

Superdegenerate Hypoelliptic Differential Operators

Denis Bell

Department of Mathematics, University of North Florida
4567 St. Johns Bluff Road South, Jacksonville, FL 32224, U. S. A.
email: dbell@unf.edu

1 Salah Mohammed: a personal reminiscence

First up, I would like to thank the organizers, Greg Budzban, Randy Hughes, and Henri Schurz for arranging this significant event, and for inviting me to speak at it.

I first met Salah while a graduate student at the University of Warwick in the late seventies. He had graduated a few years earlier, with a Ph.D. under the direction of Jim Eells and was living in Khartoum, but used to come back there periodically to visit. (Strangely, Salah and I have been based in several of the same places over the years, Warwick, Hull, Carbondale, but never at the same time!) We soon became friends. A memorable highlight of our friendship was attending an open air Bob Dylan concert in Aldershot, England, an airstrip outside of London, in 1978.

After Warwick, Salah and I did not meet again for several years until I came to SIU to give a talk in the late eighties. He was a permanent member of the department here by that time. Over dinner, I happened to mention that perhaps one could obtain a new proof of Itô's classical theorem on the existence of solution to a stochastic differential equations (sde)

$$dx = A(x(t))dw + B(x(t))dt \tag{1.1}$$

by considering the *stochastic delay equation*

$$dx^r = A(x^r(t-r))dw + B(x^r(t-r)) \tag{1.2}$$

and deriving the solution x to (1.1) as $\lim_{r \rightarrow 0} x^r$. The point being that a solution to (1.2) obviously exists since the equation is explicitly solvable in terms of iterated stochastic integrals over time intervals of length r .

A few weeks later and much to my suprise, I received in the mail a manuscript with the theorem, coauthored by the two of us (subsequently published as [BM.1]). Everything worked out to the last detail! I suggested to Salah that since he had done all the work, he might prefer to publish the paper as a singly authored work, but he wouldn't hear of it.

So began a collaboration that lasted more than a decade until divergent research interests took us in different directions, by far the most fruitful collaboration of my professional life. I hope there will be more. After all, we are still both young (at heart, anyway!).

In this paper, I would like to discuss one of our joint works, which appeared in the Duke journal in 1995. In order to simplify the exposition and highlight the essential points, the proof will be presented in outline form with many of the technical details omitted. (The full exposition can be found in the original paper [BM.2].)

2 Hörmander's Theorem and the Malliavin calculus

Definition 2.1 *A differential operator G is said to be hypoelliptic if, whenever $G u$ is smooth, for some distribution u defined on an open subset of the domain of G , then u is a smooth (i.e. C^∞) function.*

The most celebrated result in this area is Hörmander's theorem [H], published in 1967.

Let X_0, \dots, X_n denote bounded smooth vector fields on \mathbf{R}^d with bounded derivatives of all orders regarded, in the usual way, as first order differential operators on \mathbf{R}^d . Let L denote the second-order operator

$$L = \sum_{i=1}^n X_i^2 + X_0.$$

Hörmander's theorem is as follows.

Theorem 2.2 *Suppose the Lie algebra generated by X_0, \dots, X_n has full rank in an open set $U \subset \mathbf{R}^d$, i.e. the vectors*

$$\{X_i, [X_i, X_j], [[X_i, X_j], X_k], \dots, 0 \leq i, j, k, \dots \leq n\}$$

span \mathbf{R}^d at all points in U .

Then L is hypoelliptic on U .

The hypothesis in Theorem in 2.2 is known as *Hörmander's condition* (HC).

A probabilistic formulation of hypoellipticity

Let (w_1, \dots, w_n) denote a standard Wiener process in \mathbf{R}^n . Consider the Stratonovich sde

$$\xi_t^x = x + \sum_{i=1}^n \int_0^t X_i(\xi_s^x) \circ dw_i(s) + \int_0^t X_0(\xi_s^x) ds. \quad (2.1)$$

It has been known since the pioneering work of Itô, that the solution ξ to (2.1) is a time-homogeneous Markov process, whose transition probabilities $p(t, x, dy) \equiv p(\xi_t^x \in dy)$ determine a weak solution to *Kolmogorov's equation*

$$\frac{\partial p}{\partial t} = Lp. \quad (2.2)$$

We note that here, and in the sequel, L denotes the the infinitesimal generator of the process ξ , i.e. the differential operator¹

$$\frac{1}{2} \sum_{i=1}^n X_i^2 + X_0.$$

Suppose the vector fields satisfy the following slightly stronger parabolic version of HC²

$$\{X_i, [X_j, X_k], [[X_j, X_k], X_l], \dots, 1 \leq i \leq n, 0 \leq j, k, l \dots \leq n\}$$

span \mathbf{R}^d .

Then it follows from Hörmander's theorem that the operator $\frac{\partial}{\partial t} - L$ is hypoelliptic. In particular, the transition probabilities $p(t, x, dy), t > 0$ for the sde

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

admit densities $p(t, x, y)$ that are smooth in y (and in fact in all three arguments).

In the opposite direction, if it can be established *directly* that under parabolic HC, the process ξ_t defined by (2.1) admits smooth densities, then one can deduce the hypoellipticity of L and thus obtain a probabilistic proof of Hörmander's theorem. This exciting line of research was initiated by Paul Malliavin in 1976 in his seminal paper [M] and is the point of departure of the Malliavin calculus.

Modern treatments of the Malliavin calculus are framed in terms of Sobolev spaces, closures of differential operators, etc., but the underlying idea is relatively straightforward and can be understood without such technicalities.

Let H denote the *Cameron-Martin space*, i.e. the subspace of Wiener space consisting of absolutely continuous paths h with finite energy:

$$\int_0^T |h'_s|^2 ds < \infty.$$

The essential point is that, although the Itô map $g : w \mapsto \xi$ defined by the sde

$$d\xi_t = \sum_{i=1}^n X_i(\xi_t) \circ dw_i + X_0(\xi_t) dt \tag{2.3}$$

is highly irregular as a map on the space of continuous paths in the sense of the standard (Frechet) calculus, it is possible to make sense of directional derivatives of g in H -directions³.

The path $\eta \equiv D_h g(w)$ is found by formal differentiation in Eq. (2.3) wrt w . Thus η is the solution of the sde

$$d\eta_t = \sum_{i=1}^n \left\{ DX_i(\xi_t) \eta_t \circ dw_i + X_i(\xi_t) h'_i(t) dt \right\} + DX_0(\xi_t) \eta_t dt.$$

¹Acting on the y -variable in Eq. 2.1.

²Note that here, X_0 does not appear explicitly in the collection of vector fields but does occur in the Lie brackets.

³This notion of differentiability of Wiener functionals was developed by Leonard Gross prior to Malliavin's work, in the context of abstract Wiener spaces (cf. [G]).

Higher order derivatives of g can be obtained analogously.

The map $h \mapsto D_h g_t(w)$ defines a.s. an element of $L(H, \mathbf{R}^d)$. The Malliavin covariance matrix (an analogue of the Gram matrix in classical differential analysis) is defined by

$$\sigma_t \equiv Dg_t(w)^* Dg_t(w).$$

There is the following explicit form for σ_t ,

$$\sigma_t = Y_t \int_0^t Z_s A(\xi_s) A(\xi_s)^* Z_s^* ds Y_t^*$$

where $A = [X_1 \dots X_n]$, Y_t is the derivative of the stochastic flow $\xi_0 \mapsto \xi_t$, and $Z_t = Y_t^{-1}$.

Theorem 2.3 *If $\sigma_t \in GL(d)$ and*

$$\det \sigma_t^{-1} \in L^p, \forall p \geq 1$$

then ξ_t is absolutely continuous and has a C^∞ density.

This fundamental result of Malliavin was refined by Kusuoka & Stroock to the following criterion for hypoellipticity of the operator L .

Theorem 2.4 *Define $\Delta(t, y) \equiv \det \sigma_t$ where $y = \xi_0$. Suppose that for all $q \geq 1$ and $x \in \mathbf{R}^d$, there exists a neighborhood V of x such that*

$$\lim_{t \rightarrow 0^+} t \log \left(\sup_{y \in V} E |\Delta(t, y)^{-q}| \right) = 0.$$

Then L is hypoelliptic on \mathbf{R}^d .

3 Superdegenerate hypoelliptic operators

HC is known to be necessary for hypoellipticity of operators of the above form if the coefficients of L are analytic. This is not the case in the smooth non-analytic category, a situation that is strikingly illustrated by the following result of Kusuoka & Stroock [KS].

Theorem 3.1 . *Consider the class of differential operators on \mathbf{R}^3 of the form*

$$L_p \equiv \frac{\partial^2}{\partial x^2} + \exp(-|x|^p) \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}, \quad p < 0$$

Then L_p is hypoelliptic if and only if $p \in (-1, 0)$.

In particular, if $p \in (-1, 0)$ then L_p is hypoelliptic on \mathbf{R}^3 , yet fails to satisfy HC on the hyperplane $\{x = 0\}$.

Motivated by this result, Bell & Mohammed proved a sharp form of Hörmander's theorem for operators with smooth (but possibly non-analytic) coefficients that incorporates the Kusuoka-Stroock operators. This theorem allows degeneracies of exponential order on codimension-1 hypersurfaces in the ambient space. We term operators with this level of degeneracy *superdegenerate*.

The statement of the theorem requires some additional notation.

For $k \geq 0$, define $X^{(k)}$ to be a matrix with columns X_1, \dots, X_n , together with all vector fields obtained from X_0, \dots, X_n by forming iterated Lie brackets up to order k . Define

$$\lambda^{(k)} \equiv \text{smallest eigenvalue of } X^{(k)} X^{(k)*}.$$

Let H^c denote the set of points in \mathbf{R}^d where L fails to satisfy HC. Then

$$H^c = \{x \in \mathbf{R}^d : \lambda^{(k)}(x) = 0, \forall k\}.$$

The following definition is standard in PDE theory. A C^1 hypersurface $S \subset \mathbf{R}^d$ is said to be *non-characteristic* (with respect to L) at $x \in S$ if at least one of the vector fields X_1, \dots, X_n is *non-tangential* to S at x .

The condition has an intrinsic probabilistic meaning. If it fails at a point x in the domain of L then the diffusion

$$d\xi_t = \sum_{i=1}^n X_i(\xi_t) \circ dw_i$$

started from x will stay on S , hence ξ_t cannot have a density on \mathbf{R}^d . This in turn implies L is non-hypoelliptic at x (since we have seen that hypoellipticity implies the existence of densities).

Theorem 3.2 *Suppose the (non-Hörmander) set H^c of L is contained in a C^2 hypersurface S . Assume that for all $x \in H^c$*

(i) *S is non-characteristic at x .*

(ii) *There exists an integer $k \geq 0$, $p \in (-1, 0)$, and a neighborhood U of x , such that for all $y \in U$*

$$\lambda^{(k)}(y) \geq \exp(-[d(y, S)]^p).$$

where $d(y, S)$ denotes the Euclidean distance from y to S .

Then L is hypoelliptic.

In regard to the above hypotheses, it has been noted in the remark immediately preceding Theorem 3.2, that condition (i) is necessary for the hypoellipticity of L . The theorem allows HC to fail at any, or all, points on the hypersurface S . Condition (ii) controls the *rate* at which this happens in the neighborhood of S . The non-hypoelliptic Kusuoka-Stroock operators L_p show that an assumption of this type is also necessary. Furthermore the case L_{-1} shows that the allowed range of p in (ii) is optimal.

4 Proof of the theorem

We verify the Kusuoka-Stroock criterion

$$\lim_{t \rightarrow 0^+} t \log \left(\sup_{y \in V} E|\Delta(t, y)^{-q}| \right) = 0.$$

Establishing this condition under our hypotheses is somewhat delicate and requires an analysis of the interaction between the diffusion process ξ and the hypersurface S .

The strategy is as follows:

(i) We express S locally in the form

$$S = \{x \in \mathbf{R}^d / \phi(x) = 0\}$$

and translate the hypotheses of the theorem into conditions on ϕ .

(ii) Probabilistic lower bounds are obtained on the L^p -norms of the process $y_t \equiv \phi(\xi_t)$ for arbitrarily large values of p .

(iii) We study how these lower bounds are degraded under the exponential-type degeneracy allowed by the theorem. This leads to a lower bound on the integrand in the integral representation of σ_t that is then shown to imply the required property.

It is convenient to introduce the following terminology. We say a random time τ is *exponentially positive* if there exist positive constants a and b such that

$$P(\tau < \epsilon) \leq e^{-b/\epsilon}$$

for all $\epsilon < a$.

The prime example is the exit time τ of a diffusion process with bounded coefficients from a ball of fixed radius.

The proof of the theorem makes use of the following two key lemmas.

Lemma 4.1 (*Moment inequalities*) *Let y be an Itô process*

$$dy(t) = \sum_{i=1}^n a_i(t)dw_i + b(t)dt$$

where a_1, \dots, a_n and b are bounded adapted processes. Suppose at least one of $a_1(0), \dots, a_n(0)$ is deterministic and non-zero.

Let τ be an exponentially positive stopping time.

Then for every $m \geq 1$

$$P\left(\int_0^{t \wedge \tau} |y(u)|^m du = o(\epsilon^{-\frac{1}{m+1}})\right) \tag{4.1}$$

as $\epsilon \rightarrow 0^+$.

Lemma 4.2 *Let $p \in (-1, 0)$. The estimate (4.1) for $m > -\frac{p}{p+1}$ implies that there exists $q > 1$ such that*

$$P\left(\int_0^{t \wedge \tau} e^{-|y(u)|^p} du < \epsilon\right) = o(e^{-|\log \epsilon|^q})$$

as $\epsilon \rightarrow 0^+$.

The proof actually requires a somewhat more quantitative version of Lemma 4.2 but the above statement captures the spirit of the result.

Lemma 4.3 *Let $x \in H^c \subset S$. The hypotheses of the theorem imply the following*

(i) *There exists a neighborhood U of x and a C^2 map $\phi : U \mapsto \mathbf{R}$ such that*

$$S \cap U = \{y \in U / \phi(y) = 0\}.$$

(ii) *For at least one $i = 1, \dots, n$, $\nabla\phi(x) \cdot X_i(x) \neq 0$.*

(iii) *There exists $k \geq 0$ and $p \in (-1, 0)$ such that*

$$\lambda^{(k)}(y) \geq \exp(-|\phi(y)|^p), \quad \forall y \in U.$$

Recall we are trying to establish, for all $q > 1$

$$\lim_{t \rightarrow 0^+} t \log \left(\sup_{y \in V} E[|\Delta(t, y)^{-q}|] \right) = 0. \quad (4.2)$$

where $\Delta(t, y) = \det \sigma_t$.

It follows from the integral representation of σ_t that

$$E[|\Delta(t, x)^{-q}|] \leq c \sum_{j=1}^{\infty} P(Q(t, x) \leq j^{-1/(dq)})$$

where

$$Q(t, x) \equiv \inf \left\{ \sum_{i=1}^n \int_0^t \langle Z^x(u) X_i(\xi_u^x), h \rangle^2 du, |h| = 1 \right\}.$$

Since Z^x satisfies strong stochastic lower bounds, we can replace $Q(t, x)$ by

$$\int_0^{t \wedge \tau} \lambda(\xi_u^x) du$$

where τ is an exponentially positive stopping time and λ denotes $\lambda^{(0)}$, the smallest eigenvalue of the matrix AA^* where A is the matrix $[X_1 \dots X_n]$.

This yields

$$E[|\Delta(t, x)^{-q}|] \leq c \sum_{j=1}^{\infty} P\left(\int_0^{t \wedge \tau} \lambda(\xi_u^x) du \leq j^{-1/(dq)}\right). \quad (4.3)$$

In order to simplify the exposition at this point, we now assume the hypothesis of the theorem holds at x with $k = 0$. Then Lemma 4.3 (iii) holds with $k = 0$, i.e.

$$\lambda(y) \geq \exp(-|\phi(y)|^p), \quad \forall y \in U.$$

Substituting this into (4.3), we have

$$E[|\Delta(t, x)^{-q}|] \leq c \sum_{j=1}^{\infty} P\left(\int_0^{t \wedge \tau} \exp(-|\phi(\xi_u^x)|^p) du \leq j^{-1/(dq)}\right) \quad (4.4)$$

By Itô's formula the process $\phi(\xi_t^x)$ has the form

$$d\phi(\xi_t^x) = \sum_{i=1}^n \nabla\phi(\xi_t^x) \cdot X_i(\xi_t^x) dw_i(t) + G(t) dt$$

for some function G . Lemma 4.3 (ii) implies that the process $y_t \equiv |\phi(\xi_t^x)|$ satisfies the hypotheses of Lemma 4.2.

Applying Lemma 4.2 gives: there exists $\beta > 0$ and $r > 1$ such that

$$\begin{aligned} P\left(\int_0^{t \wedge \tau} \exp\{-|\phi(\xi_u^x)|^p\} du \leq j^{-1/(dq)}\right) \\ \leq \exp\{-\beta(\log j)^r\} \end{aligned} \quad (4.5)$$

for j satisfying

$$j^{-1/(dq)} \leq \exp(-t^{-1/r}),$$

i.e. for

$$j \geq \exp(qdt^{-1/r}).$$

Substituting (4.5) into (4.4), we deduce⁴

$$\begin{aligned} E[|\Delta(t, x)^{-q}|] = \\ o\left(\exp(qdt^{-1/r}) + \sum_{j=1}^{\infty} \exp\{-\beta(\log j)^r\}\right) \end{aligned} \quad (4.6)$$

as $t \rightarrow 0^+$.

Furthermore, this estimate can be seen to hold with x replaced by y , uniformly for y in a small neighborhood V of x . This implies the criterion in the Kusuoka-Stroock lemma:

$$\lim_{t \rightarrow 0^+} t \log \left\{ \sup_{y \in V} \|E[|\Delta(t, y)^{-q}|\] \right\} = 0.$$

and we are done!

In conclusion

Above all, mathematicians strive for understanding. The proof of a result might be more important than the result itself if it reveals something about the structure underlying the problem. *What is the origin of the critical exponent -1 in Theorem 3.2?* While this question has an essentially infinite-dimensional character, we show that the allowable range of p in the theorem derives directly from the form of the Gaussian function e^{-x^2} . Ultimately, this insight, relating the class of degenerate hypoelliptic Hörmander operators to the most fundamental of probabilistic objects, could prove to be a more important feature of this work than the technical contribution as a sharp form of Hörmander's theorem.

⁴Note that since $r > 1$, the infinite series in (4.6) is convergent.

References

- [BM.1] Bell, D. and Mohammed, S.-E. A. On the solution of stochastic ordinary differential equations via small delays. *Stochastics Stochastics Rep.* 28 (1989), no. 4, 295-299.
- [BM.2] Bell, D. and Mohammed, S.-E. A. An extension of Hörmander's theorem for infinitely degenerate second-order operators. *Duke Math. J.* 78 (1995), no. 3, 453-475.
- [G] Gross, L. Potential theory on Hilbert space. *J. Funct. Anal.* 1 1967 123-181.
- [H] Hörmander, L. Hypoelliptic second order differential equations. *Acta Math.* 119 1967 147-171.
- [KS] Kusuoka, S. and Stroock, D. Applications of the Malliavin calculus. II. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* 32 (1985), no. 1, 1-76.
- [M] Malliavin, P. Stochastic calculus of variation and hypoelliptic operators. *Proceedings of the International Symposium on Stochastic Differential Equations (Res. Inst. Math. Sci., Kyoto Univ., Kyoto, 1976)*, pp. 195-263, 1978.