An Introduction to Malliavin Calculus

Denis Bell

University of North Florida

Motivation -
the hypoellipticity problem

Definition. A differential operator G is *hypoelliptic* if, whenever the equation

$$Gu = f \tag{1}$$

holds in the *weak* sense, for some (Schwartz) distribution u and smooth function f, then this implies that u *is smooth*.

u *is a weak solution* to (1) means: for all test functions ϕ on $\mathbf{R^d}$

$$\int_{\mathbf{R}^d} (G^*\phi) u \, dx = \int_{\mathbf{R}^d} \phi f \, dx$$

where G^* is the formal adjoint of G in the space $L^2(dx)$.

Hypoellipticity an important property because, often weak solutions to PDEs can be shown to exist and one would like to know they are *classical* (smooth) solutions.

Let $X_0, \ldots, X_n$ be bounded smooth vector fields on $\mathbf{R}^d$ with bounded derivatives of all orders. Consider these as first-order differential operators and define the second-order differential operator

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

Hörmander proved the following fundamental result in 1967.

Hörmanders Theorem. *Suppose the vector fields*

$$\Big\{X_i, [X_i, X_j], [[X_i, X_j], X_k], \ldots; 0 \le i, j, k, \cdots \le n\Big\}$$

span $\mathbf{R}^d$ at all points in an open set $U \subset \mathbf{R}^d$. Then L is hypoelliptic in U.

The hypothesis above is known as *Hörmanders condition* (HC).

The class of operators addressed in Hörmanders theorem arise naturally in stochastic analysis:

Consider the stochastic differential equation (SDE)

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

Let $p(t,x,dy)$ denote the law of the random variable ξ_t^x, i.e.

$$P(\xi_t^x \in B) = \int_B p(t,x,dy)$$

These measures satisfy *Kolmogorov s forward and backward equations* in the weak sense

$$\frac{\partial p}{\partial t} = L_x p \tag{2}$$

$$\frac{\partial p}{\partial t} = L_y^* p. \tag{3}$$

Suppose the vector fields $X_0, \ldots X_n$ satisfy the following *parabolic* version of HC at each point in $\mathbf{R}^d$

$$\{X_i, [X_j, X_k], [[X_j, X_k], X_l], \ldots$$

$$i \leq 1 \leq n,\ 0 \leq j, k, l \cdots \leq n\}.$$

Then Hörmanders theorem yields that the operators $\frac{\partial}{\partial t} - L$ and $\frac{\partial}{\partial t} - L^*$ are hypoelliptic.

Writing equations (2) and (3) in the form

$$\Big(\frac{\partial}{\partial t} - L_x\Big)p = 0$$

$$\Big(\frac{\partial}{\partial t} - L_y^*\Big)p = 0$$

we infer that the measures $p(t, x, dy)$ *admit densities that are smooth in* t, x *and* y.

This flow of information between PDE theory and stochastic differential equations is *two-way*:

The hypoellipticity of the operator L can be *deduced* from the smoothness of the transition probabilities $p(t, x, dy)$ to the SDE

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

So if it can be proved *directly* that under HC, the transition probabilities are smooth, then this will give a *probabilistic* proof of Hörmanders theorem.

This is the problem that motivated the development of the Malliavin calculus.

A finite-dimensional version of the problem

Consider the Euclidean space $\mathbf{R}^n$ equipped the Gaussian measure

$$d\gamma(x) = (2\pi)^{\frac{-n}{2}} e^{-\frac{||x||^2}{2}} dx.$$

Let $g : \mathbf{R}^n \mapsto \mathbf{R}^d$, where $n > d$.

We want to study the regularity of the induced measure ν on $\mathbf{R}^d$ defined by

$$\nu(B) = \gamma\Big(g^{-1}(B)\Big).$$

There is the following well-known result:

Lemma. *Suppose that for all $\alpha = (\alpha_1, \dots \alpha_k)$ there exists a constant C_α such that*

$$\left| \int_{\mathbf{R}^d} D^{\alpha_1}_{x_1} \dots D^{\alpha_k}_{x_k} \phi \, d\nu \right| \leq C_\alpha ||\phi||_\infty .$$

for all test functions on $\mathbf{R}^d$. Then ν is absolutely continuous and has a smooth density.

Theorem. *Suppose $g : \mathbf{R}^n \mapsto \mathbf{R}^d$ is C^∞ and $Dg(x) \in L(\mathbf{R}^n, \mathbf{R}^d)$ is a.e. surjective (g is a submersion). Then the measure $\nu \equiv \gamma \circ g^{-1}$ is absolutely continuous wrt Lebesgue measure on $\mathbf{R}^d$.*

Proof. Consider the $d \times d$ (*Gram*) matrix

$$\sigma = Dg(x)Dg(x)^*.$$

Then $\sigma(x) \in GL(d)$. Let e_i be the i-th standard basis vector in $\mathbf{R}^d$ and define

$$r = Dg(x)^*\sigma^{-1}e_i.$$

Then

$$Dg(x)r = e_i. \tag{4}$$

We have

$$\int_{\mathbf{R}^d} \frac{\partial \phi}{\partial x_i} d\nu = \int_{\mathbf{R^n}} \frac{\partial \phi}{\partial x_i} \circ g(x) d\gamma$$

$$= \int_{\mathbf{R^n}} D(\phi \circ g)(x) r d\gamma. \tag{5}$$

We can now integrate by parts to remove the derivative in (5).

This gives

$$\int_{\mathbf{R^n}} D(\phi \circ g)(x) r d\gamma = \int_{\mathbf{R^n}} (\phi \circ g) X d\gamma$$

where

$$X(x) = < r, x > - \sum_{k=1}^{n} \frac{\partial r_k}{\partial x_k}.$$

We have shown

$$\int_{\mathbf{R}^d} \frac{\partial \phi}{\partial x_i} d\nu = \int_{\mathbf{R^n}} (\phi \circ g) X d\gamma.$$

This yields

$$\left| \int_{\mathbf{R}^d} \frac{\partial \phi}{\partial x_i} d\nu \right| \leq ||\phi||_\infty ||X||_{L^1(\gamma)}$$

and absolute continuity of ν follows as a special case of the Lemma.

By iterating this argument we can derive the higher-order estimates in the Lemma and hence prove that the density of ν is *smooth*.

Note the key components in this derivation:

(i) surjectivity of $Dg(x)$. This was used to define a *lift* r of e_i to $(\mathbf{R}^n, \gamma)$ (cf. eq. (4)).

(It follows from Sard s theorem that the surjectivity condition is *necessary* for the absolute continuity of ν.)

(ii) Integration by parts in $(\mathbf{R}^n, \gamma)$.

Regularity of the law of an SDE

We want to study the law ν of the random variable ξ_t where

$$\xi_s = x + \sum_{i=1}^{n} \int_0^s X_i(\xi_u) dw_i + \int_0^s X_0(\xi_u) du.$$

Let γ denote the Wiener measure on the space of paths $C_0 = \{\sigma : [0,1] \mapsto \mathbf{R}^n : \sigma(0) = 0\}$.

Note that ν is the induced measure $\gamma \circ g_t^{-1}$ where

$$g_t : w \mapsto \xi_t.$$

Wiener measure has a well-understood analytic structure. In particular, there is a theorem that permits *integration by parts* over the Wiener space. The setup here looks similar to the finite-dimensional case. However, *there is an essential problem*.

The map $w \mapsto \xi$ defined by the SDE

$$\xi_s = x + \sum_{i=1}^{n} \int_0^s X_i(\xi_u) dw_i + \int_0^s X_0(\xi_u) du$$

is only defined up to a set of full Wiener measure and is discontinuous in the C_0-topology.

In particular, the map $g_t(w) = \xi_t$ is *non-differentiable* in the classical sense.

To understand the problem, suppose h is an arbitrary path in C_0. Then

$$Dg_t(w)h \equiv \lim_{u \to 0} \frac{g_t(w+uh) - g_t(w)}{u}. \qquad (6)$$

Let $B \subset C_0$ denote the set of w for which $g_t(w)$ exists.

Then $\gamma(B) = 1$, however $B \neq C_0$. In general the translations $w + uh \notin B$. So the difference quotient in (6) will not make sense.

Cameron-Martin Theorem. *Let $H \subset C_0$ denote the subspace of absolutely continuous paths h with finite energy, i.e.*

$$\int_0^T ||h'(s)||^2 ds < \infty.$$

If $\gamma(A) = 1$ and $h \in H$, then $\gamma(A+h) = 1$.

The C-M theorem implies that if $h \in H$ then $w + uh \in B$ on a set of full Wiener measure and the above problem does not arise.

It is possible to make sense of the derivative $Dg_t(w)$ as an operator in $L(H, \mathbf{R}^d)$.

The next calculation shows how it can be computed:

Example. Let $h \in H$. Then $Dg_t(w)h$ is obtained by formally differentiating wrt w in the equation

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

Thus $Dg_t(w)h = \eta_t$ where η satisfies the linear SDE

$$\eta_t = \sum_{i=1}^{n} \int_0^t DX_i(\xi_s)\eta_s dw_i(s) + \int_0^t \Big\{ \sum_{i=1}^{n} X_i(\xi_s)h_i' + DX_0(\xi_s)\eta_s \Big\} ds.$$

Define the *Malliavin covariance matrix*

$$\sigma = Dg_t(w)Dg_t(w)^*$$

By repeated *integration by parts* on the Wiener space, one can prove the following analogue of the finite-dimensional result:

Theorem (Malliavin, 1976). *If $\sigma \in GL(d)$ a.s., then the law of ξ_t^x admits a density wrt Lebesgue measure. Furthermore if*

$$(Det\,\sigma)^{-1} \in L^p,\ \forall p \geq 1 \qquad (7)$$

then the density is smooth.

The probabilistic proof of Hörmanders theorem then follows from:

Theorem. *If the vector fields $X_0, \ldots, X_n$ in the SDE satisfy the parabolic HC, then (7) holds.*

The proof of this result is difficult and requires delicate stochastic estimates. It was proved by Kusuoka & Stroock in 1985.

The integration by parts calculation on the Wiener space alluded to earlier is effected by means of the following infinite-dimensional divergence theorem.

Theorem. *Suppose $Z : C_0 \mapsto H$ is an H-differentiable map satisfying certain regularity conditions. Then for test functions Φ on C_0*

$$\int_{C_0} D\Phi(w) Z d\gamma = \int_{C_0} \Phi\, Div(Z) d\gamma, \quad where$$

$$Div(Z)(w) = < Z, w >_H - trace_H DZ(w).$$

For any basis vector $e_i \in \mathbf{R}^d$, define $r \in H$ by

$$r = Dg_t(w)^* \sigma^{-1} e_i.$$

Assuming (7) holds, we can use the above theorem to obtain

$$\left| \int_{\mathbf{R}^d} \frac{\partial \phi}{\partial x_i} d\nu \right| = \left| \int_{C_0} (\phi \circ g_t) Div(r) d\gamma \right|$$

$$\leq ||\phi||_\infty ||Div(r)||_{L^1(\gamma)}$$

and higher-order estimates by iteration. Thus we conclude that ν has a smooth density.

The Malliavin-Stroock formulation

The previous introduction of $Dg_t(w)$ as an element of $L(H, \mathbf{R}^d)$ has glossed over a serious technical difficulty. *The map g_t is not H-differentiable in a classical sense.*

The main technical achievement of the Malliavin calculus is to overcome this problem.

The essential ingredients, as formulated by Malliavin and reinterpreted and rigorized by Stroock and others, are as follows.

Consider the *Wiener Laplacian*

$$A\phi \equiv (Div \nabla_H)\phi$$

defined on H-C^1 functions $\phi : C_0 \mapsto \mathbf{R}$.

Remarkably, it turns out that A is *closable* and the domain of the closure $\bar{A}$ of A *contains maps defined as solutions of SDEs with smooth coefficients*.

The operator $\bar{A}$ is known as the *number operator*. It existed in the quantum physics literature prior to Malliavin s work, but its use in the present context was new.

For functions ϕ and $\psi \in Dom(\bar{A})$, the symmetric bilinear form

$$(\phi, \psi) = \frac{1}{2}\Big\{\bar{A}(\phi\psi) - \phi\bar{A}(\psi) - \psi\bar{A}(\phi)\Big\}.$$

is introduced.

If ϕ and ψ are H-C^1 functions then

$$(\phi, \psi)(w) =< D\phi(w)^*, D\psi(w)^* >_H .$$

The covariance matrix σ is defined by

$$\sigma_{ij} = \Big((g_t(w)_i, g_t(w)_j\Big).$$

The estimates required to prove regularity of the law of ξ_t are obtained using the operations $\bar{A}$ and $(\cdot, \cdot)$.

An elementary approach

Observation:

The *Ito map* $g : w \mapsto \xi$ defined by the SDE

$$\xi_s = x + \sum_{i=1}^{n} \int_0^s X_i(\xi_u) dw_i + \int_0^s X_0(\xi_u) du$$

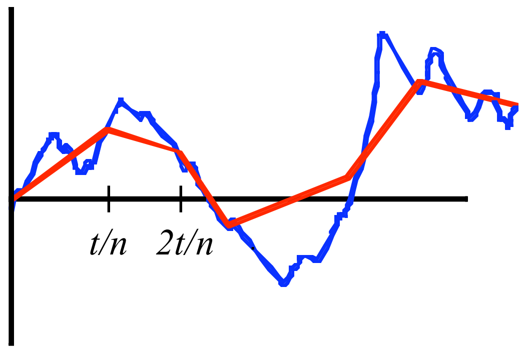
is smooth *when restricted to the Cameron-Martin space* H.

The restriction $\bar{g}$ is the map: $v \in H \mapsto \eta$, defined by the *classical* integral equation

$$\eta_s = x + \int_0^s \Big(\sum_{i=1}^{n} X_i(\xi_u) v_i' + X_0(\xi_u) \Big) du.$$

Since ξ is only defined on a set of γ-measure 1 and $\gamma(H) = 0$, the above statement needs clarifying.

Let $P_n w$ denote the piecewise linearization of w between the time-points $0, t/n, 2t/n, \ldots, t$.



Then as $n \to \infty$, $P_n w \to w$ in C_0 and

$$\bar{g}(P_n w) \to g(w), \quad a.s.$$

(Say $\bar{g}$ is the *skeleton* of g.)

This permits an elementary treatment of the problem. We perform classical integration by parts calculations, having replaced g by $\bar{g}$ and w by $P_n w$. Since the projections P_n have finite rank this reduces to *finite-dimensional* analysis of exactly the type we saw earlier.

The desired estimates are obtained by showing that the quantities that emerge are bounded as $n \to \infty$. (This where the work lies!)