)
$$

where $R_k$ are as defined previously. As before, the assumption $\sigma \in GL(d)$, a.s. implies $\nu_k \to \nu$ in variation, so it suffices to prove that each $\nu_k$ is absolutely continuous.
Let $e$ denote any unit vector in $\mathbb{R}^d$ and define $h^m : C_0 \to H$ by

$$h_i(w) = \begin{cases} 
Dg_t^m(w)\sigma^m(w)^{-1}e, & \sigma^m \in GL(d) \\
0, & \sigma^m \notin GL(d).
\end{cases}$$

Arguing as before and integrating by parts with respect to the measures $P_m(\gamma)$ (note these are Gaussian measures on the finite-dimensional vector spaces $V_m$) yields

$$\int_{\mathbb{R}^d} D_y\phi \, dv_k = \lim_{m \to \infty} \int_{C_0} \phi \circ g_m^m \text{Div} \left[h^m R_k \circ \sigma^m\right] \, d\gamma$$

where $\text{Div}$ is now defined with respect to the Cameron-Martin space $H$

$$\text{Div} G(w) = \langle G(w), w > H - \text{Trace}_H DG(w).$$

To complete the proof, we must thus show that

$$\sup_m E|\text{Div} [h^m R_k \circ \sigma^m]| < \infty \quad (2.10)$$

First consider the inner product term in the divergence. This is non-zero only if $\sigma^m \in GL(d)$ and $||\sigma^m||^{-1} \leq k + 1$. In this case

$$E|<h^m R_k \circ \sigma^m, w > H| = E|<(\sigma^m)^{-1}e, Dg_t^m(w)w>| \leq (k + 1)E|\eta^m_t| \quad (2.11)$$

where $\eta^m = Dg_t^m(w)w$ satisfies the equation

$$\eta^m_{kt/m} = \sum_{i=0}^{k-1} \left\{ A(v_{it/m})\Delta_i w + DA(v_{it/m})(\eta^m_{it/m}) \right\} + t/m DB(v_{it/m}) \eta^m_{it/m},$$

$k = 1, \ldots, m$.

Since $A, DA$, and $DB$ are bounded, Lemma 2.6 implies that $E|\eta^m_t|$ is bounded and it follows from (2.11) that $E|<h^m R_k \circ \sigma^m, w > H|$ is bounded.

It remains to show that the same holds for the second term in the divergence, i.e.

$$\sup_m E|\text{Trace}_H D[h^m R_k \circ \sigma^m]| < \infty. \quad (2.12)$$

Let $f_1, \ldots, f_n$ be an orthonormal basis of $\mathbb{R}^n$. For $1 \leq r \leq n$ and $0 \leq l \leq m - 1$, define $f^{rl} \in H$ by

$$f^{rl}_s = \begin{cases} 
0, & 0 \leq s < lt/m \\
\sqrt{m/t}(s - lt/m)f_r, & lt/m \leq s < (l + 1)t/m \sqrt{t/m}f_r, \\
(l + 1)t/m < s \leq 1
\end{cases}$$
The set $S_m = \{ f^l, 1 \leq r \leq n, 0 \leq l \leq m-1 \}$ is orthonormal in $H$, thus can be extended to an orthonormal basis $B_m$ of $H$. Note that for any $f \in B_m \cap S_m, Dg^m(w)f = 0$. Evaluating the Trace in (2.12) on the basis $B_m$ gives

$$\text{Trace}_H D[h^m R_k \circ \sigma^m] = \sum_{r=1,l=0}^{n,m-1} < D[h^m R_k \circ \sigma^m](w)f^l, f^r >$$

$$= \sum_{r=1,l=0}^{n,m-1} \left\{ R_k \circ \sigma^m(w) \left[ < (\sigma^m)^{-1}e, D^2 g^m_l(w)(f^l, f^r) > 
- < (\sigma^m)^{-1}D\sigma^m(w)f^r(\sigma^m)^{-1}e, Dg^m_l(w)f^l > 
+ D R_k(\sigma^m)D\sigma^m(w)f^r < (\sigma^m)^{-1}e, Dg^m_l(w)f^l > \right] \right\}.$$ 

In view of the definition of $R_k$, it suffices to show that

$$\sup_m E \left| \sum_{l=0}^{m-1} D^2 g^m_l(w)(f^l, f^l) \right| < \infty$$ (2.13)

and

$$\sup_m E \left[ \sum_{l=0}^{m-1} |D\sigma^m(w)f^l| \times |Dg^m_l(w)f^l| \right] < \infty.$$ (2.14)

Let $\eta^l$ denote the path $\sqrt{m}Dg^m_l(w)f^l$. Differentiation in (2.8) yields $\eta^l_{jt/m} = 0$ if $j \leq l$ and

$$\eta^l_{jt/m} = \sqrt{t}A(v_{lt/m}) f_t + \sum_{p=l}^{j-1} DA(v_{pt/m}) (\eta^l_{pt/m}, \Delta_p w) + \frac{t}{m} \sum_{p=l}^{j-1} DB(v_{pt/m}) \eta^l_{pt/m}, j \geq l+1$$

It follows from Lemma 2.6 that

$$\sup_{j,k,m} ||\eta^k_{jt/m}||_4 < \infty$$ (2.15)

Now let $\rho^l$ denote the path $mD^2 g^m_l(w)(f^l, f^l)$. This satisfies the equation

$$\rho^l_{jt/m} = t DA(v_{lt/m}) (\eta^l_{lt/m}, f_t)$$

$$+ \sum_{p=l}^{j-1} \left\{ t D^2 A(v_{pt/m}) (\eta^l_{pt/m}, \eta^l_{pt/m}, \Delta_p w) + DA(v_{pt/m}) (\rho^l_{pt/m}, \Delta_p w) \right\}$$

$$+ \frac{1}{m} \sum_{p=l}^{j-1} \left\{ t D^2 B(v_{pt/m}) (\eta^l_{pt/m}, \eta^l_{pt/m}) + DB(v_{pt/m}) \rho^l_{pt/m} \right\}, j \geq l+1$$

11
\((ρ_{jt/m}^l = 0 \text{ for } j ≤ l)\). Lemma 2.6 together with (2.15) now give

\[
\sup_{r,j,k,m} \|ρ_{jt/m}^r \|_4 < \infty
\]

and this implies (2.13). Condition (2.14) can be established by a similar argument. With this, the proof of the theorem is complete.

We now return to the sde (1.2) defined in terms of vector fields. It follows from (2.9) that the Malliavin covariance matrix \(σ \) for \(ξ_t \) now has the form

\[
σ = Y_t \sum_{i=1}^n \int_0^t [Z_s X_i(ξ_s)] \otimes [Z_s X_i(ξ_s)]^* ds Y_t^*
\]

(2.16)

where \(Z \) satisfies is the \(d \times d \) matrix-valued sde

\[
Z_s = I - \sum_{i=1}^n \int_0^s Z_u DX_i(ξ_u) \circ dw_i - \int_0^s Z_u DX_0(ξ_u) du
\]

(2.17)

As before, \(Y_t \) is the inverse matrix of \(Z_t \). Note that the stochastic integral in (2.17) is of Stratonovich type.

The following result and its proof (which I first learned from Bismut’s paper [Bi1]) makes transparent the relationship between Hörmander’s Lie bracket condition and the non-degeneracy of \(σ \).

**Theorem 2.7.** Suppose the Lie algebra generated by the vector fields \(X_1, \ldots, X_n \) span \(\mathbb{R}^d \) at \(ξ_0 \). Then \(σ \in GL(d), a.s. \) (hence, by Theorem 2.5 \(ξ_t \) is absolutely continuous with respect to Lebesgue measure on \(\mathbb{R}^d \)). The proof will require the following

**Lemma 2.8.** Suppose \(y \in \mathbb{R}^d, B \) is a smooth vector field on \(\mathbb{R}^d \) and \(τ \) is a stopping time such that

\[
< Z_s B(ξ_s), y > = 0, \forall s ∈ [0, τ].
\]

Then for \(i = 1, \ldots, n, \)

\[
< Z_s [X_i, B](ξ_s), y > = 0, \forall s ∈ [0, τ].
\]

**Proof.** Applying Itô’s formula and using (2.17), we obtain, for \(s ∈ [0, τ], \)

\[
0 = d < (Z_s B(ξ_s)), y >
\]

\[
= (Z_s [B, X_0] ds + \sum_{i=1}^n Z_s [B, X_i](ξ_s) \circ dw_i, y).
\]

(2.19)
Converting the Stratonovich integrals to Itô form, we see that the infinitesimal term in (2.19) has the form
\[ G(s)ds + \sum_{i=1}^{n} Z_s[B, X_i](\xi_s)dw_i \]
for some continuous adapted process \( G \). Thus
\[ < G(s)ds + \sum_{i=1}^{n} Z_s[B, X_i](\xi_s)dw_i, y > = 0, \forall s \in [0, \tau]. \]
The conclusion now follows from the computational rules of Itô calculus: \( dw_i dt = 0 \), while \( dw_i dw_j = \delta_{ij} dt \).

**Proof of Theorem 2.7.** Define \( V \) to be the set of vectors
\[ \{ X_i(\xi_0), [X_i, X_j](\xi_0), [X_i, [X_j, X_k]](\xi_0), \ldots, 1 \leq i, j, k, \ldots, \leq n \}. \]
We will show that \( V \subseteq \sigma \), which clearly implies the result. For each \( 0 < s < t \), define
\[ R_s = \text{span}\{ Z_u X_i(\xi_u/0 \leq u \leq s, 0 \leq i \leq n) \} \]
and
\[ R = \cap_{0 < s < t} R_s. \]
Then \( R_t = \text{Range} \sigma \). Furthermore by the Blumenthal zero-one law, there exists a deterministic set \( \tilde{R} \) such that \( R = \tilde{R} \) a.s. Suppose \( y \in \tilde{R}^\perp \). Then with probability 1, there exists \( \tau > 0 \) such that \( R_s = \tilde{R} \) for all \( s \in [0, \tau] \). Then for \( i = 1, \ldots, n \)
\[ < Z_s X_i(\xi_s), y > = 0, \forall s \in [0, \tau]. \quad (2.20) \]
Iterating Lemma 2.8 on (2.20) gives
\[ < Z_s X_i(\xi_s), y >, < Z_s [X_i, X_j](\xi_s), y >, < Z_s [[X_i, X_j], X_k](\xi_s), y >, \ldots = 0 \]
for \( s \in [0, \tau] \), and setting \( s = 0 \) shows that \( y \in (\text{span} V)^\perp \). Thus span \( V \subseteq \tilde{R} \subseteq R_t = \text{Range} V \), as required.

In [Ma1], Malliavin proves further

**Theorem 2.9.** If
\[ (\det \sigma) \in \cap_{p \geq 1} L^p(\gamma) \quad (2.21) \]
then the density of \( \xi_t \) is \( C^\infty \).
This is also proved in [Be1], by iterating the argument used to prove Theorem 2.5. Kusuoka and Stroock prove in [KS2] that (2.21) holds under Hörmander’s general condition and they use this to give a complete probabilistic proof of Hörmander’s theorem. We do not include this work here since we will prove a more general result in the next section.
3. A Hörmander theorem for infinitely degenerate operators

The material in this section comes from [BM3]. Let $X_0, \ldots, X_n$ denote smooth vector fields defined on an open subset $D$ of $\mathbb{R}^d$. As before, we consider the second-order differential operator

$$L \equiv \frac{1}{2} \sum_{i=1}^{n} X_i^2 + X_0.$$  \hfill (3.1)

Let $\text{Lie}(X_0, \ldots, X_n)$ be the Lie algebra generated by the vector fields $X_0, \ldots, X_n$. According to the theorem of Hörmander ([H], Theorem 1.1), $L$ is hypoelliptic on $D$ if the vector space $\text{Lie}(X_0, \ldots, X_n)(x)$ has dimension $d$ at every $x \in D$. It can be shown that this is a necessary condition for hypoellipticity for operators of the form (3.1) with analytic coefficients.

This is not the case if the vector fields $X_0, \ldots, X_n$ defining $L$ are allowed to be (smooth) non-analytic. This fact was strikingly illustrated by Kusuoka and Stroock, who studied differential operators on $\mathbb{R}^3$ of the form

$$L_a \equiv \frac{\partial^2}{\partial x_1^2} + a^2(x_1) \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2}.$$  

They assume $a$ to be a $C^\infty$ real-valued even function, non-decreasing on $[0, \infty)$, which vanishes (only) at zero. It is shown in ([KS2], Theorem 8.41) that $L_a$ is hypoelliptic on $\mathbb{R}^3$ if and only if $a$ satisfies the condition $\lim_{a \to 0^+} s \log a(s) = 0$. In particular, in the case $a(s) = \exp(-|s|^p)$, $L_a$ is hypoelliptic provided $p \in (-1, 0)$. However, it is clear that any such operator fails to satisfy Hörmander’s condition on the hyperplane $x_1 = 0$.

In this section we present a sharp criterion for hypoellipticity that implies Hörmander’s theorem and encompasses the class of superdegenerate hypoelliptic operators of Kusuoka and Stroock.

We introduce the following notation. For a positive integer $m$, let $E^{(m)}$ denote any matrix whose columns consist of $X_0, \ldots, X_n$ together with all (iterated) Lie brackets of the form

$$[X_{i_1}, X_{i_2}]_{i_1, i_2=0}^n ; \cdots ; [X_{i_1}, [X_{i_2}, [X_{i_3}, \cdots, [X_{i_m}, X_{i_m}]]] \cdots]_{i_1, i_2, \ldots, i_m=0}^n ,$$

For $x \in D$ and $m \geq 1$, define

$$\mu^{(m)} \equiv \text{smallest eigenvalue of } [E^{(m)} E^{(m)*}].$$

Observe that $\mu^{(m)}(x) > 0$ for some $m \geq 1$ if and only if Hörmander’s (general) condition holds for the operator $L$ at $x \in D$. In this case we will say that $x$ is a Hörmander point for $L$. We denote the set of all such points by $H$ (note that $H$ is an open subset of $D$). The set $D \cap H^c$ of non-Hörmander points of $L$ will be denoted simply by $H^c$ in the sequel.

The main result of this section is the following
Theorem 3.1. Suppose the non-Hörmander set $H^c$ of $L$ is contained in a $C^2$ hypersurface $S$. Assume that at every point $x$ in $H^c$

(i) at least one of the vector fields $X_1, \cdots, X_n$ is transversal to $S$.

(ii) There exists an integer $m \geq 1$, an open neighborhood $U$ of $x$, and an exponent $p \in (-1, 0)$ such that

$$\mu^{(m)}(y) \geq \exp\{-[\rho(y, S)]^p\}, \quad \forall y \in U$$

(3.3)

where $\rho(y, S)$ denotes the Euclidean distance between $y$ and $S$.

Then $L$ is hypoelliptic on $D$.

We note that hypotheses (ii) in Theorem 3.1 controls the rate at which the Hörmander condition fails in a neighborhood of non-Hörmander points. It is clear that some such condition is necessary, since the Kusuoka-Stroock result cited above shows that the operator $L_{-1}$ is non-hypoelliptic if $p \leq -1$. Furthermore, the non-hypoellipticity of the operator $L_{-1}$ shows that the allowed range $(-1, 0)$ for $p$ in (3.3) is optimal. Hypothesis (ii) has the following probabilistic interpretation in terms of the diffusion process $\xi_t$ corresponding to $L$. It implies that if $\xi$ starts at a non-Hörmander point $x$ on $S$, then it will escape from $x$ at a fast enough rate to acquire a non-singular distribution.

One can see that a hypothesis such as (i) is also necessary for the hypoellipticity of $L$ (at least in the case where $X_0 = 0$) by looking at the probabilistic picture. For if $X_i$ is tangential to $S$ for each $i = 1, \ldots, n$ then, started from a point $x \in S$, $\xi$ will stay on $S$. Hence $\xi_t$ will not have a density in $\mathbb{R}^d$ at positive times $t$, which implies that $L$ cannot be hypoelliptic (as we remarked earlier hypoellipticity of $L$ implies the existence of a density for $\xi_t$, $t > 0$).

The ideas underlying the proof of Theorem 3.1 are as follows. In Section 2 we introduced the Malliavin covariance matrix $\sigma$ corresponding to the random variable $\xi_t$ and showed that non-degeneracy of $\sigma$ implies the existence of a density for $\xi_t$. Kusuoka and Stroock have proved that if the inverse moments of $\sigma$ do not explode too quickly as $t \downarrow 0$, then the parabolic operator $L + \partial/\partial t$ is hypoelliptic, they also showed how to deduce hypoellipticity of $L$ from this. More specifically, let $\sigma$ denote the matrix defined in (2.16) and (2.17) (recall that the process $\xi$ in these formulas is defined by equation (1.2)) and let $\Delta(t, \xi_0)$ denote the determinant of $\sigma$. The following result is proved in [KS2].

Let $D$ be an open set in $\mathbb{R}^d$. Suppose that for every $q \geq 1$ and every $x$ in $D$, there exists a neighborhood $V \subseteq D$ of $x$ such that

$$\lim_{t \to 0^+} t \log \left( \sup_{y \in V} \|\Delta(t, y)^{-1}\|_q \right) = 0. \quad (3.4)$$
Then the differential operator \( L + \frac{\partial}{\partial t} \) is hypoelliptic on \( \mathbb{R} \times D \). We will prove that (3.4) is satisfied under a (parabolic version of) the hypotheses of Theorem 3.1. There are two stages to this argument:

(a) a local parameterization \( \phi \) of the hypersurface \( S \) is introduced. Throughout this section, let \( \xi \) denote the diffusion process defined in (1.2). The quantity \( \phi(\xi_t) \) measures the distance between \( \xi_t \) and \( S \). We obtain probabilistic lower bounds on the \( L^q \)-norms of the paths \( \phi(\xi) \). These lower bounds are asymptotically sharp as \( q \to \infty \).

(b) We study the way the lower bounds in (a) are degraded under hypothesis (ii) of Theorem 3.1. This allows us to obtain sharp lower bounds on \( \Delta \) from which we are able to verify condition (3.4).

The proof of Theorem 3.1 uses basic stochastic analytic tools, e.g. Itô’s formula, Girsanov’s theorem, and the time-change theorem for stochastic integrals. In particular, this work establishes a precise connection between the maximal class of hypoelliptic operators of the form (1.1) and the space-time scaling property of the Wiener process. This provides new insight into hypoellipticity that is not available through a classical analysis of the problem.

Before proving Theorem 3.1, we state and prove a parabolic version of the theorem. To this end, denote by \( F(m) \) the matrix obtained by deleting the column \( X_0 \) from the matrix \( E(m) \) defined on Page 14 and define \( \lambda(m) \) to be the smallest eigenvalue of the matrix \( [F(m)F(m)^*] \). Define \( K \equiv \{ x \in D / \lambda(m)(x) > 0 \} \).

Suppose the set \( K^c \) is contained in a \( C^2 \) hypersurface \( N \) of \( D \). Assume that at every point in \( K^c \), at least one of the vector fields \( X_1, \cdots, X_n \) is transversal to \( N \). Assume further that for every \( x \in K^c \), there exists an integer \( m \geq 1 \), an open neighborhood \( U \) of \( x \), and \( p \in (-1, 0) \) such that \( \lambda(m)(y) \geq \exp\{ -[\rho(y,N)]^p \} \) for all \( y \in U \). Then the operator \( L + \frac{\partial}{\partial t} \) is hypoelliptic on \( \mathbb{R} \times D \).

The proof of Theorem 3.3 will require a definition and several preliminary results which we now state.

**Definition.** A non-negative random variable \( T \) is exponentially positive if there exist positive constants \( c_1 \) and \( c_2 \) (which we will refer to as the characteristics of \( T \)) such that

\[
P(T < \epsilon) < e^{-c_1/\epsilon}
\]

for all \( \epsilon \in (0, c_2) \).

We will make frequent use of the following well-known result ([IW], Lemma 10.5).

**Lemma 3.4.** Let \( y : [0, 1] \times \Omega \to \mathbb{R}^d \) be an Itô process of the form

\[
dy(t) = \sum_{i=1}^{n} a_i(t) dW_i(t) + b(t) dt, \quad 0 \leq t \leq 1,
\] (3.5)
where \( a_1, \ldots, a_n, b : [0,1] \times \Omega \to \mathbb{R}^d \) are measurable adapted processes, all bounded a.s. by a deterministic constant \( c_3 \). Let \( r > 0 \) and define

\[
\tau \equiv \inf \{ s > 0 : |y_s - y_0| = r \} \land 1.
\] (3.6)

Then \( \tau \) is an exponentially positive stopping time, and the characteristics of \( \tau \) depend only on \( r, c_3, n \) and \( d \).

The next two lemmas are central to our argument (in order to facilitate the exposition, we delay their proofs to the end of the section). The first yields sharp probabilistic lower bounds when applied to diffusion processes with at least one non-zero initial time diffusion coefficient.

**Lemma 3.5.** Let \( y : [0,1] \times \Omega \to \mathbb{R}^d \) be the Itô process in (3.5). Suppose that \( \tau \leq T \) is an exponentially positive stopping time such that at least one diffusion coefficient \( a_i \) satisfies the condition: a.s., \(|a_i(s)| \geq \delta\), for all \( 0 \leq s \leq \tau \), for some deterministic \( \delta > 0 \). Then for every \( m \geq 2 \), there exist positive constants \( c_4, c_5 \) and \( T_0 \) such that for all \( t \in (0,T_0) \) and \( \epsilon \in (0, c_4 t^{m+1}) \), the following holds

\[
P\left( \int_0^{t \land \tau} |y(u)|^m \, du < \epsilon \right) < \exp\left\{ -c_5 \epsilon^{-\frac{1}{m+1}} \right\}.
\] (3.7)

The constants \( c_4 \) and \( c_5 \) can be chosen to depend only on \( m, c_3, \delta, \) and the characteristics of \( \tau \). The constant \( T_0 \) depends only on the characteristics of \( \tau \).

The following result describes how the estimate (3.7) transforms under composition of the integrand with a function that vanishes at zero, at an appropriate exponential rate.

**Lemma 3.6.** Let \( \tau \) be an exponentially positive stopping time. Suppose \( y \) is an Itô process as in Lemma 3.4. Suppose further that \( y \) and \( \tau \) satisfy an estimate of the form (3.7) for some \( m > -\frac{p}{p+1} \), where \( p \in (-1,0) \). Then there exist positive constants \( T_1, c_6, c_7 \) and \( q > 1 \) such that for all \( t \in (0,T_1) \) and all \( \epsilon < \exp\{-c_6 t^{-\frac{1}{2}}\} \), the following holds

\[
P\left( \int_0^{t \land \tau} \exp(-|y_u|^p) \, du < \epsilon \right) < \exp\{ -c_7 |\log \epsilon|^q \}.
\] (3.8)

Furthermore, the constants \( T_1, c_6, c_7 \) and \( q \) are completely determined by \( c_3 \) in Lemma 3.4, \( c_4 \), \( c_5 \), and \( m \) in (3.7), \( p \), and the characteristics of \( \tau \).

Finally, we will need the following two technical Lemmas (since the proofs are straightforward we omit them)
Lemma 3.7. For every \( q \geq 1 \) and every bounded set \( V \subset R^d \) there exists a positive constant \( c_8 \) such that for all \( t \in (0, 1) \) and \( x \in V \)

\[
\|\Delta(t, x)^{-1}\|_{2q}^2 \leq c_8 \left\{ 1 + \sum_{j=1}^{\infty} P\left( Q(t, x) < j^{-\frac{1}{2q}} \right) \right\},
\]

(3.9)

where

\[
Q(t, x) \equiv \inf \left\{ \sum_{i=1}^{n} \int_{0}^{t} < Z_x^u X_i(\xi_x^u), h >^2 \ du : h \in R^d, |h| = 1 \right\}.
\]

(3.10)

Lemma 3.8. Suppose that the hypotheses of Theorem 3.3 are satisfied. Then for every \( x \in D \) there exists an integer \( m \geq 1 \) such that exactly one of the following two conditions holds:

(a) \( <\lambda^{(m)}(x) > 0. \)

(b) There exists an open neighborhood \( U \subseteq D \) of \( x \), a \( C^2 \) function \( \phi : U \rightarrow R \), and an exponent \( p \in (-1, 0) \) such that

(i) \( \phi(x) = 0 \) and \( \nabla\phi(x) \cdot X_i(x) \neq 0, \) for at least one \( i = 1, \ldots, n. \)

(ii) \( <\lambda^{(m)}(y) \geq \exp(-|\phi(y)|^p), \) for all \( y \in U. \)

We now assume the hypotheses and notations of Theorem 3.3. Without loss of generality, the vector fields \( X_0, \cdots, X_n \) may be supposed to be defined on the whole of \( R^d \) and to have compact support (this follows from a simple argument using a partition of unity and the fact that hypoellipticity is a local property). We will be assume this from now on.

Let \( x_0 \in D \) and choose \( m \) so that either (a) or (b) of Lemma 3.8 hold for \( x_0. \) Let \( t \in (0, 1) \) and suppose \( x \) lies in a fixed bounded neighborhood \( W \) of \( x_0. \) Define

\[
\tau_1 \equiv \inf \left\{ s > 0 : |\xi^s_x - x| \vee \|Z^s_x - I\| = 1/2 \right\} \wedge 1.
\]

(3.11)

By Lemma 3.4, \( \tau_1 \) is an exponentially positive stopping time with characteristics independent of \( x \in W. \)

Let \( S^d \equiv \{ h \in R^d : |h| = 1 \} \) denote the unit sphere in \( R^d. \) Suppose \( h \in S^d \) and \( \alpha = 1/18. \) Then

\[
P\left( \sum_{i=1}^{n} \int_{0}^{t \wedge \tau_1} < Z^u_x X_i(\xi^u_x), h >^2 \ du < \epsilon \right) \leq P(A \cap E) + P(A \cap E^c),
\]

where

\[
A \equiv \left( \sum_{i=1}^{n} \int_{0}^{t \wedge \tau_1} < Z^u_x X_i(\xi^u_x), h >^2 \ du < \epsilon \right)
\]

18
and

\[ E \equiv \left( \sum_{i=1}^{n} \int_{0}^{t \wedge \tau_1} \left( \sum_{j=1}^{n} <Z_u^x[X_i, X_j](\xi_u^x), h \right)^2 + \right. \]

\[ <Z_u^x\left\{ [X_i, X_0] + \frac{1}{2} \sum_{j,k=1}^{n} [X_i, [X_j, X_k]] \right\}(\xi_u^x), h \right)^2 \] \] \)

du < \epsilon^\alpha \),

By a lemma of Kusuoka-Stroock and Norris (cf. e.g. [B], Lemma 6.5; cf. [KS2], Theorem A.24), there exist positive constants \( c_9 \) and \( c_{10} \) such that

\[ P(A \cap E^c) \leq c_9 \exp(-c_{10} \epsilon^{-\alpha}). \]

The constants \( c_9 \) and \( c_{10} \) are independent of \( h \in S^d \). Note that \( E \subseteq F \cap G \), where

\[ F \equiv \left( \sum_{i,j=1}^{n} \int_{0}^{t \wedge \tau_1} <Z_u^x[X_i, X_j](\xi_u^x), h \right)^2 du < \epsilon^\alpha \) \]

\[ G \equiv \left( \sum_{i=1}^{n} \int_{0}^{t \wedge \tau_1} <Z_u^x[X_i, X_0] + \frac{1}{2} \sum_{j,k=1}^{n} [X_i, [X_j, X_k]] \right\}(\xi_u^x), h \right)^2 du < \epsilon^\alpha \). \]

Thus

\[ P\left( \sum_{i=1}^{n} \int_{0}^{t \wedge \tau_1} <Z_u^x[X_i, X_0](\xi_u^x), h \right)^2 du < \epsilon \) \leq c_9 \exp(-c_{10} \epsilon^{-\alpha}) + P(A \cap F \cap G). \] (3.12)

Applying a similar argument to \( P(A \cap F \cap G) \) gives

\[ P(A \cap F \cap G) \leq c_{11} \exp(-c_{12} \epsilon^{-\alpha^2}) + P(A \cap F \cap G \cap H) \] (3.13)

where

\[ H := \left( \sum_{i,j,k=1}^{n} \int_{0}^{t \wedge \tau_1} <Z_u^x[X_i, [X_j, X_k]](\xi_u^x), h \right)^2 du < \epsilon^{r_1} \). \]

It is easy to check that

\[ G \cap H \subseteq \left( \sum_{i=1}^{n} \int_{0}^{t \wedge \tau_1} <Z_u^x[X_i, X_0](\xi_u^x), h \right)^2 du < \epsilon^{r_1} \)

for some \( r_1 \in (0, 1) \) and sufficiently small \( \epsilon > 0 \).

Thus

\[ A \cap F \cap G \cap H \subseteq \]

\[ \left( \int_{0}^{t \wedge \tau_1} \left\{ <\sum_{i=1}^{n} (Z_u^x[X_i](\xi_u^x), h \right)^2 + \sum_{i,j=0}^{n} <Z_u^x[X_i, X_j](\xi_u^x), h \right)^2 \) du < \epsilon^{r_2} \) \]

19
for some \( r_2 \in (0, 1) \) and sufficiently small \( \epsilon > 0 \). Combining this with (3.12) and (3.13), one obtains

\[
P\left( \sum_{i=1}^{n} \int_{0}^{t} <Z_u^x X_i(\xi_u^x), h >^2 \, du < \epsilon \right) \leq c_9 \exp(-c_{10} \epsilon^{-\alpha}) + c_{11} \exp(-c_{12} \epsilon^{-\alpha^2})
\]

\[+ P\left( \int_{0}^{\tau_1} \left\{ \sum_{i=1}^{n} <Z_u^x X_i(\xi_u^x), h >^2 + \sum_{i,j=0}^{n} <Z_u^x [X_i, X_j](\xi_u^x), h >^2 \right\} \, du < \epsilon^2 \right).
\]

(3.14)

Iterating the argument used to derive (3.14) yields the following:

For each \( m \geq 1 \), there exist positive constants \( c_{13} \) and \( c_{14} \) and exponents \( r_3 \) and \( r_4 \in (0, 1) \), all independent of \( h \in S^d \), such that for all \( t \in (0, T) \), \( x \in W \), and \( \epsilon \in (0, c_{14}) \), one has

\[
P\left( \sum_{i=1}^{n} \int_{0}^{t} <Z_u^x X_i(\xi_u^x), h >^2 \, du < \epsilon \right) \leq \exp(-c_{14} \epsilon^{-r_3}) + P\left( \sum_{j=1}^{N} \int_{0}^{\tau_1} <Z_u^x K_j(\xi_u^x), h >^2 \, du < c_{17} \epsilon^{r_4} \right).
\]

(3.15)

Here the vector fields \( K_1, \ldots, K_N \) are the columns of the matrix function \( X^{(m)} \). Applying a straightforward compactness argument (cf.[B], Lemma 6.8) to (3.15), one obtains

\[
P(Q(t, x) < \epsilon) \leq \exp(-c_{15} \epsilon^{-r_3})
\]

\[+ c_{16} \epsilon^{-d} \sup \left\{ P\left( \sum_{j=1}^{N} \int_{0}^{\tau_1} <Z_u^x K_j(\xi_u^x), h >^2 \, du < c_{17} \epsilon^{r_4} \right) : |h| = 1 \right\}
\]

(3.16)

for \( \epsilon \in (0, c_{18}) \) and positive constants \( c_{15}, c_{16}, c_{17}, c_{18} \).

Since (3.8) implies \( \|Z^x(u) - I\| \leq 1/2 \), for all \( 0 \leq u \leq \tau_1 \), it is easy to deduce from (3.16) that

\[
P(Q(t, x) < \epsilon) \leq \exp(-c_{15} \epsilon^{-r_3}) + c_{16} \epsilon^{-d} P\left( \int_{0}^{\tau_1} \lambda^{(m)}(\xi_u^x) \, du < c_{16} \epsilon^{r_4} \right).
\]

(3.17)

We now consider each of the two cases (a) and (b) delineated in the conclusion of Lemma 3.8. Suppose first that (a) holds at \( x_0 \) for some \( m \geq 1 \). Then by continuity of \( \lambda^{(m)} \) there exist \( \rho > 0 \) and \( \delta > 0 \) such that

\[
\lambda^{(m)}(y) \geq \delta
\]

(3.18)
for all \( y \in B_\rho(x_0) \), where \( B_\rho(x_0) \) denotes the open ball in \( \mathbb{R}^d \) with center \( x_0 \) and radius \( \rho \). Let \( V \equiv B_{\rho/2}(x_0) \), assume \( x \in V \), and let \( \tau_2 \) denote the first exit time of \( \xi^x \) from \( V \). Then (3.17) and (3.18) imply

\[
P(Q(t,x) < \epsilon)
\]

\[
\leq \exp(-c_{15} \epsilon^{-\tau_3}) + c_{16} \epsilon^{-d} P\left( \tau_1 \wedge \tau_2 \wedge t < \frac{c_{18} \epsilon^{\tau_4}}{\delta} \right)
\]

\[
\leq c_{19} \exp(-c_{20} \epsilon^{-c_{21} r_5})
\]

provided \( t > \frac{c_{14} \epsilon^{\tau_4}}{\delta} \), where \( r_5 \equiv r_3 \wedge r_4 \) and \( c_{19}, c_{20} \) and \( c_{21} \) are positive constants, independent of \( (t,x) \in (0,T) \times V \). Substituting (3.20) into (3.9) yields, for every \( q \geq 1 \), the following inequality

\[
\|\Delta(t,x)^{-1}\|_{2q}^2 \leq c_8 \left\{ \left( \frac{\delta t}{c_{18}} \right)^{-\frac{2dq}{\tau_4}} + A(t) \right\},
\]

where

\[
A(t) \equiv 1 + \sum_{j=k}^\infty c_{19} \exp(-c_{20} j^r_8), \leq 1 + \sum_{j=1}^\infty c_{19} \exp(-c_{20} j^r_8) \leq \infty,
\]

where \( r_6 \equiv c_{21} r_5 / 2dq \) and \( k \) is the integer part of \((\delta t/c_{18})^{-2dq/r_4}\). We conclude that \( \|\Delta(t,x)^{-1}\|_{2q}\) grows no faster than a power of \( t \) as \( t \downarrow 0 \), uniformly with respect to \( x \in V \). Hence (3.4) is satisfied.

We now turn to the case where (b) of Lemma 3.8 holds at the point \( x_0 \). By the transversality condition (i) we may choose \( \rho > 0 \) small enough to ensure that \( B_\rho(x_0) \cap U \) and such that

\[
|\nabla \phi(x).X_i(x)| \geq \frac{1}{2} |\nabla \phi(x).X_i(x_0)| > 0
\]

for some \( 1 \leq i \leq n \) and every \( x \in B_\rho(x_0) \). Let \( V \equiv B_{\rho/2}(x_0) \). Assume \( x \in V \) and let \( \tau_3 \) denote the first exit time of \( \xi^x \) from \( B_{\rho/2}(x) \). In view of Lemma 3.4 (b)(ii), (3.17) implies

\[
P(Q(t,x) < \epsilon) \leq \exp(-c_{15} \epsilon^{-\tau_3}) + c_{16} \epsilon^{-d} P\left( \int_0^{t \wedge \tau_1 \wedge \tau_5} \exp(-|\eta^x_i|^p) \, du < c_{18} \epsilon^{\tau_4} \right)
\]

where \( \eta^x(t) \) denotes the process \( \phi(\xi^x(t)) \), \( t \leq \tau_3 \). Applying Itô’s lemma to compute \( \eta^x(t) \) gives

\[
d\eta^x(t) = \sum_{i=1}^n \nabla \phi(\xi^x(t)).X_i(\xi^x(t))dW_i(t) + (L - c)\phi(\xi^x(t)) \, dt.
\]

Lemma 3.8 (b)(i), Lemma 3.1, and (3.22) imply that the process \( y \equiv \eta^x \) and the stopping time \( \tau \equiv \tau_1 \wedge \tau_2 \) satisfy the hypotheses of Lemma 3.5. Hence (3.7) is satisfied.
for every \( m > 1 \) with \( \tau = \tau_1 \wedge \tau_3 \) and \( y = \eta^x \). Thus, by Lemma 3.6 there exist positive constants \( c_6, c_7, T_1 \) and \( q' > 1 \), all independent of \( x \in V \), such that for all \( t \in (0, T_1) \) and \( \epsilon < \exp( -c_6 t^{-q'} ) \)

\[
P\left( \int_0^{t \wedge \tau_1 \wedge \tau_3} \exp(-|\eta^x_u|^p) \, du < \epsilon \right) < \exp\{-c_7 |\log \epsilon|^{q'}\}.
\]  

(3.23)

Substituting this into (3.17) gives

\[
P(\Omega(t, x) < \epsilon) \leq \exp\{-c_{15} \epsilon^{-q_3} + c_{16} \epsilon^{-d} \exp(-c_7 |\log \epsilon|^{q'})\}.
\]  

(3.24)

for \( t \in (0, T_1) \) and \( \epsilon < \exp( -c_6 t^{-q'} ) \). Combining (3.21) with (3.5), we arrive at

\[
\| \Delta(t, x) \|_{L^2}^2 \leq c_8 \left\{ \exp(2d q_6 t^{-q'} ) + c_{22} \right\}, \quad 0 < t < T_1
\]  

(3.25)

where

\[
c_{22} \equiv 1 + \sum_{j=1}^{\infty} \left\{ \exp\left( -c_{15} j^{2q_4} \right) + c_{16} j^{1/2q} \exp(-c_7 |\log j|^{q'}) \right\} < \infty.
\]

Note that the constants \( c_6, c_8 \) and \( c_{22} \) can all be chosen to be independent of \( x \in V \). The right hand side of (3.25) explodes exponentially fast as \( t \downarrow 0 \). However, since \( q' > 1 \) we conclude that (3.4) holds also for this case, and the proof of Theorem 3.3 is complete.

**Proof of Theorem 3.1.**

Assume \( L \) satisfies the hypotheses of Theorem 3.1. We borrow yet another technique from Kusuoka and Stroock [KS2]. The idea is to imbed the operator \( L \) in another operator \( \tilde{L} \) defined on a \( (d + 1) \)-dimensional domain and satisfying the hypotheses of Theorem 3.3. This will prove the hypoellipticity of the parabolic operator \( \tilde{L} + \frac{\partial}{\partial s} \), which in turn will imply hypoellipticity of \( L \).

Choose a smooth non-negative real-valued function \( \rho \) on \( (0, 1) \) such that both \( \rho(s) \) and its derivative \( \rho'(s) \) are bounded away from zero for all \( s \in (0, 1) \). Define the operator \( \tilde{L} \) on the domain \( D \times (0, 1) \) by

\[
\tilde{L} \equiv \rho(s) L + \frac{1}{2} \frac{\partial^2}{\partial s^2}.
\]

Then \( \tilde{L} \) has the form

\[
\tilde{L} \equiv \frac{1}{2} \sum_{i=1}^{n+1} \tilde{X}_i^2 + \tilde{X}_0.
\]

where \( \tilde{X}_0(x, s) \equiv \rho(s) X_0(x) \), \( \tilde{X}_i(x, s) \equiv \rho(s)^{1/2} X_i(x), \quad 1 \leq i \leq n \), and \( \tilde{X}_{n+1}(x, s) \equiv \frac{\partial}{\partial s}, \quad (x, s) \in D \times (0, 1) \). Define \( \tilde{F}^{(m)}, \tilde{\lambda}^{(m)}, \tilde{H}, \) and \( \tilde{K} \) similarly to before, using the
vector fields $\tilde{X}_i$ in place of the $X_i$’s. Since $\rho$ and $\rho'$ are bounded away from zero on $(0,1)$, it is easy to see that there are positive constants $\delta_m$ such that
\[
\tilde{\lambda}^{(m)}(x,s) \geq \delta_m \mu^{(m)}(x), \quad \forall (x,s) \in D \times (0,1).
\] (3.26)

This implies $\tilde{K}^c \subseteq H^c \times (0,1)$. Since $H^c$ is contained in a $C^2$ hypersurface $S$ in $D$, then $\tilde{H}^c$ is contained in the $C^2$ hypersurface $S \times (0,1)$ in $D \times (0,1)$. By assumption, at least one the vector fields $X_1, \ldots, X_n$ is transversal to $S$ at every point of $H^c$. Hence one of the vectors fields $\tilde{X}_1, \ldots, \tilde{X}_{n+1}$ is transversal to $S \times (0,1)$ at every point of $\tilde{H}^c$. Hypothesis (ii) of Theorem 3.1 together with (3.26) imply that $\tilde{\lambda}^{(m)}$ satisfies the corresponding hypothesis in Theorem 3.3 (with respect to the set $\tilde{K}^c$). We conclude from Theorem 3.3 that the operator $\tilde{L} + \partial / \partial t$ is hypoelliptic on $\mathbb{R} \times D \times (0,1)$. Consequently $\tilde{L}$ is hypoelliptic on $D \times (0,1)$, and this implies $L$ is hypoelliptic on $D$.

We conclude this section by proving Lemmas 3.5 and 3.6, which provided the key steps in the preceding argument. The proof of Lemma 3.5 requires two preliminary results which we first state and prove.

**Proposition 3.9.** Suppose $m \geq 2$ and $a > 0$. Let $B : [0, \infty) \times \Omega \to \mathbb{R}$ be a one-dimensional Brownian motion. Then there exists a positive constant $c_{23}$ such that
\[
P\left(\int_0^a |B(u)|^m \, du < \epsilon\right) \leq \sqrt{2} \exp\left(-c_{23} a^{1+\frac{2}{m}} \epsilon^{-\frac{m}{2}}\right)
\]
for every $\epsilon > 0$. The constant $c_{23}$ may be chosen to be $2^{-7}$.

*Proof.* The result is known to hold for $m = 2$, with $c_{23} = 2^{-7}$ (cf. [IW], Lemma V.10.6, p. 399).

For $m > 2$ we apply Hölder’s inequality and use the result for $m = 2$ to obtain
\[
P\left(\int_0^a |B(u)|^m \, du < \epsilon\right) \leq P\left(\int_0^a |B(u)|^2 \, du < a^{1-\frac{2}{m}} \epsilon^{\frac{2}{m}}\right) \leq \sqrt{2} \exp(-c_{23} a^{1+\frac{2}{m}} \epsilon^{-2})
\]
for every $\epsilon > 0$. This proves the proposition.

**Proposition 3.10.** Assume the notation and hypotheses of Lemma 3.5. Then for every $m \geq 2$ and $q > 0$, there exist positive constants $c_{24}$ and $c_{25}$ such that for all $\epsilon > 0$,
\[
P\left(\int_0^\tau |y_n|^m \, du < \epsilon, \tau \geq \epsilon^q\right) < c_{24} \exp\left(-c_{25} \epsilon^{q + \frac{2(q-1)}{m}}\right).
\]
The constants $c_{24}$ and $c_{25}$ depend only on $c_3$ (the bound for the drift and diffusion coefficients of $y$), $\delta$, $m$, and $q$. In particular, they are independent of $y_0$ and the characteristics of $\tau$. 

23
Proof. Expressing (3.5) in components, it is sufficient to treat the case \( d = 1, n \geq 1 \). In this case, the process \( y \) may be written in the form

\[
y_t = B(\tau_4(t)) + \int_0^t b(u) \, du, \quad 0 \leq t \leq T,
\]

where

\[
\tau_4(t) = \int_0^t |a(u)|^2 \, du, \quad 0 \leq t \leq T,
\]

and \( B : [0, \infty) \times \Omega \to \mathbb{R} \) is a one-dimensional adapted Brownian motion started at \( B(0) = y_0 \). By assumption, we may take \(|a(u)| \geq \delta > 0\) for \( 0 \leq u \leq \tau \). Hence \( \tau \geq \epsilon^q \) implies \( \tau_4(\tau) \geq \delta^2 \epsilon^q \). Furthermore, the function \( \tau_4(t) \) is strictly increasing on \((0, \tau)_T\) and changes of the time variable yield

\[
\int_0^\tau |y_u|^m \, du \geq c_{26}^{-1} \int_0^{\tau_4(\tau)} |y(\tau_4^{-1}(s))|^m \, ds
\]

\[
= c_{26}^{-1} \int_0^{\tau_4(\tau)} \left| B(s) + \int_0^s \frac{b(\tau_4^{-1}(u))}{|a(\tau_4^{-1}(u))|^2} \, du \right|^m \, ds,
\]

where \( c_{26} \equiv nc_{23}^3 \). Thus

\[
P \left( \int_0^\tau |y_u|^m \, du < \epsilon, \tau \geq \epsilon^q \right)
\]

\[
\leq P \left( \int_0^{\delta^2 \epsilon^q} \left| B(s) + \int_0^s \frac{b(\tau_4^{-1}(u))}{|a(\tau_4^{-1}(u))|^2} \, du \right|^m \, ds < c_{26} \epsilon, \tau_4(\tau) \geq \delta^2 \epsilon^q \right) \tag{3.27}
\]

We now define a (bounded) process \( h : [0, \infty) \times \Omega \to \mathbb{R} \) by

\[
h(u) \equiv \begin{cases} 
\frac{b(\tau_4^{-1}(u))}{|a(\tau_4^{-1}(u))|^2}, & u \in (0, \tau_4(\tau)) \\
\frac{b(\tau)}{|a(\tau)|^2}, & u \geq \tau_4(\tau). 
\end{cases}
\]

and we denote by \( B' \) the process

\[
B'(s) \equiv B(s) + \int_0^s h(u) \, du, \quad 0 \leq s \leq \tau_4(T).
\]

By the Girsanov theorem, \( B' \) is a Brownian motion on \( \Omega \) with respect to the measure

\[
dP' \equiv \exp \left( - \int_0^{\tau_4(T)} h(u) \, dB(u) - \frac{1}{2} \int_0^{\tau_4(T)} h^2(u) \, du \right) \, dP.
\]

Denote by \( \Omega_\epsilon \) the event

\[
\Omega_\epsilon \equiv \left( \int_0^{\delta^2 \epsilon^q} |B'(s)|^m \, ds < c_{26} \epsilon \right).
\]
and by $G$ the Girsanov density

$$G \equiv \exp \left( -\int_0^{\tau_4(T)} h(u) dB(u) - \frac{1}{2} \int_0^{\tau_4(T)} h^2(u) du \right).$$

We now apply Hölder’s inequality to (3.27) to obtain

$$P \left( \int_0^\tau |y_u|^m du < \epsilon, \tau \geq \epsilon q \right) \leq P(\Omega_\epsilon) \leq \sqrt{E(G^{-2})} P'(\Omega_\epsilon).$$

By Proposition 3.9, we have

$$P'(\Omega_\epsilon) \leq \sqrt{2} \exp \left( -2c_{25}^q + 2(q-1)m \right),$$

(3.28)

where $c_{25} \equiv \frac{1}{2} c_{23}^{-\frac{q}{2}} \delta^2 (1+\frac{q}{2})$. The boundedness of $h$ and $\tau_4(T)$ imply the existence of a constant $c_{27}$, depending only on the bounds of the foregoing quantities, such that

$$G^{-2} \leq c_{27} \exp \left( 2 \int_0^{\tau_4(T)} h(u) dB(u) - 2 \int_0^{\tau_4(T)} h^2(u) du \right).$$

(3.29)

The desired conclusion follows from (3.28) and (3.29), together with the fact that the exponential on the right hand side of (3.29) is a Girsanov density (note that it is obtained by replacing $h$ by $(-2h)$ in the relation defining $G$), and therefore has expectation equal to 1.

**Proof of Lemma 3.2.**

Note that for every $q > 0$, we may write

$$P \left( \int_0^{t \wedge \tau} |y_u|^m du < \epsilon \right) \leq P_1 + P_2,$$

(3.30)

where

$$P_1 \equiv P \left( \int_0^{t \wedge \tau} |y_u|^m du < \epsilon, \ t \wedge \tau \geq \epsilon q \right)$$

and

$$P_2 \equiv P(t \wedge \tau < \epsilon q).$$

By Proposition 3.10,

$$P_1 < c_{24} \exp \left( -c_{25}^q + 2(q-1)m \right),$$

(3.31)

where $c_{24}$ and $c_{25}$ are independent of $t$. Now $\tau$ is exponentially positive; so if $T_0 > t > \epsilon q$, then

$$P_2 < \exp(-c_{27}^q),$$

(3.32)

25
where \( c_{27} \) and \( T_0 \) denote the characteristics of \( \tau \). Combining (3.30) - (3.32), we obtain

\[
P\left( \int_0^{t \wedge \tau} |y_u|^m \, du < \epsilon \right) \leq c_{24} \exp\left( -c_{25} \epsilon^{q + \frac{2(q-1)}{m}} \right) + \exp\left( c_{27} \epsilon^{-q} \right)
\]

for \( t \in (0, T_0) \) and \( 0 < \epsilon < t^{\frac{1}{2}} \). The lemma now follows by choosing as \( q \) the value for which the two exponents \( \left\{ q + \frac{2(q-1)}{m} \right\} \) and \( -q \) coincide, namely \( \frac{1}{m+1} \).

**Proof of Lemma 3.6.**

Choose and fix \( m \geq \max\left\{ -\frac{p}{1+p}, 2 \right\} \) and set \( q \equiv -\frac{m}{p(m+1)} \); so \( q > 1 \). Define a function \( \psi : [0, \infty) \to [0, 1) \) by

\[
\psi(z) \equiv \begin{cases} 
\exp(-z^m) & z > 0 \\
0 & z = 0.
\end{cases}
\]

Note that

(i) \( \psi \) is strictly increasing;

(ii) \( \psi \) is convex in an interval \((0, c_{28})\), for some positive constant \( c_{28} \).

Furthermore

\[
P\left( \int_0^{t \wedge \tau} \exp(-|y_u|^p) \, du < \epsilon \right) = P\left( \int_0^{t \wedge \tau} \psi(|y_u|^m) \, du < \epsilon \right).
\]

We break the proof of the lemma into two cases. Firstly, suppose that \( |y_0| \geq c_{29} \) \( (\equiv \left( \frac{1}{2} c_{28} \right)^{\frac{1}{m}}) \). Let \( \tau_5 \equiv \inf\{ s > 0 : |y_s - y_0| = \frac{1}{2} c_{29} \} \wedge \tau \). Then there exists a positive constant \( c_{30} \), determined by \( m \), \( p \), and \( \delta \), such that

\[
P\left( \int_0^{t \wedge \tau} \exp(-|y_u|^p) \, du < \epsilon \right) \leq P(t \wedge \tau \leq c_{30} \epsilon).
\]

(3.33)

Applying Lemma 3.5 to the right hand side of (3.33), we deduce the existence of positive constants \( c_{31} \) and \( c_{32} \), such that if \( t \in (0, c_{31}) \) and \( 0 < \epsilon < \frac{1}{c_{30}} \), then

\[
P\left( \int_0^{t \wedge \tau} \exp(-|y_u|^p) \, du < \epsilon \right) \leq \exp\left( -\frac{c_{32}}{\epsilon} \right).
\]

The constants \( c_{30}, c_{31}, c_{32} \) depend only on \( m \), \( p \), \( c_3 \), and the characteristics of \( \tau \). Thus the conclusion of the lemma holds in this case.

On the other hand, suppose \( |y_0| < c_{29} \). We now set

\[
\tau_6 \equiv \inf\{ s > 0 : |y_s|^m = c_{28} \} \wedge \tau.
\]

26
Jensen’s inequality yields

\[
P\left( \int_0^{t \wedge \tau} \psi(|y_u|^m) du < \epsilon \right) \leq P\left( \int_0^{t \wedge \tau_6} |y_u|^m du \leq (t \wedge \tau_6)\psi^{-1}\left( \frac{\epsilon}{t \wedge \tau_6} \right) \right) \leq P_1 + P_2
\]

where

\[
P_1 \equiv P\left( t \wedge \tau_6 \leq \frac{\epsilon}{e_{28}} \right)
\]

and

\[
P_2\left( \int_0^{t \wedge \tau} |y_u|^m du \leq (t \wedge \tau_6)\psi^{-1}\left( \frac{\epsilon}{t \wedge \tau_6} \right), \ t \wedge \tau_6 > \frac{\epsilon}{e_{28}} \right).
\]

Note that \(P_1\) is of the same form as the probability on the right hand side of (4.7), and hence satisfies a similar estimate.

We now consider \(P_2\). An elementary argument shows that the convexity of \(\psi\) in the interval \((0, e_{28})\) implies that the function \(\theta(u) \equiv u\psi^{-1}\left( \frac{\epsilon}{u} \right)\) is increasing for \(u > \frac{\epsilon}{e_{28}}\).

In particular, if \(t \wedge \tau_6 > \frac{\epsilon}{e_{28}}\) then

\[
(t \wedge \tau_6)\psi^{-1}\left( \frac{\epsilon}{t \wedge \tau_6} \right) \leq T\psi^{-1}\left( \frac{\epsilon}{T} \right)
\]

where \(T\) is any upper bound for \(t\). This implies

\[
P_2 \leq P\left( \int_0^{t \wedge \tau_6} |y_u|^m du \leq T\psi^{-1}\left( \frac{\epsilon}{T} \right) \right).
\]  

(3.34)

We now apply Lemma 3.5 to estimate the right hand side of (3.34). Thus

\[
P_2 < \exp\left( -c_5 \left\{ T\psi^{-1}\left( \frac{\epsilon}{T} \right) \right\}^{-\frac{1}{m+1}} \right) \leq \exp(-c_7 \log \epsilon^{|q|}),
\]  

(3.35)

for all \(0 < t < c_{34}\) and \(\epsilon < \exp\left( -c_{33}t^{-\frac{1}{q}} \right)\), where \(c_7, c_{33}, \) and \(c_{34}\) are positive constants exhibiting the appropriate dependence. Then (3.35) gives an estimate for \(P_2\) of the required form, and the proof of the lemma is complete.

4. A study of a class of degenerate functional stochastic differential equations

In the previous section we derived very sharp sufficient conditions for the existence of smooth densities for the class of Itô processes

\[
d\xi_t = \sum_{i=1}^{n} X_i(\xi_t)dw_i + X_0(\xi_t)dt.
\]  

(4.1)
The arguments relied heavily on the existence of an invertible stochastic flow on $\mathbb{R}^d$ determined by the equation. Furthermore, as is apparent from Theorem 2.7 and its proof, the Lie bracket condition of Hörmander appears as a result of the interaction between $Z$, the derivative of the stochastic flow, and the vector fields $X_i$. The existence of a stochastic flow is the analytic counterpart of the probabilistic fact that the solution to the classical Itô equation (4.1) is a Markov process. In this section (which comes from [BM2]) we study a class of stochastic differential equations with far less analytic and probabilistic structure.

In order to describe this class of equations, we introduce the following notation and assumptions. Denote by:

- $\eta$ a continuous (deterministic) $\mathbb{R}^d$-valued path defined on a time interval $[-r, 0]$, for a fixed $r > 0$.
- $g$, a bounded smooth map from $\mathbb{R}^d$ to $\mathbb{R}^d \otimes \mathbb{R}^n$ with bounded derivatives of all orders.
- $C$ the space of continuous paths $\{x : [-r, \infty) \mapsto \mathbb{R}^d\}$, equipped with the uniform norm on compact time sets.
- $B : C \mapsto C$ a smooth bounded map, with bounded derivatives of all orders. Assume also that $B$ is non-anticipating, i.e. for all $t \geq -r$ and $x \in C$, $B(t, x) \equiv B(x)_t$ depends only on $\{x(u) : u \leq t\}$.
- $H_d[0,s]$, the $d$-dimensional Cameron-Martin space (cf. Section 2) with paths and norm defined on the time interval $[0, s]$. Suppose that for all $x \in C$, the map $DB(x)/H_d : H_d \mapsto H_d$ and that

$$\alpha \equiv \sup_{s \in [0,t], x \in C} ||DB(x)||_{H_d[0,s]} < \infty.$$  

We consider the following stochastic functional differential equation

$$dx_t = g(x_{t-r})dw_t + B(t, x)dt, \quad t \geq 0$$

where, as before, $w$ is a standard Wiener process in $\mathbb{R}^n$. Owing to the past-state dependence of the coefficients in the equation, the solution $x$ will not generally be a Markov process.

In addition to their theoretical interest, equations of the form (4.2) have important applications. For example, physical systems influenced by noise are often modeled as diffusion processes (4.1). However, this choice of model assumes that the evolution of the system at any instant in time depends only on the position of the system at that instant, together with the noise input. Since physical systems are always subject to some degree of inertia, this assumption is somewhat unrealistic. It would seem
that equations of the form (4.2), where the coefficients are allowed to depend on past behaviour, more accurately model such physical systems.

We establish the existence of smooth densities for \( x_t \) under hypotheses that allow degeneracy of the matrix-valued function \( g \). The general scheme of Chapter 2 is used. As before, we construct a Malliavin covariance \( \sigma \) for each \( x_t, (t > 0) \) and deduce the existence of a smooth density for \( x_t \) by demonstrating the strong invertibility of \( \sigma \). However the argument for doing this is, by necessity, quite different from that in Chapter 3. Because of its non-Markovian nature*, \( x \) does not generate a stochastic flow on \( \mathbb{R}^d \). Thus the analysis used in Chapter 3 to study the non-degeneracy of \( \sigma \) breaks down. We salvage the situation by proving and exploiting a property that seems peculiar to equations of the form (4.2); namely we show (cf. Lemma 4.5) that for all times \( a \geq 0 \) and \( k \geq 1 \), \( x \) satisfies estimates of the form

\[
P\left( \int_a^{a+r} |x_t|^2 dt < \varepsilon \right) = o(\varepsilon^k) \text{ as } \varepsilon \to 0.
\]

This probabilistic lower bound on \( x \) is propagated from the initial condition \( \eta \) by the time delay \( r \) in equation 4.2 **. We use this property to deduce non-degeneracy of the Malliavin covariance matrices, under appropriate hypotheses on \( g \) and \( B \).

The main result of this section is

**Theorem 4.1.** Suppose there exist positive constants \( \rho, \delta \), an integer \( p \geq 2 \) and a function \( \phi : \mathbb{R}^d \to \mathbb{R} \) such that

(i) the Lebesgue measure of the set \( \{ s \in [-r, 0] : \phi(\eta(s)) = 0 \} \) is zero.

(ii) \( g \) satisfies

\[
g g^*(x) \geq \begin{cases} |\phi(x)|^p I, & |\phi(x)| < \rho \\ \delta I, & |\phi(x)| \geq \rho. \end{cases}
\]  

(iii) \( \phi \) is \( C^2 \) with bounded first and second derivatives and there is a positive constant \( c \) such that

\[
||\nabla \phi(x)|| \geq c, \text{ whenever } |\phi(x)| \leq \rho.
\]  

Suppose \( 0 < \alpha \sqrt{2} t e^{-\alpha t} < 1 \). Then the random variable \( x_t \) defined by (4.2) is absolutely continuous and has a \( C^\infty \) density.

* In fact, the process \( x \) can be embedded as a Markov process with an infinite-dimensional state space. However, the generator of this process is a highly degenerate operator whose analysis is way beyond the reach of existing PDE techniques.

** There is no reason to believe that similar estimates hold for classical diffusion processes.
We note that condition (ii) above allows degeneracy of \( g \) on the hypersurface \( D \subseteq \mathbb{R}^d \) where \( \phi \) vanishes, while condition (ii)) implies that \( D \) does not contain any singular points.

Denote by \( F \) the map \( w \mapsto x_t \). In order to prove the theorem we must compute the Malliavin covariance matrix, defined formally by \( \sigma \equiv DF(w)DF(w)^* \). This can be done as follows. Let \( H \) denote the Cameron-Martin space and \( P_m \) the sequence of piecewise linear projections defined in Chapter 2, we consider the natural restriction \( \tilde{F} \) of the map \( F \) to \( H \), i.e. the map \( h \in H \mapsto k_t \) where \( k \equiv \eta \) on \([-r,0]\) and

\[
k'_t = g(k_{s-r})h'_s + B(s,k), \ s > 0.
\]

The matrix \( \sigma \) is then obtained as the limit in probability of the matrix sequence \( D\tilde{F}(P_m)D\tilde{F}(P_m)^* \).

The result of this computation is

**Lemma 4.2.** The Malliavin covariance matrix \( \sigma \) for the random variable \( x_t \) is

\[
\sigma = \int_0^t Z_u g(x_{u-r})g(x_{u-r})^* Z_u^* du \tag{4.5}
\]

where the process \( Z \) satisfies

\[
Z_s = I + \int_{(s+r)^\wedge t}^t Dg(x_{s-r})(.,dw_s)^* Z_s + D_{H_d}B(x)^* \left( \int_0^s Z \right)_s. \tag{4.6}
\]

It is at this point that the non-Markovian nature of the problem manifests itself analytically; in contrast to the situation in Chapter 3, there is now no reason to suppose that the matrix process \( Z_s \) is non-degenerate for small values of \( s \). We show in the next Lemma that it is, however, non-degenerate for values of \( s \) close enough to \( t \).

**Lemma 4.3.** Under the hypotheses of Theorem 4.1, \( Z_s \in GL(d) \) for \( s \in [t-r,t] \) and there exists a deterministic constant \( c < \infty \) such that \( ||Z_s^{-1}|| \leq c \) for \( s \in [t-r,t] \).

**Proof.** For notational convenience, write the operator \( D_{H_d}B(x)^* \) as \( K(x) \). For \( s \in [t-r,t] \) and any unit vector \( e \in \mathbb{R}^d \), (4.6) gives

\[
Z_se = e + K(x)_s(\int_0^s Z_u e \ du). \tag{4.7}
\]

Now

\[
|K(x)_s(\int_0^s Z_u e \ du)| \leq ||K(x)(\int_0^s Z_u e \ du)||_{H_d[0,s]} \\
\leq \alpha ||\int_0^s Z_u e \ du||_{H_d[0,s]} = \alpha (\int_0^s |Z_u e|^2)^{1/2}. \tag{4.8}
\]
Substituting this back into (4.7), we deduce
\[ ||Z_s|| \leq 1 + \alpha \left( \int_0^s ||Z_u||^2 du \right)^{1/2} \]
which implies
\[ ||Z_s||^2 \leq 2 + 2\alpha \int_0^s ||Z_u||^2 du. \]
Applying Gronwall’s inequality to this yields
\[ ||Z_s|| \leq \sqrt{2} e^{\alpha s} \leq \sqrt{2} e^{\alpha t}, \quad \forall s \leq t. \]
Combining this with (4.8) gives
\[ ||K(x_s)\left( \int_0^s Z_u \, du \right)|| \leq \sqrt{2} e^{\alpha t} \alpha < 1. \]
It follows from (4.7) that \( Z_s \in GL(d) \),
\[ Z_s^{-1} = \sum_{j=0}^{\infty} (-1)^j \left[ K(x_s)\left( \int_0^s Z_u \, du \right) \right]^j \]
and
\[ ||Z_s^{-1}|| \leq (1 - \alpha \sqrt{2} e^{\alpha t})^{-1}. \]

Let \( \lambda \) denote the smallest eigenvalue of \( \sigma \) and set \( \xi_s \equiv \phi(x_s), \ s \geq -r \). It follows from (4.5) and Lemma 4.3 that
\[ \lambda \geq \frac{1}{\varepsilon} \int_{(t-r)\wedge 0}^t (||\xi_{u-r}||^p \wedge \delta) \, du. \quad (4.9) \]
The following two results are used to produce lower bounds on the Malliavin covariance matrices corresponding to equation 4.2. These lower bounds comprise the essential part of the proof of Theorem 4.1.

**Lemma 4.4.** Fix \( b > a > 0 \). Let \( y : [0, \infty) \times \Omega \rightarrow \mathbb{R} \) be a measurable stochastic process such that \( E \sup_{a \leq t \leq b} |y(t)|^p \) is finite for every positive integer \( p \). Suppose that \( y \) satisfies
\[ P \left( \int_a^b |y(t)|^2 \, dt < \varepsilon \right) = o(\varepsilon^k). \]
Then for every positive constant \( \alpha \),
\[ P \left( \int_a^b \{ y(t)^2 \wedge \alpha \} \, dt < \varepsilon \right) = o(\varepsilon^k). \]
Proof This result was originally proved as Lemma 3 in ([BM1], pp. 91–94). We give here a simplified proof of the result.

Let $A$ and $B$ denote, respectively, the sets \{ $s \in [a, b] : y(s)^2 \leq \alpha$ \} and \{ $s \in [a, b] : y(s)^2 > \alpha$ \}. Then

$$
P \left( \int_a^b \{ y(t)^2 \wedge \alpha \} \, dt < \varepsilon \right) = P \left( \int_A y(t)^2 \, dt + \alpha \lambda(B) < \varepsilon \right)
$$

$$
\leq P \left( \int_A y(t)^2 \, dt < \varepsilon, \alpha \lambda(B) < \varepsilon \right) = P \left( \int_A y(t)^2 \, dt < \varepsilon + \int_B y(t)^2 \, dt, \lambda(B) < \varepsilon/\alpha \right) \leq P_1 + P_2.
$$

Here

$$
P_1 := P \left( \int_a^b y(t)^2 \, dt < \varepsilon + \int_B y(t)^2 \, dt, \lambda(B) < \varepsilon/\alpha, \sup_{a \leq t \leq b} y(t)^2 \leq 1/\sqrt{\varepsilon} \right)
$$

and

$$
P_2 := P( \sup_{a \leq t \leq b} y(t)^2 > 1/\sqrt{\varepsilon}).
$$

Note that

$$
P_1 \leq P \left( \int_a^b y(t)^2 \, dt < \varepsilon + \sqrt{\varepsilon}/\alpha \right) = o(\varepsilon^k)
$$

by hypothesis.

Using the finite-moment hypothesis on $y$ and applying Markov Chebyshev's inequality to the probability $P_2$, we obtain $P_2 = o(\varepsilon^k)$. This completes the proof of the lemma.

**Lemma 4.5 (Propagation lemma).** Suppose that, for some $-r < a < b$, $\xi$ satisfies

$$
P \left( \int_a^b \xi_s^2 \, ds < \varepsilon \right) = o(\varepsilon^k). \quad (4.10)
$$

Then

$$
P \left( \int_{a+r}^{b+r} \xi_s^2 \, ds < \varepsilon \right) = o(\varepsilon^k). \quad (4.11)
$$

Proof. Write $g = (g_1 \ldots g_n)$, where $g_i, 1 \leq i \leq n$, are column $d$-vectors. Computing $\xi_s = \phi(x_s), s > 0$, by Itô's formula gives

$$
d\xi_s = \sum_{i=1}^n \nabla \phi(x_s) \cdot g_i(x_{s-r}) \, dw_i(s) + G(s) \, ds, \quad s > 0 \quad (4.12)
$$
where $G$ is a bounded adapted real-valued process. We write

$$P\left(\int_{a+r}^{b+r} \xi_s^2 \, ds < \varepsilon\right) = P_1 + P_2$$

where

$$P_1 \equiv P\left(\int_{a+r}^{b+r} \xi_s^2 \, ds < \varepsilon, \sum_{i=1}^{n} \int_{a+r}^{b+r} [\nabla \phi(x_s) \cdot g_i(x_{s-r})]^2 \, ds \geq \varepsilon^{1/18}\right)$$

and

$$P_2 \equiv P\left(\int_{a+r}^{b+r} \xi_s^2 \, ds < \varepsilon, \sum_{i=1}^{n} \int_{a+r}^{b+r} [\nabla \phi(x_s) \cdot g_i(x_{s-r})]^2 \, ds < \varepsilon^{1/18}\right).$$

In view of 4.12, an inequality of Kusuoka and Stroock (cf. [KS2], Lemma 6.5) implies that $P_1 = o(\varepsilon^k)$ Thus it is sufficient to show that $P_2$ also has this property.

Write

$$a(s) = \sum_{i=1}^{n} [\nabla \phi(x_s) \cdot g_i(x_{s-r})]^2.$$

Then by (4.3) and (4.4), it follows that

$$a(s) \geq c_2 |\xi_s - r|^{p \wedge \delta} \text{ if } |\xi_s| \leq \rho.$$

Define

$$A \equiv \{s \in [a+r, b+r] : |\xi_s| \leq \rho\} \text{ and } B \equiv \{s \in [a+r, b+r] : |\xi_s| > \rho\}.$$

Then

$$P_2 = P\left(\int_{a+r}^{b+r} \xi_s^2 \, ds < \varepsilon, \int_{a+r}^{b+r} a(s) \, ds < \varepsilon^{1/18}\right)$$

$$\leq P\left(\int_{a+r}^{b+r} \xi_s^2 \, ds < \varepsilon, \int_A c_2 (|\xi_s-r|^{p \wedge \delta}) \, ds < \varepsilon^{1/18}\right)$$

$$= P\left(\int_{a+r}^{b+r} \xi_s^2 \, ds < \varepsilon, \int_{a+r}^{b+r} c_2 (|\xi_s-r|^{p \wedge \delta}) \, ds < \varepsilon^{1/18} + \int_B c_2 (|\xi_s-r|^{p \wedge \delta}) \, ds\right)$$

$$\leq P\left(\int_{a+r}^{b+r} \xi_s^2 \, ds < \varepsilon, \int_{a+r}^{b+r} c_2 (|\xi_s-r|^{p \wedge \delta}) \, ds < \varepsilon^{1/18} + c_2 \delta \lambda(B)\right).$$

However $\int_{a+r}^{b+r} \xi_s^2 \, ds < \varepsilon$ implies $\lambda(B) < \varepsilon / \rho^2$. Thus the preceding probability is

$$\leq P\left(\int_{a+r}^{b+r} c_2 (|\xi_s-r|^{p \wedge \delta}) \, ds < \varepsilon^{1/18} + c_2 \delta \varepsilon / \rho^2\right)$$

$$\leq P\left(\int_a^b (|\xi_s|^{p \wedge \delta}) \, ds < c \varepsilon^{1/18}\right)$$
for some positive constant \( c' \), and for small enough \( \varepsilon \). Assumption 4.10, Lemma 4.4, and Jensen’s inequality allow us to conclude that the probability on the right hand side of the above inequality is \( o(\varepsilon^k) \). This implies that \( P_2 = o(\varepsilon^k) \), and the proof of Lemma (4.2) is complete.

We are now in a position to complete the proof of Theorem 4.1. Recall that we need to verify the condition

\[
(\det \sigma)^{-1} \in \cap_{p \geq 1} L^p(\gamma).
\]  

(4.13)

As before let \( \lambda \) denote the smallest eigenvalue of \( \sigma \). Let \( n \) denote the integer such that \( t \in ((n-1)r, nr] \). First, suppose \( n = 1 \). Hypothesis (i) of Theorem 4.1 and (4.9) imply

\[
\lambda \geq \frac{1}{c} \int_{-r}^{t-r} (|\xi|^p \wedge \delta) \, du = \frac{1}{c} \int_{-r}^{t-r} (|\eta|^p \wedge \delta) \, du > 0.
\]

Since the second integral is deterministic, 4.13 trivially holds in this case.

On the other hand, suppose \( n > 1 \). Since \( [t - (n+1)r, t - nr] \subset [-r, 0] \), hypotheses (i) implies

\[
P\left( \int_{t-(n+1)r}^{t-r} |\xi|^2 \, ds < \varepsilon \right) = o(\varepsilon^k).
\]

We now iterate Lemma 4.5 \( n \) times on this estimate to obtain

\[
P\left( \int_{t-r}^{t} |\xi|^2 \, ds < \varepsilon \right) = o(\varepsilon^k).
\]

Applying Lemma 3.4 yields

\[
P\left( \int_{t-r}^{t} |\xi|^2 \wedge \delta^{2/p} \, ds < \varepsilon \right) = o(\varepsilon^k)
\]

(4.14)

Using Jensen’s inequality in (4.14) gives

\[
P\left( \int_{t-r}^{t} |\xi|^p \wedge \delta \, ds < \varepsilon \right) = o(\varepsilon^k).
\]

Combining this with (4.9), we finally have

\[
P(\lambda < \epsilon) = o(\varepsilon^k).
\]

This implies (4.13) and completes the proof of the theorem.
5. Some open problems

In this section we describe a few open problems that we hope will serve as a stimulus to further research.

1. It would be interesting to generalize Theorem 4.1 to the class of fully hereditary functional differential equations

\[ dx_t = A(t, x_t)dw + B(t, x_t)dt \]  

(5.1)

where both \( A(t, x) \) and \( B(t, x) \) are allowed to depend on the whole history of the path \( \{x_s : 0 \leq s \leq t\} \). Kusuoka and Stroock [KS1] addressed this problem under a strong ellipticity assumption, i.e. they have shown that \( x_t \) has a smooth density for all positive \( t \) if there exists \( \delta > 0 \) such that \( A(t, x)A(t, x)^* \geq \delta I, \forall (t, x) \in [0, \infty) \times C \). As far as I am aware, Theorem 4.1 is the only result establishing the existence of densities for a general class of non-Markov Itô processes under hypotheses that allow degeneracy of the diffusion coefficient. An analogous result in the fully general setting (5.1) would therefore be of considerable interest and importance.

2. The non-hypoellipticity of the operators

\[ \frac{\partial^2}{\partial x_1^2} + \exp\{-|x_1|^p\} \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2}, \quad p \leq -1 \]

described in Section 2, contrasts strikingly with a result proved by Fedii in 1971. Fedii showed that the operator on \( \mathbb{R}^2 \)

\[ \frac{\partial^2}{\partial x^2} + \exp(-|x|^p) \frac{\partial^2}{\partial y^2}, \quad p < 0 \]

is hypoelliptic for all negative values of \( p \). It would be interesting to gain a deeper understanding, either by classical or probabilistic means, of the role that dimension is playing in these results.

3. The probabilistic methods employed above can also be used to study quasilinear partial differential equations. For example, let \( \psi \) denote a smooth non-linear and non-negative function defined on \( [0, \infty) \times \mathbb{R}^d \times \mathbb{R} \) and consider the initial-value problem

\[ \frac{\partial u}{\partial t} = Lu + \psi(t, x, u), \quad (t, x) \in (0, \infty) \times \mathbb{R}^d \]

\[ u(0, x) = 0, \quad x \in \mathbb{R}^d \]  

(5.2)

where \( L \) is defined in (1.1). In particular, A continuous weak solution to (5.2) is given by the probabilistic representation

\[ u(t, x) = E\left[ \int_0^t \psi(s, \xi_{s-t} u(s, \xi_{s-t}^x)) \, ds \right] \]  

(5.3)
where $\xi^x$ denotes process $\xi$ in (1.2) with initial point $x \in \mathbb{R}^d$. It should be possible to use this probabilistic representation together with the ideas in of Section 3 to show that, under suitable conditions, the solution $u$ to (5.2) is smooth for $t > 0$. We note, however, that the problems is considerably more difficult than for the linear case treated in Section 3, owing to the presence of $u$ in the right hand side of (5.3).

Quasilinear problems of this type have received a great deal of attention in recent years. Dynkin [D] and others have discovered a remarkable link between operators of the form $L + \psi$ and a collection of stochastic processes called superprocesses.

4. A related issue is the study of the quasilinear Dirichlet problem. Suppose $D$ is a bounded regular open subset of $\mathbb{R}^d$ with a $C^2$ boundary, $f$ is a smooth non-negative function defined on $\bar{D} \times \mathbb{R}$, and $g$ is a smooth non-negative function on $\partial D$

$$\begin{align*}
Lu &= f(x, u), \quad x \in D \\
u(x) &= g(x), \quad x \in \partial D
\end{align*}$$

(5.4)

Under mild further conditions, a weak continuous solution of equation (5.4) exists, given implicitly by

$$u(x) + E \left[ \int_0^\tau f(\xi^x(s), u(\xi^x(s))) \, ds \right] = E[g(\xi^x(\tau))]$$

(5.5)

where $\tau = \tau(x)$ is the first exit time of the diffusion $\xi^x$ from $D$. Again, one would hope to be able to study the regularity of the solution $u$ to (5.4) via the representation (5.5) by using the methods of Section 3. The goal would be to establish smoothness of $u$ on $\bar{D}$ under conditions that allow degeneracy of the operator $L$ (e.g. Hörmander’s condition or even the superdegeneracy condition introduced in Theorem 3.1.).

References


Kusuoka, S; and Stroock, D; Applications of the Malliavin calculus, I, Taniguchi SymposiA Katata (1982), 271–306.

Kusuoka, S; and Stroock, D; Applications of the Malliavin calculus, Part II, Journal of Faculty of Science, University of Tokyo, SecI A, Vol32, No1 (1985) 1-76.


Malliavin, P; C^k-hypoellipticity with degeneracy, part II, Stochastic Analysis, A. Friedman and M. Pinsky (eds), 1978, 327-340.

