3

AN INTRODUCTION TO MALLIAVIN CALCULUS AND SOME OF ITS APPLICATIONS

Dominique Michel

and

Etienne Pardoux

INTRODUCTION

A new part of the theory of stochastic processes, initiated by P. Malliavin, and developed among others by D. Stroock, J. M. Bismut, S. Watanabe and many others has recently emerged under the name of "the Malliavin Calculus." This is a stochastic calculus of variation which is in particular able to give powerful criterions for the law of a given functional of the Brownian motion to possess a density. One can distinguish three parts in the Malliavin Calculus. First of all, it relies on the use of differential operators which apply to functionals on Wiener space, associated Sobolev spaces, and an integration by parts formula which relates the derivations on Wiener Space and the Itô integral. Second, it contains a criterion, essentially in terms of the "Malliavin covariance matrix," or a similar quantity in Bismut's approach for a random vector defined on Wiener space to possess a density, or even a smooth density. The third circle of ideas concerns the way in which one can give in specific examples sufficient conditions for the above criterion to be satisfied. The most studied example is that of the solution of a stochastic differential equation, for which the Malliavin Calculus produces a probabilistic proof of Hörmander's "sum of squares theorem."

It has become clear during recent years that the Malliavin Calculus is not just a fancy way of making the probabilistic theory of diffusion processes self-contained (i.e. suppressing the need for borrowing results from analysis), but that it is a powerful tool for analyzing Wiener functionals, which is of interest to theoreticians as well as to specialists of stochastic control and nonlinear filtering.

This set of notes, which reflects the contents of a series of lectures given by the authors at the Systems Research Center, University of Maryland at College Park, aims at introducing and motivating the study of the Malliavin Calculus.

Section 3.4.1 explains the ideas of Malliavin and Bismut in the simple case of a finite dimensional probability space. In Section 3.4.2, we present the basic definitions and results of the Malliavin calculus on Wiener space, following Bismut and Zakai. Section 3.4.3 explains how the sufficient condition for the existence of a density follows from Hörmander's condition, in the case of the solution of a stochastic differential equation. In Section 3.4.4, we study the same problem for the conditional law in a filtering problem. Section 3.4.5 presents the application of the Malliavin calculus to stochastic partial differential equations, and how it allows to prove the non existence of finite dimensional statistics in certain nonlinear filtering problems. Finally in Section 3.4.6 we show how the differential calculus on Wiener space can be used to generalize Itô's and Stratonovich's stochastic integrals and differential calculus.

The aim of this text is to state the main results and introduce some of the relevant techniques. No detailed proof is given; only some proofs are sketched. We have included an extended bibliography at the end. Each Section is followed by bibliographical comments.

We want to thank the Systems Research Center and in particular its director John Baras for inviting us to give this series of lectures and for their warm hospitality.

PRELIMINARY

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\Phi : \Omega \to \mathrm{R}^n$ a measurable map. If P^Φ is the image measure of P under Φ the question is:

(i) does P^Φ have a density w.r.t. Lebesgue measure on R^n?

(ii) is this density regular?

In the case where Ω is finite dimensional, some answers can be given using the differential calculus on Ω. When $\Omega = C_0(\mathrm{R}_+, \mathrm{R}^m)$, the m dimensional Wiener space of continuous paths in R^m starting at 0, the Banach structure on Ω is too strong to give an answer to the problem, essentially because most of the Wiener functionals are not even continuous for this structure. The idea of Malliavin was to define a new notion of regularity of a Wiener functional and to use it to perform an integration by parts on Ω.

The answer to the above question is then given by the following lemma.

Lemma. Let μ be finite Radon measure on R^n. Assume that there exist $m \in \mathrm{N}$, $m \geq 1$ and $C \geq 0$ such that:

$$\forall \alpha, |\alpha| \leq m, \quad \forall \varphi \in C_b^\infty(\mathrm{R}^n), \quad \left| \int \mathrm{D}^\alpha \varphi(\mathrm{x}) \mu(\mathrm{dx}) \right| \leq \mathrm{C} \|\varphi\|_\infty$$

then μ has a density w.r.t. Lebesgue measure and, if $m > n$, this density is in $C_b^k(\mathrm{R}^n)$ with $k = m - n - 1$.

3.1 A FINITE DIMENSIONAL ANALOG OF MALLIAVIN CALCULUS

In this paragraph, $(\Omega, \mathcal{F}, \mathcal{P})$ is $(\mathrm{R}^\mathrm{N}, \mathrm{B}, \ \mathrm{g}(\mathrm{x})\,\mathrm{dx})$ where B is the Borel field on R^N and g a C^∞ positive function on R^N with integral 1.

$$\Phi \text{ is a } C^\infty \text{ map from } \mathrm{R}^\mathrm{N} \text{ to } \mathrm{R}^n$$

A natural assumption for the existence of a density for $P^\Phi = \Phi_*(P)$ is that Φ be a submersion, i.e. that the differential of Φ be of maximal rank, namely n. So, from now on, we assume:

$$\Phi \text{ is a submersion.} \tag{H_1}$$

If Φ is a submersion and is proper (i.e. B is a bounded set of $\mathrm{R}^n \Longrightarrow \Phi^{-1}(\mathrm{B})$ is a bounded set of R^N), it is not difficult to prove that P^Φ has a smooth density. If Φ is not proper, there is a density but nothing can be said about its regularity.

In that case, we look for necessary conditions which ensure that the following integration by parts formula is true:

$$\forall \varphi \in C_b^\infty(\mathrm{R}^n), \quad \forall i,\ 0 < i \leq n, \quad \exists\, B^\varphi \in L^1(\Omega),$$

$$\int D_i\, \varphi(x)\ P^\Phi(dx) = E(D_i\, \varphi \circ\ \Phi) = E(\varphi \circ \Phi \cdot B^\varphi) \qquad (3.1.1)$$

From (3.1.1) and Lemma 0, we get the answer.

In the following, we give two methods which lead to this integration by parts formula and which can be generalized to the infinite dimensional case of the Wiener Space.

3.1.1 The Malliavin Method

It is adapted to the particular case where P is the canonical gaussian measure on R^N, i.e. $g(y) = (2\pi)^{-\frac{n}{2}} \exp\left(-\frac{|y|^2}{2}\right)$.

The Ornstein-Uhlenbeck operator

$$L = \Delta - x \cdot \nabla$$

is self adjoint with respect to the gaussian measure i.e. :

$$E(LF \cdot G) = E(F \cdot LG), \qquad (3.1.2)$$

provided the above quantities are well defined.

Furthermore, let F, G and $F\,G \in D(L)$, where $D(L) \triangleq \{\Phi \in L^2(\mathrm{R}^N);$ $L\,\Phi \in L^2(\mathrm{R}^N)\}$, and define:

$$\Gamma(F,\ G) = \frac{1}{2}\ (L(FG)\ -\ FLG\ -\ GLF)$$

If F and G are C^1 functions on R^N, it holds:

$$\Gamma(F,\ G) = \nabla F \cdot \nabla G \qquad (3.1.3)$$

Formulas (3.1.2) and (3.1.3) are the keys for the integration by parts formula.

Let $J = D\,\Phi$ be the differential of Φ . J is an $(N,\ n)$ matrix and assumption (H_1) is equivalent to the invertibility of the $(n,\ n)$ matrix

$$A = J^* J = \left((\nabla\Phi^i \cdot \nabla\Phi^j)\right)_{i,j=1 \text{ to } n} \qquad (H_1)$$

which is called the Malliavin matrix.

The way of getting the integration by parts formula is the following.

$$E(D_i\,\varphi \circ \Phi) = \sum_{j,k} E(D_j\ (\varphi \circ \Phi)\ D_j\,\Phi_k\,A^{-1}_{k\,i})$$

Assume:

$$\varphi \circ \Phi,\ \Phi,\ \Phi(\varphi \circ \Phi),\ A_{ki}^{-1},\ A_{ki}^{-1}\,\Phi_k \text{ are in } D(L) \qquad (H_2)$$

Then

$$\begin{aligned} E(D_i\varphi \circ \Phi) &= \sum_k E\left(\nabla\varphi \circ \Phi \cdot \nabla\Phi_k \cdot A_{ki}^{-1}\right) \\ &= \frac{1}{2}\sum_k E\left(\{L(\varphi \circ \Phi \cdot \Phi_k) - \varphi \circ \Phi L\Phi_k - \Phi_k L\varphi \circ \Phi\}\, A_{ki}^{-1}\right) \\ &= E(\varphi \circ \Phi \cdot B^\varphi) \end{aligned}$$

where $B^\varphi = \frac{1}{2}\{\Phi_k\ L\ A_{ki}^{-1} - A_{ki}^{-1}\ L\ \Phi_k - L(\Phi_k\ A_{ki}^{-1})$.

It is possible to get, in a similar way, an integration by parts formula up to any order with allows to prove the existence of a smooth density for P^Φ.

Comments on the generalization to Wiener space. The natural way of generalizing the operator L to an infinite dimensional setting is to define it by its spectral decomposition.

If $\alpha = (\alpha_1, \ldots, \alpha_N)\ \in rmN^N$, define:

$$H_\alpha(y) = \prod_{i=1}^{N} H_{\alpha_i}\ (y_i),\ y \in \mathrm{R}^N$$

$$\text{where} \quad H_k(x) = \frac{(-1)^k}{(k!)^{\frac{1}{2}}}\ e^{\frac{x^2}{2}}\ \frac{\partial^k}{\partial x^k}\left(e^{\frac{-x^2}{2}}\right)$$

is the k-th Hermite polynomial on R.

The Hermite polynomials $(H_\alpha)_{\alpha\ \in\ rmN^N}$ are eignevectors of L. More precisely:

$$LH_\alpha = -|\alpha|\ H_\alpha.$$

Furthermore, the set $\{H_\alpha,\ \alpha \in rmN^N\}$ is an orthonormal basis of $L^2(\mathrm{R^N},\mathrm{P})$. So, the domain of L in $L^2(\mathrm{R^N},\mathrm{P})$ is:

$$D(L) = \{\Phi \in L^2(\mathrm{R^N},\mathrm{P}) \;/\; \Phi = \sum_{\alpha \in rmN^N} \lambda_\alpha H_\alpha,$$

$$\sum_{\alpha \in rmN^N} |\alpha|^2 \lambda_\alpha^2 < +\infty\}$$

and if $\Phi \in D(L)$, by definition

$$L\,\Phi = -\sum_{\alpha \in rmN^N} |\alpha| \lambda_\alpha H_\alpha.$$

All this program can be developed in the Wiener space setting: the multiple Wiener integrals stand for the Hermite polynomials, and the fact that they form an orthonormal basis of $L^2(\Omega, P)$ is the Wiener chaos decomposition of a square integrable functional of the brownian motion.

So, if Φ is in $L^2(\Omega, P)$ with its values in R^n, the assumptions for P^Φ to have a smooth density are: (H_1') the Mallavian matrix:

$$A = \left(\left(\Gamma\left(\Phi^i,\ \Phi^j\right)\right)\right)_{i,j=1 \ to \ n}$$

is invertible.

(H_2) Φ and A^{-1} are "regular" and have good properties of integrability.

This can be rewritten in the following way:

(H_1'') the Malliavin matrix A is invertible and $|\det A|^{-1} \in L^p(\Omega,\ P)\ \forall p \geq 1$.

(H_2') Φ is "regular" and has good properties of integrability as well as its "derivatives."

3.1.2 The Bismut method

The idea of Bismut is not related to the Gaussian character of the measure but to the quasi invariance property of the measure with respect to some transformations on Ω.

If $g = 1$, (of course, P is not a probability measure but, any-way, we can look for an integration by parts formula), P is Lebesgue measure and, so, is invariant under translations.

Take $H:\ \mathrm{R^N} \longrightarrow \mathrm{R}$ a C^1 map, P integrable. $\int_{\mathrm{R^N}} \varphi \circ \Phi(y+a)\ H(y+a)\,dy$ is independant of a. By differentiating $w.r.t.a$, we get, under some more regularity assumptions,

$$\int_{\mathrm{R^N}} \sum_{j=1}^{n} D_j\, \varphi \circ \Phi(y)\ D_i\, \Phi_j(y)\ H(y)\,dy = -\int_{\mathrm{R^N}} \varphi \circ \Phi(y)\ D_i\, H(y)\,dy \quad (3.1.4)$$

Under assumption (H_1'), define:

$$H_{ik}(y) = \sum_{j=1}^{N} D_j\ \Phi_i(y)\ A_{jk}^{-1}(y)$$

Write (3.1.4) with $H = H_{ik}$ and sum over i. We get:

$$\int_{\mathrm{R^N}} D_k\, \varphi \circ \Phi(y)\,dy = -\int_{\mathrm{R^N}} \varphi \circ \Phi(y) \sum_{i,j=1}^{N} D_i\left(D_j\ \Phi_i\ A_{jk}^{-1}\right)(y)\,dy$$

This is a good formula of integration by parts if:

$$\sum_{i,j=1}^{N} D_i\left(D_j\ \Phi_i\ A_{jk}^{-1}\right) \in L^1(\mathrm{R^N},\mathrm{dx}) \qquad (H_3)$$

When $g \neq 1$, the measure P has a priori no invariance property but it is quasi invariant under any C^1 differomorphism Ψ (i.e. $\Psi_* P$ is absolutely continuous $w.r.t.\ P$) because of the change of variables formula:

$$dP^\Psi(y) = \frac{g\left(\Psi^{-1}(y)\right)}{g(y)}\ |\det \Psi|^{-1}(y)\ dP(y)$$

So, the integral is:

$$I(\Psi) = \int_{\mathrm{R^N}} \varphi \circ \Phi \circ \Psi(y)\ dP^\Psi(y)$$

In order to differentiate $w.r.t\ \Psi$, we choose a one parameter family of diffeomorphisms, for example the flow associated to a vector field X an $\mathrm{R^N}$ i.e $\Psi_t(y)$ is the solution at time t of the equation:

$$\begin{cases} dy_t = X(y_t) \\ y_0 = y \end{cases}$$

Now, we write: $\frac{d}{dk} I(\Psi_t)_{t=0} = 0$:

$$\int_{\mathrm{R^N}} X(\varphi \circ \Phi)(y)\,dP(y) = -\int_{\mathrm{R^N}} \varphi \circ \Phi(y)\ \frac{\operatorname{div} gX}{g}(y)\,d\,P(y) \qquad (3.1.5)$$

$\frac{\text{div}\ gX}{g}$ is the divergence of the vector field X under the measure P.

In a similar way that we have chosen H in the Malliavin method, we have to choose X in order to get the integration by parts formula.

Since

$$X(\varphi \circ \Phi) = \sum_{i=1}^{N}\sum_{j=1}^{n} X_i\, D_j\, \varphi \circ \Phi\, D_i\, \Phi_j$$
$$= \sum_{j=1}^{n} D_j\, \varphi \circ \Phi\, (X \cdot \nabla \Phi_j), \tag{3.1.6}$$

we need the following assumption:

(H_4) There exist vector fields $Y_1, \ldots, Y_n$ such that the matrix $((Y_i \cdot \nabla \Phi_j))$ is invertible.

Actually, if (H_4) is valid, let B be the inverse matrix, and define:

$$X^i = \sum_{j=1}^{n} B_{i\,j}\, Y_j$$

Then: $X^i(\varphi \circ \Phi) = D_i\varphi \circ \Phi$

So, if

$$\forall i = 1 \ldots n, \quad \frac{\text{div}\ gX_i}{g} \in L^1(\Omega,\ P).$$

the integration by parts formula holds.

In the finite dimensional setting, it is easy to show that (H_1') and (H_4) are equivalent. In the infinite dimensional setting, this is no longer the case.

Comments about the generalization. As we are only interested in the behaviour of the flow Ψ_t near $t = 0$, it is equivalent and easier to take for Ψ_t the linearized flow $y + t\,X(y)$. When Ω is the Wiener Space, we need transformations of $\Omega, \omega \longrightarrow \omega + t\,X(\omega)$, leaving the Wiener measure quasi invariant. Some of them are given by Girsanov theorem.

Girsanov's Theorem. Let u: $[0,\ 1] \times W_1 \longrightarrow$ R be an adapted process such that,

$$E\left(\exp\left(\lambda \int_0^1 u_s^2\, ds\right)\right) < \infty \quad \text{for a } \lambda > \frac{1}{2}.$$

Then the law of the process $\omega_. + \int_0^. u_s\, ds$ is absolutely continuous with respect to Wiener measure, the density being:

$$exp\left(-\int_0^1 u_s\, d\omega_s - \frac{1}{2}\int_0^1 u_s^2\, ds\right)$$

By means of this theorem, it is possible to get an integration by parts formula analog to (3.1.5) on Wiener Space. By the way, it is necesary to define the analog $X \cdot \nabla \Phi$ when Φ is a Wiener satisfies the assumptions of Girsanov theorem.

3.1.3 Bibliographical comments

This section was inspired by similar expositions in Bismut $[A1]$ and Stroock $[A13]$.

3.2 THE BISMUT - ZAKAI METHOD FOR MALLIAVIN CALCULUS

Let Ω denote the space $C(\mathrm{R}_+,\ \mathrm{R}^m)$ equipped with the topology of uniform convergence on compact sets, $\mathcal{F}$ the Borel σ-field on Ω, P the standard Wiener measure, and let $\{W_t(\omega) = \omega(t), t \geq 0\}$.

3.2.1 The polynomial functionals

Definition. A polynomial Wiener functional is a map $\Phi:\ \Omega \longrightarrow$ R defined by:

$$\Phi = f\left(\delta_{i_1}(h_1),\ .,\delta_{i_p}\ (h_p)\right)$$

Where:

(i) f is a polynomial function on R^p

(ii) $\{h_1, \ldots, h_p\}$ is an orthonormal set of $L^2(\mathrm{R}_+)$

(iii) $\delta_i\ (h) = \int_0^{+\infty} h(s)\ dW^i(s)$

$\mathcal{P}(\Omega)$ is the set of polynomial Wiener functionals.

Proposition. The set of polynomial functionals on Ω is dense in $L^2(\Omega,\ P)$.

3.2.2 The directional derivative of a Wiener functional

If $\Phi(\omega) = f\left(\delta_{i_1}\ (h_1), \ldots, \delta_{i_p}(h_p)\right)$ is a polynomial functional and the h_i are C^1 functions with compact support, then:

$$\delta_i(h_j) = -\int_0^{+\infty} W_s^i\, \dot{h}_s^j\, ds$$

and Φ is Frechet differentiable i.e. , if $g \in L^2(\mathrm{R}+;\ \mathrm{R}^m)$

$$D_G\ \Phi = \lim_{\varepsilon \to 0} \frac{\Phi\left(\omega + \varepsilon \int_0^{\cdot} g(s)\ ds\right) - \Phi(\omega)}{\varepsilon}$$
$$= \sum_{j=1}^{p} \sum_{\{k;\ i_k = j\}} \frac{\partial f}{\partial x_k}\ (\delta(h))\ (h_k,\ g_j)$$

where (,) denotes the scalar product in $L^2(\mathrm{R}_+;\ \mathrm{R}^m)$ and $G_s = \int_0^s g(u)\ du$.
This motivates the following definition:

Definition. If Φ is a polynomial functional on Ω and $g \in L^2(\mathrm{R}_+;\ \mathrm{R}^m)$, define:

$$D_G\ \Phi\ (\omega) = \sum_{j=1}^{p} \sum_{\{k;\ i_k = j\}} \frac{\partial f}{\partial x_k}\ (\delta(h))\ (h_k,\ g_j)$$
$$\text{where}\quad G_s = \int_0^s g(u)\ du.$$

Notation. Denote by $(D_t\ \Phi)_{t \in \mathrm{R}+}$ the element of $L^2(\mathrm{R}^+ \times \Omega,\ \mathrm{R}^m)$ such that: $D_G\ \Phi = (D \cdot \Phi, g)$ i.e. $D_t^j \Phi = \sum_{\{k;\ i_k = j\}} \frac{\partial f}{\partial x_k}\ (\delta(h))\ h_k\ (t)$
On $\mathcal{P}(\Omega)$, we define the following norm:

$$\|\Phi\|_{2,\ 1} = \left\{ E\left(\Phi^2 + \sum_{i=1}^{m} \int_o^{+\infty} \left(D_t^i\ \Phi\right)^2\ dt \right) \right\}^{\frac{1}{2}}$$
$$= \left[E\left(f^2\left(\delta(h)\right) + \|\nabla f\|^2\left(\delta(h)\right)\right)\right]^{\frac{1}{2}}$$

Definition. $H^{2,1}$ is the closure of $\mathcal{P}(\Omega)$ in $L^2(\Omega)$ for the norm $\|\ \|_{2,1}$

It can be shown that the gradient operator D extends to $H^{2,1}$. The extension is called the derivative on Wiener Space.

3.2.3 The integration by parts formula

Theorem. If Φ is in $H^{2,1}$ and $u \in L^2((0,\ t) \times \Omega,\ \mathrm{R}^m)$ is adapted, then:

$$E\left(\sum_{i=1}^{m} \int_0^t D_s^i\ \Phi\ u_s^i\ ds \right) = E\left(\Phi \sum_{i=1}^{m} \int_0^t u_s^i\ dw_s^i \right)$$

This is an extension to Wiener space of the formula (5) of Section one.

Remark. The gradient operator D can be thought of as a linear continuous operator from $H^{2,1} \subset L^2(\Omega)$ into $L^2(\mathrm{R}_+ \times \Omega;\ \mathrm{R}^m)$. Clearly, D has an adjoint D^* which maps

$L^2(\mathrm{R}_+ \times \Omega; \mathrm{R}^m)$ into $(H^{2,1})'$, where $(H^{2,1})'$, the dual space of $H^{2,1}$, is a space of "distributions over Wiener space." Let us now define the operator δ as follows.

$$\text{Dom}\ \delta = \{u \in L^2(\mathrm{R}_+ \times \Omega;\ \mathrm{R}^m);\ \mathrm{D}^*\mathrm{u} \in \mathrm{L}^2(\Omega)\}$$
$$\text{For } u \in \text{Dom}(\delta),\quad \delta\ u \triangleq D^* u$$

It follows from Riesz's representation theorem that an equivalent definition of δ is as follows.

Definition. Dom $\delta = \{u \in L^2(\mathrm{R}_+ \times \Omega, \mathrm{R}^m)$, such that there exists a constant c with: $|E(D \cdot \Phi, u)| \leq c\|\Phi\|_2,\ \forall\ \Phi \in H^{2,1}\}$.

For $u \in$ Dom $\delta, \delta u$ is the unique random variable which satisfies:

$$E(D\ .\ \Phi,\ u) = E(\Phi\ \delta u) \qquad \forall\ \Phi \in H^{2,1}$$

The operator δ is called the Skorohod integral. It will be studied in § 6 below.

3.2.4.1 Theorem

Let Φ be in $L^2(\Omega)$. Assume that Φ is in $H^{2,1}$ and that there exists $u \in$ Dom δ such that:

(i) $(D\ \cdot\ \Phi,\ u) \in H^{2,\ 1}$

(ii) $(D\ \cdot\ \Phi, u) > 0$ a.s.

Then P^{Φ} has a density.

Proof: Let $\varphi \in C_b^{\infty}(\mathrm{R})$. Then $\varphi \circ \Phi \in H^{2,1}$.

Furthermore, if $\Psi = \dfrac{1}{\varepsilon + (D\ .\ \Phi,\ u)}$, then $\varphi \circ \Phi\ .\ \Psi$ is in $H^{2,\ 1}$ by assumption.
Then, the following integration by parts holds:

$$E\left(\left(D\left(\varphi \circ \Phi \cdot \Psi\right),\ u\right)\right) = E\left(\left(\varphi \circ \Phi\right) \Psi \delta u\right)$$
$$\Longleftrightarrow E\left(\varphi' \circ \Phi(D \cdot \Phi,\ u)\Psi\right) = E\left[(\varphi \circ \Phi)(\Psi\ \delta u) - (D \cdot \Psi,\ u)\right]$$

So, if P_ϵ is the probability law on Ω such that $\dfrac{dP_\epsilon}{dP} = \dfrac{(D \cdot \Phi, u)}{\varepsilon + (D \cdot \Phi, u)}$, then P_ϵ^{Φ} possesses a density. Now, if B is a borel set of Lebesgue measure zero: $P_\epsilon^{\Phi}\ (B) = P_\epsilon\left(\Phi \in B\right) = 0$. By the monotone convergence theorem, $P\left(\Phi \in B\right) = 0$. So, P^{Φ} is absolutely continuous w.r.t Lebesgue measure.

Further integrations by parts lead to the smoothness of the density. Let us now state the multidimensional version of Theorem 3.2.4.1.

3.2.4.2 Theorem

Let $\Phi = (\Phi_1, \dots, \Phi_d)' \in (L^2(\Omega))^d$. Assume that Φ_i is in $H^{2,1}, i = 1, \dots, d$, and that there exists $u_1, \dots, u_d \in \text{Dom}\ \delta$ such that:

(i) $(D \cdot \Phi_i,\ u_j) \in H^{2,1}; 1 \le i, j \le d$

(ii) $(D \cdot \Phi_i,\ u_j)_{1 \le i,j \le d} > 0$ a.s.

Then P^Φ has a density with respect to Lebesgue measure on R^n.

3.2.5 Bibliographical comments

The abstract presentation of Bismut's methodology (independently of the particular application to SDE's) can be found (in the case of one- dimensional processes) in Zakai [A17], from which Theorem 2.4.1 is borrowed. For a systematic exposition of the theory of Sobolev spaces over Wiener space, we refer to Watanabe [A15].

3.3 THE APPLICATION TO STOCHASTIC DIFFERENTIAL EQUATIONS

In this paragraph, we study the particular case where Φ is the solution at time t of a stochastic differential equation, i.e.

$$dx_t = X_0(x_t)\ dt + \sum_{i=1}^{m} X_i(x_t) \circ dw_t^i \tag{3.3.1}$$

$X_0,\ X_1, \dots,\ X_m$ are C_b^∞ maps from R^n to R^n that we consider as vector fields on R^n in the sense that we identify X_i with the first order partial differential operator $\sum_{j=1}^n X_i^j \frac{\partial}{\partial x_j}$. Equation (1) is understood in the sense of Stratonovich.

It is natural to associate to (1):

its infinitesimal generator: (3.3.2)

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

a controlled ordinary differential equation: (3.3.3)

$$\dot{X}_t^u = X_0(x_t^u) + \sum_{i=1}^{m} X_i(x_t^u)\, \dot{u}_t^i$$

$$x_0^u = x_0$$

where $u \in H^1(\mathrm{R},\ \mathrm{R}^m)$

The links between (3.1), (3.2) and (3.3) are the following:

3.3.1 Theorem

a. If μ_t is the law of x_t, solution of (3.3.1), then μ_t is the solution of the following PDE written in the distributional sense:

$$\frac{d\mu_t}{dt} - L^*\ \mu_t = 0$$

where L^* is the adjoint of L in the L^2 sense.

b. Support theorem:

If $A(t, x_0) = \{X_t^u,\ u \in H^1\}$ and

$A(x_0) = \cup_{t>0} A(t,\ x_0)$, then :

support $\mu_t = \overline{A(t,\ x_0)}$

support $\left(\int_0^\infty e^{-\alpha t}\ \mu_t\ dt\right) = \overline{A(x_0)}$.

So, the existence of a smooth density for μ_t is in relation with:

1. the hypoellipticity of the operation $\frac{\partial}{\partial t} - L^*$. Recall that a partial differential operator A is said to be hypoelliptic if:

$$A\varphi = \Psi \text{ on } U$$

$$\Psi_{|U} \in C^\infty(U) \Longrightarrow \varphi_{|U} \in C^\infty(U)$$

U being an open set of R^n.

2. the accessibility property for the controlled ODE (3.3.3).

The assumptions needed for solving this problem are in terms of the following Lie algebras of vector fields:

1. $\mathcal{A} = LA(X_0,\ X_1, \dots, X_m)$
2. $\mathcal{B} = LA(X_1, \dots, X_m)$
3. $\mathcal{J}$ = ideal generated by $\mathcal{B}$ in $\mathcal{A}$

The main results concerning these problems are presented in the following picture:

3.3.2 Theorem

$\dim \mathcal{A}(x) = n,\ \forall x \in \mathrm{R}^n$	$\dim \mathcal{J}(x) = n,\ \forall x \in \mathrm{R}^n$	$\dim \mathcal{B}(x) = n,\ \forall x \in \mathrm{R}^n$
L is hypoelliptic	$\frac{\partial}{\partial t} - L^*$ is hypoelliptic	$\frac{\partial}{\partial t} - L^*$ is hypoelliptic
$\int_0^{+\infty} e^{-\alpha t}\mu_t\, dt$ has a smooth density	μ_t has a smooth density.	μ_t has a smooth density, everywhere strictly positive.
$\overset{\circ}{A}(x) \neq \emptyset,\ \forall x \in \mathrm{R}^n$	$\overset{\circ}{A}(t,x) \neq \emptyset,\ \forall x \in \mathrm{R}^n$ $t > 0$	$A(t,x) = \mathrm{R}^n,$ $\forall x \in \mathrm{R}^n$ $t > 0$

3.3.3 Comments

1. The reason why the Lie algebras occur in the accessibility problem is essentially the following: if the vector fields X_i commute, for a control u with piecewise constant derivative, x_t^u is a composition of the flows associated to each X_i. If they don't, the flows associated to the brackets occur because of the parallelogram law: follow the vector field X_i during time ε, then X_j during the same time, then $-X_i$ and $-X_j$; up to the second order, it is as if you had followed the field $[X_i,\ X_j]$ during time 4ε, up to a term of order ε^2.

2. The implication $\{\dim A(x) = n, \forall x \in \mathrm{R}^n \implies \mathrm{L}$ is hypoelliptic$\}$ is Hörmander's theorem. To obtain the implication $\{\dim \mathcal{J}(x) = n, \forall\ x \in \mathrm{R}^n \implies \frac{\partial}{\partial t} - \mathrm{L}^*$ is hypoelliptic$\}$, it suffices to apply Hörmander's theorem to the operator on R^{n+1} $\frac{\partial}{\partial t} - \mathrm{L}^*$: then occurs the Lie algebra $\mathcal{C} = LA\left(\frac{\partial}{\partial t} + \widetilde{X}_0, X_i \ldots X_n\right)$ where $\widetilde{X}_0$ is the vector field defined by:

$$\frac{\partial}{\partial t} - L^* = \frac{\partial}{\partial t} + \widetilde{X}_0 - \frac{1}{2}\,\Sigma X_i^2 + c$$

c being a zero degree operator is a function. Actually, $\widetilde{X}_0 = -X_0 + \sum_{i=1}^m \lambda_i X_i$, and the two following conditions are equivalent:

a. $\dim\ \mathcal{C}(t,\ x) = n+1 \quad \forall (t,\ x)\ \in \mathrm{R}^{n+1}$

b. $\dim\ \mathcal{J}(x) = n \quad \forall_x \in \mathrm{R}^n$

3. Beware that the fact that the support of μ_t is the whole space does not imply that μ_t has a smooth density. The following example is in R^2.

$$X_1 = \frac{\partial}{\partial x}$$
$$X_2 = \frac{\partial}{\partial x} + \varphi(x)\ \frac{\partial}{\partial y}$$

where $\varphi : \mathrm{R} \longrightarrow \mathrm{R}$ is a smooth function whose value is 0 on R$-$.

$$\dim \quad \mathcal{B}(x,\ y) = 1 \qquad \text{if } x \leq 0$$
$$\dim \quad \mathcal{B}(x,\ y) = 2 \qquad \text{if } x > 0$$

So, on $\mathrm{R}^- \times \mathrm{R}$, the system $(X_1,\ X_2)$ does not satisfy the range condition. Nevertheless:

$$A(t,\ x,\ y) = \mathrm{R}^2, \qquad \forall t\ \in \mathrm{R}^+.$$

and certainly μ_t does not have a density when $x_0 < 0$. But, we shall see, using Malliavin's calculus, that μ_t has a smooth density when $x_0 > 0$ because the rank condition is needed only at the starting point.

Stroock showed that, if the rank condition fails on a submanifold N of codimension $n-1$, under some technical assumptions on the degeneracy, then μ_t has a smooth density even if the starting point is on N.

3.3.4 The Malliavin, Bismut method

Theorem. Let x_t be the solution of the SDE (1) at time t and μ_t its law. Then, if:

$$\dim \mathcal{J}(x_0) = n,$$

μ_t has a smooth density.

Proof: The proof proceeds in three steps.

First step: $x_t \in H^{2,1}$

This is done by solving (1) by the Picard iteration method, using the rules of calculus for the gradient and passing to the limit. The gradient $D_s x_t$ is a solution of the following SDE:

$$D^i_s x_t = X_i(x_s) + \int_s^t X_0'(x_r)\, D^i_s x_r\, dr + \int_s^t X_j'(x_r)\, D^i_s x_r \circ dw^j_r$$

Let Φ_t be the $n \times n$ matrix valued process solution of:

$$\Phi_t = I + \int_0^t X_0'(x_s)\, \Phi_s\, ds + \int_0^t X_j'(x_s)\, \Phi_s \circ dw^i_s.$$

Remark that Φ_t is nothing else but the jacobian matrix of x_t w.r.t. x_0.

$$\text{Then:} \quad D^i_s x_t = \Phi_t\, \Phi_s^{-1}\, X_i(x_s)$$

Second step: Choose u such that $((Dx^i_t,\ u_j))_{i,j} > 0$. Look at the bilinear form associated to $((Dx^i_t,\ u_j))_{i,j}$,

$$((D_{u_j} x^i_t) \cdot p,\ q) = \sum_{i,\ j} Du_j\, x^i_t\, p_j\, q_i$$

$$= \int_0^t \Phi^j_t\, \Phi_s^{-1}\, X_r\, u^k_j\, p_j\, q_i\, ds$$

$$= \int_0^t \sum_k (\Phi_t\, \Phi_s^{-1}\, X_k, q)\, (u^k,\ p)\, ds$$

We get the best rank if $\operatorname{span}((u^1)', \ldots, (u^m)') = \operatorname{span}(\Phi_t \Phi_s^{-1} X_1, \ldots, \Phi_t \Phi_s^{-1} X_m)$ and, so in particular, if $(u^k)' = \Phi_t\, \Phi_s^{-1} X_k$, where $(u^k)'$ is the column vector whose j-th component is u^k_j.

In that case, the Bismut approach leads to the same condition as the Malliavin approach i.e. to check the invertibility of the Malliavin covariance matrix $A_t = ((Dx^i_t, Dx^j_t))_{i,j}$.

Third step: The invertibility of A_t

Remark first that:

$A_t = \Phi_t C_t \Phi_t^{-1}$

with $C_t = \int_0^t \sum_k (\Phi_s^{-1} X_k)\ (\Phi_s^{-1} X_k)^t (x_s)\, ds$

From the invertibility of Φ_t and its boundedness properties, it is clear that it is equivalent to prove $E\left((\det\ A_t)^{-p}\right) < +\infty$ or $E\left((\det\ C_t)^{-p}\right) < +\infty$. We first consider the elliptic case and then the hypoelliptic case.

3.3.4.1 Theorem.

Assume $X_1 \ldots X_m$ generate R^n at x_0 then C_t is positive definite a.s. and so μ_t has a density.

Proof: For each $q \in \mathrm{R}^n$, $\sum_{k=1}^m (X_k(x_0), q)^2$ is strictly positive. So, by continuity, $(C_t\, q, q)$ is strictly positive a.s.

3.3.4.2 Theorem.

Assume $\mathcal{J}(x_0) = \mathrm{R}^n$. Then C_t is positive definite a.s. and so μ_t has a density

Proof: Define $U_s = \operatorname{span}\{\Phi_s^{-1} X_i,\ 1 \le i \le m\}$, $V_t = \ \cup\ U_s$ and $V_t^+ = \ \cap\ V_s$. By the 0.1. law, V_0^+ is a non random space a.s. Assume $V_0^+ \neq \mathrm{R}^n$ and take q in $(V_0^+)^\perp$. Define $\tau = \inf\{s > 0,\ V_s \neq V_0^+\}$. It is clear that, a.s., $\tau > 0$. Then, for $t < \tau$:

$$(q,\ \Phi_t^{-1} X_i) = 0, \quad \forall\, i = 1, \ldots, m.$$

By Itô's formula it holds:

$$(\Phi_t^{-1} X_i, q) = (X_i, q) + \int_0^t (\Phi_s^{-1} [X_0, X_i]\,; q) ds$$

$$+ \int_0^t (\Phi_s^{-1} [X_j, X_i]\,, q) \circ dW^j$$

Annihilating the martingale part, we get:

$$(q,\ \Phi_t^{-1}[X_j, X_i]) = 0, \quad \forall t < \tau.$$

and by iteration: $(q, \Phi_t^{-1}[X_k, [X_j, X_i]]) = 0$. Then, annihilating the bounded variation part, it comes, $(q, \Phi_t^{-1}[X_0, X_i]) = 0$

Recursively, we can prove that q is orthogonal to every bracket which contradicts the assumption. So, $V_0^+ = \mathrm{R}^n$, and this implies that each V_s is R^n.

The problem of the smoothness of the density is more complicated because we need to prove that $E\left((\det A_t)^{-p}\right) < +\infty$. But, the first idea is the same, namely to give a kind of Taylor expansion of $\Phi_t^{-1}X_i$ by means of the brackets. Before looking at other properties of the density we give a classical example of a hypoelliptic diffusion.

3.3.5 Example: the Heisenberg group

The Heisenberg group H_3 is $\mathrm{C} \times \mathrm{R}$ together with the group law:

$$(z,t)\ (z',t') = (z+\bar{z}',\ t+t'+2\ \mathrm{Im}\ z\bar{z}').$$

X, Y, T are the left invariant vector fields on H_3 defined by:

$$X = \frac{\partial}{\partial x} + 2y\frac{\partial}{\partial t},\quad Y = \frac{\partial}{\partial y} - 2x\frac{\partial}{\partial t},\quad T = \frac{\partial}{\partial t}.$$

They form a basis of the tangent space. Let ξ be the diffusion on H_3 associated to the Kohn laplacian: $\Delta_K = X^2 + Y^2$. Because, $[X,Y] = 4T$, ξ is a hypoelliptic diffusion. It is a solution of the system:

$$d\xi = \begin{pmatrix} 1 \\ 0 \\ 2\xi_2 \end{pmatrix} dw_1 + \begin{pmatrix} 0 \\ 1 \\ -2\xi_1 \end{pmatrix} dw_2$$

So, if $\xi_0 = 0$, we get,

$$\xi = \begin{pmatrix} w_1(t) \\ w_2(t) \\ 2\left(\int_0^t w_2\, dw_1 - \int_0^t w_2 dw_1\right) \end{pmatrix}.$$

The last component is 4 times the area swipped out by the vector $(w_1,\ w_2)$. Paul Levy gave a formula for the density of the law of this area knowing $(w_1,\ w_2)$.

3.3.6 Further properties of the density

3.3.6.1 Positivity Theorem.

If $\mathcal{B}(x) = \mathrm{R}^n$, $\forall x \in \mathrm{R}^n$, then the density of μ_t, $p(t,\ x)$, is strictly positive for all $(t,\ x) \in \mathrm{R}_+^* \times \mathrm{R}^n$.

Proof: We use the Feynmann-Kac formula and a time reversal. Let $\alpha \in (0,\ t)$ and define

$$q(s,\ x) = p(t-s,\ x), \quad \text{for } s \le t-\alpha$$

q is a solution of the following PDE:

$$\left\{\begin{array}{ll} \dfrac{\partial q}{\partial s} + L^* q = 0, & s \le t-\alpha \\ q(t-\alpha) \text{ given.} & \end{array}\right\}.$$

where $L^* = \frac{1}{2}\sum_1^m X_i^2 - \widetilde{X}_0 - c$ as we saw before. Let y_s^x be the diffusion associated to the operator $\frac{1}{2}\sum_1^m X_i^2 - \widetilde{X}_0$, starting at x.

By the Feynmann-Kac formula, it holds:

$$p(t,x) = q(0,x) = E\left(q(t-\alpha, y_{t-\alpha}^x)\exp\left(\int_o^{t-\alpha} c(y_s^x)\, ds\right)\right)$$

Since $x \to q(t-\alpha, x) = p(\alpha, x)$ is the smooth density of a probability measure, there exists a ball $B \subset \mathrm{R}^n$ and a strictly positive constant $k \in \mathrm{R}$, such that:

$$q(t-\alpha, x) \ge k, \quad \forall x \in B$$

Moreover, as $|c|$ is bounded, there exists $\bar{c} \in \mathrm{R}$, such that, $c(x) \ge \bar{c}$, $\forall x \in \mathrm{R}^n$. We get then:

$$p(t,x) \ge k\exp(\bar{c}(t-\alpha))\ P\left(y_{t-\alpha}^x \in B\right)$$

Because $\mathcal{B}(x) = \mathrm{R}^n$, $\forall x \in \mathrm{R}^n$, the support of the law of $y_{t-\alpha}^x$ is R^n. This implies that:

$$P\left(y_{t-\alpha}^x \in B\right) > 0.$$

Remark. The statement of the theorem can be improved as follows.

Let $t > 0$. Suppose that for some $s \in (0,t)$, $A(s,x) = \mathrm{R}^n$ for all $x \in \mathrm{R}^n$. Then $p(t,\ x) > 0$ for all $x \in \mathrm{R}^n$.

Indeed, the hypothesis has been used only to insure that for some $\alpha \in (0,t)$ and all $x \in \mathrm{R}^n$, the support of the law of $y^x_{t-\alpha}$ is R^n. That support coincides with $\bar{B}(t-\alpha, x)$, where $B(t,x)$ is the accessible set at time t, starting from x, for the control system:

$$\dot{y}^u_t = -X_0(y^u_t) + \sum_{i=1}^{m} a_i(y^u_t) X_i(y^u_t) + \sum_{i=1}^{m} X_i(y^u_t)\dot{u}^i_t$$

which is the same as that associated to:

$$\dot{y}^u_t = -X_0(y^u_t) + \sum_{i=1}^{n} X_i(y^u_t)\dot{u}^i_t$$

It finally follows easily from a time reversal argument that $A(s,x) = \mathrm{R}^n$ for all $x \in \mathrm{R}^n$ if and only if $B(s, x) = \mathrm{R}^n$ for all $x \in \mathrm{R}^n$.

3.3.6.2 The regularity with respect to the starting point

Let $p(t, x_0, x)$ be the density of μ_t, x_0 being the starting point of the diffusion. As a function of (t, x_0), p satisfies the following backward equation:

$$\frac{\partial p}{\partial t} + Lp = 0.$$

Then, if $\Im(x) = \mathrm{R}^n$, $\forall \mathrm{x} \in \mathrm{R}^n$, $\mathrm{p}(\mathrm{t}, \mathrm{x}_0, \mathrm{x})$ is smooth w.r.t. x_0, by Hörmander's theorem. One way of recovering this result by using the Malliavin calculus is to proceed as in the previous paragraph i.e. if φ is a smooth function with compact support:

$$\int p(t, x_0, x)\varphi(x_0)\, dx_0 = E\left(\varphi(y^x_t)\exp\left(\int_0^t c(y^x_s)\, ds\right)\right)$$

by Feynmann-Kac's formula.

Malliavin's calculus applied on the right hand side allows one to get estimates of the type:

$$|\int p(t, x_0, x) D^\alpha \varphi(x_0)\, dx_0| \leq c_\alpha \|\varphi\|$$

and then to conclude.

3.3.6.3 The hypoellipticity of L

If $\dim \Im(x) = n, \forall x \in \mathrm{R}^n, \mu_t$ has a density which is smooth with respect to the forward and backward variables. Using localization and integration w.r.t. x_0 Stroock showed that L is hypoelliptic and, so, recovered the result of Hörmander's theorem.

Assume now the more general condition, $\dim \mathcal{A}(x) = n \quad \forall x \in \mathrm{R}^n$.

Stroock's method. It uses the following trick to go back to the previous case. Let W^{m+1} be a Wiener process independent of $W^1, \dots, W^m$, and define:

$$\tilde{x}_t = \begin{pmatrix} x \int_0^t \varrho(W^m(s))ds \\ W^{m+1}(t) \end{pmatrix}$$

where ϱ is a C_b^∞ strictly positive function. Then $\tilde{x}_t$ is a solution of the following SDE:

$$d\tilde{x}_t = \widetilde{X}_0(\tilde{x}_t)\, dt + \sum_1^m \widetilde{X}_k(\tilde{x}_t) \circ dW^k_t + \widetilde{X}_{m+1}(\tilde{x}_t) \circ dW^{m+1}_t$$

with:

$$\widetilde{X}_0 = \begin{pmatrix} \rho X_0 \\ 0 \end{pmatrix}, \quad \widetilde{X}_k = \begin{pmatrix} \sqrt{\rho} X_k \\ 0 \end{pmatrix}, \quad \widetilde{X}_{m+1} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

and $\Im\left(\widetilde{X}_0, \widetilde{X}_1, \dots, \widetilde{X}_{m+1}\right) = \mathrm{R}^{m+1}$ at each point. So, $\widetilde{L} = \frac{1}{2}\sum_1^{m+1} \widetilde{X}_i^2 + \widetilde{X}_0$ is hypoelliptic. As $\widetilde{L}_{|\mathrm{R}^m} = \rho(x_{m+1})L$, the same property holds for L.

Bismut's method. This is a geometric method. We shall only give an idea of the method in the case where moreover:

$$\dim \Im(x) = n - 1, \ \forall x \in \mathrm{R}^n$$

By Frobenius theorem, R^n is a disjoint union of maximal integral submanifolds of $\Im$. The idea of Bismut is to write the flow associated to x_t as a composition of a diffusion flow living in an integral submanifold and a deterministic transverse flow.

More explicitly, let φ_t be the flow associated to X_0. Then:

Proposition. $x_t = \varphi_t(X_t)$ where,

$$dX_t = \varphi_t^{*-1} X_i(X_t) \circ dW^i_t$$

Denote by M_{x_o} the integral submanifold of $\Im$ which contains x_o. Then:

Proposition. $X_t(x_0) \in M_{x_0}$, $\forall t \in \mathrm{R}^+$ and the law of X_t has a density on C M_{x_0}.

The hypoellipticity of L is obtained by integration along the flow φ_t.

3.3.7 Bibliographical comments

The equation in Theorem 3.1.a is often called the forward equation, or the Fokker-Planck equation. It can be found in most standard textbooks on stochastic differential equations and diffusion processes. Theorem 3.1.b. is due to Stroock-Varadhan [A 15].

The first two lines of Theorem 3.2 is Hörmander's sum of squares theorem, whose probabilistic proof is essentially the subject of the rest of the section. The third line of Theorem 3.2 is borrowed from results in control theory, see e.g. A. Isidori: Nonlinear Control Systems, Lecture Notes in Control and Info. Sci. 72, Springer Verlag (1985) or C. Lobry: Bases mathématiques de la théorie des systèmes asservis non linéaires, mimeographed, Univ. de Bordeaux (1976).

Bismut's version of the Malliavin calculus applied to SDEs appears in [A1]. In his paper, Bismut makes a heavy use of the theory of flows. This may obscur the exposition for the reader who is not familiar with stochastic flows. Norris [A7] presents a very clear and complete account of Bismut's approach.

The proof of the existence of the density in the one-dimensional case can also be found in Zakai [A16].

The original approach of Malliavin [A5], [A6] has been developed by Stroock in [A11], [A12], [A13] and [A14]. Other expositions appear in Ikeda-Watanabe [A3], Kusuoka-Stroock [A4], Ocone [A8], Shigekawa [A9], [A10] and Watanabe [A15].

The result in 3.6.1 seems to be new. The regularity with respect to the starting point is studied in Stroock [A14] and the hypoellipticity of L in Stroock [A 14] and Bismut [A1].

3.4 AN APPLICATION TO NON LINEAR FILTERING: EXISTENCE AND REGULARITY OF THE CONDITIONAL DENSITY (following Bismut-Michel)

We consider in this paragraph, a "signal-observation" process (x, y) with values in $\mathrm{R}^n \times \mathrm{R}^p$, solution of a system:

$$dx_t = X_o(x_t)dt + X_i(x_t) \circ dw_t^i + Y_j(x_t) \circ dy_t^j$$
$$dy_t = h(x_t)dt + dv_t$$

where w and v are independant Wiener processes with values in R^m and R^p respectively, $X_0, X_1, \ldots, X_n$, $Y_1, \ldots, Y_p, h$ are in C_b^∞.

The problem is: does the filter associated to that system possess a smooth density, i.e. does there exist a smooth function $q(t, x)$ such that for every continuous function with compact support $\varphi : \mathrm{R}^n \to \mathrm{R}$:

$$\pi_t\varphi = E(\varphi(x_t)|\mathcal{Y}_t) = \int_{\mathrm{R}^n} \varphi(x)q(t,x)dx$$

where $\mathcal{Y}_t = \sigma(y_s, s \leq t)$

If $h = 0$ and $Y_1 = \ldots = Y_p = 0$, the conditional law is the law of x and the answer to the question is in §3. We shall in a certain sense, go back to that situation by using Girsanov theorem and the theory of stochastic flows. But first remark that, heuristically, if $h = 0$ the conditional law is the law of the diffusion associated to the operator $\frac{1}{2}\Sigma X_i^2 + X_0 + \Sigma Y_i \frac{dy^i}{dt}$ whose drift term has $p+1$ independant components, namely $X_0, Y_1, \ldots, Y_p$. This is why it is natural to introduce the ideal $\mathcal{J}$ generated in $\mathcal{A} = LA(X_0, X_1, \ldots, X_m, Y_1, \ldots, Y_p)$ by the Lie algebra $\mathcal{B} = LA(X_1, \ldots, X_m)$.

3.4.1 The Girsanov theorem

$Z_t = \exp\left(\int_0^t h(xs)dy_s - \frac{1}{2}\int_0^t |h(xs)|^2 ds\right)$ is an exponential martingale and it allows to define a new probability measure $\overset{\circ}{P}$ on $(\Omega, \mathcal{F})$ by:

$$\frac{d\overset{\circ}{P}}{dP}|F_t = Z_t^{-1}, \quad t \geq 0.$$

then:

a. by Girsanov's theorem, y and w are independant Wiener processes under $\overset{\circ}{P}$.

b. by Kallianpur-Striebel's formula.

$$\pi_t\varphi \equiv E(\varphi(x_t)|\mathcal{Y}_t) = \frac{\mathring{E}(\varphi(x_t)Z_t|\mathcal{Y}_t)}{\mathring{E}(Z_t|\mathcal{Y}_t)} \equiv \frac{\sigma_t\varphi}{\sigma_t 1}$$

We get a first result in the uncorrelated case.

Theorem. Assume $Y_1 = \ldots = Y_p = 0$ and $\mathcal{J}(x) = \mathrm{R}^n$, $\forall x \in \mathrm{R}^n$, then π_t has a density.

Proof: Define $\rho(t,x) = \mathring{E}(Z_t|x_t = x, Y_t)$, then, by the independance of x and y under $\mathring{P}$ we have:

$$\sigma_t\varphi = \mathring{E}(\varphi(x_t)\rho(t,x_t))$$

Under the assumption, the law of x_t, which is the same under P or $\mathring{P}$, has a density p and: $\sigma_t\varphi = \int \varphi(x)\rho(t,x)\,p(t,x)dx$. Therefore, π_t has the density:

$$q(t,x) = \frac{p(t,x)\rho(t,x)}{\int p(t,x)\rho(t,x)dx}.$$

3.4.2 The stochastic flows

In the case of correlated noises, more work is needed. In order to be able to fix the trajectory y in x_t, we write x_t as a composition of stochastic flows (cf §3.6.3).

Denote by ψ_t the stochastic flow associated to the following SDE.

$$du_t = Y_j(u_t) \circ dy_t^j$$

It is characterized by the following properties:

(i) $t \to \psi_t(y, u_0)$ is the essentially unique solution of the SDE

(ii) $(t, u_0) \to \psi_t(y, u_0)$ is a continuous map for every y, a.s.

(iii) $u_0 \to \psi_t(y, u_0)$ is a C^∞ diffeomorphism, $\forall (t,y)$, a.e.

Proposition. For $a.e.y, \bar{x}_t = \psi_t^{-1}(y, x_t)$ is the essentially unique solution of the SDE:

$$d\bar{x}_t = \psi_t^{*-1}X_i(\bar{x}_t) \circ dw_t^i + \psi_t^{*-1}X_0(\bar{x}_t)dt$$

where

$$\psi_t^{*-1}X(x) \equiv \left[\frac{\partial\psi_t}{\partial x}(x)\right]^{-1} X(\psi_t(x)).$$

Since ψ_t is a C^∞ diffeomorphism, to prove that the conditional law of x_t given y_t under $\mathring{P}$ has a smooth density, it suffices to prove that the law of $\bar{x}_t$ has a smooth density, for $a.e.y$. Of course, we cannot apply directly the results of §3 to the equation satisfied by $\bar{x}_t$ because it is inhomogeneous and with unbounded coefficients. But, by first assuming that the vector fields $Y_1, \ldots, Y_p$ have compact support which insures the boundedness of ψ_t and its derivatives and then passing to the limit, one can show that everything works in a similar way. If τ_t (resp. $\bar{\tau}_t$) is the stochastic flow associated to x_t (resp. $\bar{x}_t$), one gets:

$$D_s\bar{x}_t = \bar{\tau}_t^*\bar{\tau}_s^{*-1}(\psi_s^{*-1}X_i)(x_0)$$

Since $\tau_t = \psi_t \circ \bar{\tau}_t$, we have:

$$D_s\bar{x}_t = \bar{\tau}_t^*\tau_s^{*-1}X_i(x_0).$$

Since $\bar{\tau}_t^*$ is an invertible linear map, the invertibility of the Malliavin matrix is equivalent to that of:

$$C_t = \sum_{k=1}^m \int_0^t \left(\tau_s^{*-1}X_k(x_0)\right)\left(\tau_s^{*-1}X_k(x_0)\right)^t ds$$

Theorem. Assume $\mathcal{J}(x_0) = \mathrm{R}^n$. then C_t is a.s. invertible and π_t has a smooth density.

Proof: It is similar to that of Theorem 3.4.2. We just write the stochastic differential of $\tau_t^{*-1}X_i(x_0)$ which shows how the ideal $\mathcal{J}$ enters.

$$d\tau_t^{*-1}X_i(x_0) = \tau_t^{*-1}[X_0, X_i](x_0)dt + \tau_t^{*-1}[X_j, X_i](x_0) \circ dw_t^j + \tau_t^{*-1}[Y_j, X_i](x_0) \circ dy_t^j$$

Remarks.

a. We have presented here a simplified version of the Bismut-Michel approach, which yields only the existence of a density for Π_t. Indeed, in order to obtain the smoothness of that density along the above lines, we would have to prove both the smoothness of the conditional density under $\mathring{P}$, and the smoothness o $x \to p(t,x)$. Bismut-Michel obtain the smoothness of the density of Π_t via a different and more direct approach.

b. The same kind of result can be obtained under more general assumptions.

- X_0, X_i, Y_j, h can depend on y_t.
- the coefficient of dv_t can be a function of y_t : $dy_t = h(x_t, y_t)dt + l(y_t) \circ dv_t$

3.4.3 Bibliographical comments

Our reference here has been Bismut-Michel [B1], which is an improvement over the original work of Michel [B7]. The same result can be obtained by an adaptation of the PDE hypoellipticity argument. For a summary of the two approaches and further references, we refer to Michel [B8]. Let us finally mention the works of Ferreyra [B4] and Kusuoka-Stroock [B6] who present other versions of the proof of existence and smoothness of the conditional law in nonlinear filtering, using the Malliavin calculus.

3.5 ANOTHER APPLICATION TO NON LINEAR FILTERING: MALLIAVIN CALCULUS APPLIED TO ZAKAI EQUATION AND NON EXISTENCE OF FINITE DIMENSIONAL FILTERS (following D. Ocone)

We consider a filtering problem as in §4 and assume that the signal noise and the observation noise are independent i.e. that the vector fields $Y_1 \ldots, Y_p$ are identically zero. So (x, y) is a solution of:

$$dx_t = X_0(x_t)dt + X_i(x_t) \circ dw_t^i$$
$$dy_t = h(x_t)dt + dv_t$$

and $X_0, X_1, \ldots, X_m, h$ are C^∞ and bounded as well as all their derivatives.

From §4, we know that, under Hörmander's condition $(\mathcal{J}(x_0) = \mathrm{R}^n)$, the unnormalized filter has a density p_t. Assume: $(X_1, \ldots, X_m)$ is a uniformly elliptic system of vector fields, i.e., there exists $d > 0$ such that

$$\forall v \in \mathrm{R}^m, \quad \sum_{i=1}^{m} (\mathrm{X_i}, \mathrm{v})^2 \geq \mathrm{d}\|\mathrm{v}\|^2$$

then, Krylov Rozovskii and Pardoux observed that p_t is in $L^2(\Omega \times [0,T], H^{1,2}(\mathrm{R}^n)) \cap \mathrm{L}^2(\Omega, \mathrm{C}([0,\mathrm{T}], \mathrm{L}^2(\mathrm{R}^n)))$ and that it is the unique solution in this space of Zakai's equation

$$dp_t = L_0^* p_t dt + L_i^* p_t \circ dy_t^i$$

where

$$L_0 = \frac{1}{2}\left(\sum_{i=1}^{m} X_i^2\right) + X_0 - \frac{1}{2}h^2$$
$$L_i = h_i$$

From now on, we consider p_t as the unique process with values in $L^2(\mathrm{R}^n)$ solution of Zakai's equation.

The idea is to use Malliavin's calculus to determine whether this process "fills up" $L^2(\mathrm{R}^n)$ or not. Of course, we cannot speak of a density for the law of p_t in $L^2(\mathrm{R}^n)$ because there is no Lebesgue measure in $L^2(\mathrm{R}^n)$ but Ocone looks at all the finite dimensional projections of p_t in $L^2(\mathrm{R}^n)$ and finds a condition of Hörmarder's type under which every such projection has a density.
Denote by $\wedge$ the ideal of differential operators in R^n generated by Lie $(L_1^*, \ldots, L_p^*)$ in Lie $(L_0^*, L_1^*, \ldots, L_p^*)$ and, if $p \in L^2(\mathrm{R}^m)$, by $\wedge(p)$ the space $\{Dp, D \in \wedge\}$. Then, we have a result which is an infinite dimensional analog of that in §3.2.

3.5.1 Theorem.

Assume $\wedge p_0$ is dense in $L^2(\mathrm{R}^n)$. Then for any $t > 0$, the law of every finite dimensional projection of p_t has a density.

This result has an important consequence in terms of finite dimensional filters.

3.5.2 Definition.

For any $t > 0$, p_t is said to be finite dimensionally computable if there exists a finite dimensional random vector $Z(t, y)$ evolving in a finite dimensional vector space such that: $p_t(x; y) = F_t(x; Z(t, y))$ for some smooth function $F_t(\cdot)$.

3.5.3 Theorem.

If $\wedge p_0$ is dense in $L^2(\mathrm{R}^n)$, then for any t, p_t is not finite dimensionally computable.

Proof: Suppose p_t is finite dimensionally computable, then $F_t(\cdot\,; Z(t, y))$ stays in a finite dimensional sub manifold of $L^2(\mathrm{R}^n)$ when y evolves in $C([0, t], \mathrm{R}^p)$. So if S is a sub-space of $L^2(\mathrm{R}^n)$ of dimension larger than that of the state space of Z, then the projection of p_t on S cannot have a density, which contradicts Theorem 5.1.

3.5.4 Sketch of the proof of Theorem 3.5.1

Denote by S a finite dimensional subspace of $L^2(\mathrm{R}^n)$ and by π_S the orthogonal projection on S.

3.5.4.1 Lemma

$\forall t \in \mathrm{R}_+, p_t$ is in the space $H^{2,1}$ of functionals from Ω into $L^2(\mathrm{R}^n)$ having a gradient.

This gradient is a Hilbert-Schmidt operator on $L^2(\mathrm{R}^n)$ whose kernel $D_s p_t$ satisfies;

$$D_s p_t = L_1^* p_s + \int_0^t L_0^* D_s p_r dr + \int_0^t L_1^* D_s p_r \circ dy_r$$

(we assume here for simplicity that the dimension of y is one).

Denote by $\Phi(t,s)$ the H.S. operator on $L^2(\mathrm{R}^n)$ solution of the linear equation: $\forall \varphi \in L^2(\mathrm{R}^n)$:

$$\Phi(t,s)\varphi = \varphi + \int_s^t L_0^* \Phi(r,s)\varphi dr + \int_s^t L_1^* \Phi(r,s)\varphi \circ dy_r$$

Then $D_s p_t = \Phi(t,s) L_1^* P_s$.

Let A be the Malliavin covariance matrix of p_t:

$$A_t = \int_0^t D_s p_t \otimes D_s p_t ds.$$

3.5.4.2 Lemma

If A_t is positive definite a.s., then the Malliavin covariance matrix associated to $\pi_S p_t$ is positive definite a.s., for every finite dimensional subspace S of $L^2(\mathrm{R}^n)$

Proof: Remark that the projection π_S and the gradient operator commute, ie.

$$D_S \pi_S p_t = \pi_S D_S p_t$$

So, if $A_{S,t}$ is the Malliavin covariance matrix associated to $\pi_S p_t$, one gets:

$$A_{S,t} = \pi_S A_t \pi_S^*.$$

and the result follows.

3.5.4.3 Lemma

If $\wedge p_0$ is dense in $L^2(\mathrm{R}^n)$, then A_t is positive definite a.s.

Proof: The difference with the finite dimensional case is that we cannot write $\Phi(r,s) = \Phi(r)\Phi(s)^{-1}$ because $\Phi(s)^{-1}$ would be an unbounded operator whose

domain is not known. So, instead of working on $\Phi(s)^{-1}\varphi$ we work directly on $\langle A_t, \varphi, \varphi\rangle$ that we write as:

$$\int_0^t \sum_i < L_i^* p_s, \Phi^*(t,s)\varphi >^2 ds$$

Let $v(s,t,\varphi) = \Phi^*(t,s)\varphi$. Then, as a function of s, for (t,φ) fixed, v solves the following backward SPDE:

$$dv_s = -L_0 v_s ds - L_i v_s \circ dy_s^i$$

$$v_t = \varphi$$

and the starting point of the proof is the following equality

$$\langle L_i^* p_s, \Phi^*(t,s)\varphi\rangle = \langle L_i^* p_0, \Phi^*(t,0)\varphi\rangle + \int_0^s \langle [L_i^*, L_0^*] p_r, \Phi^*(t,r)\varphi\rangle dr$$
$$\int_0^s \langle [L_i^*, L_j^*] p_r, \Phi^*(t,r)\varphi\rangle \circ dy_r^j$$

As in the classical case, this implies that, if $\langle A_t\varphi, \varphi\rangle = 0$, then $\Phi^*(t,0)\varphi$ is orthogonal to $\wedge p_0$. So: $\Phi^*(t,0)\varphi = v(0) = 0$. This implies $\varphi = 0$, from a backward uniqueness Theorem for parabolic PDEs.

Of course, what we need is stronger because we want the exceptional set in y to be independent of φ. This is done by using a pathwise interpretation and a technical Lemma.

Another difficulty we have overlooked is that v_t is not adapted: this can be solved by using either a pathwise interpretation or the extended stochastic calculus.

3.5.5 Bibliographical comments

The contents of this section is based on the work of Ocone [B 9]. Ocone has applied his results to the cubic sensor problem in [B 10].

Let us now explain in what sense the results of Ocone are stronger than earlier results about nonexistence of finite dimensional sufficient statistics, due in particular to Hazewinkel-Marcus [B5], Sussmann [B11] and Chaleyat Maurel-Michel [B3]. What the results of these authors preclude is the existence of a smooth finite dimensional manifold M, smooth vector fields over $M : Z_0, Z_1, \ldots, Z_p$, and a smooth mapping $F : \mathrm{R}_+ \times \mathrm{M} \to \mathrm{L}^2(\mathrm{R}^n)$ s.t.:

$$p_t = F(t, V_t), \qquad t \geq 0; \qquad \text{where :}$$
$$dV_t = Z_0(V_t)dt + \sum_{i=1}^{p} Z_i(V_t) \circ dy_t^i$$

On the other hand, Ocone shows that under appropriate conditions, for each fixed $t > 0$, there does not exist a finite dimensional manifold M, an M-valued random variable V_t, and a smooth mapping $F(t) : M \to L^2(\mathrm{R}^n)$, such that for that t:

$$p_t = F(t, V_t).$$

We finally note that the stochastic calculus of variation has also been used by Chaleyat-Maurel [B2] in order to study the smoothness of the mapping $\{y(s); 0 \leq s \leq t\} \to p_t$.

3.6 EXTENDED STOCHASTIC INTEGRALS AND STOCHASTIC CALCULUS (Following Skorohod; Gaveau-Trauber; Nualart-Zakai; Nualart-Pardoux).

For simplicity, all processes in this section will be one dimensional, $(\Omega, \mathcal{F}, P)$ is the canonical space $C([0,1], \mathrm{R})$ equipped with Wiener measure and $w_t(\omega) = w(t)$.

3.6.1 The Skorohod integral

In this paragraph we develop the idea of Remark 3.2.3.

We saw in §2, that, if $\Phi \in H^{2,1}$ and $u \in L^2([0,1] \times \Omega, \mathrm{R})$ and is adapted, the following integration by parts formula holds:

$$E(\Phi \int_0^1 u_s dw_s) = E\left(\int_0^1 D_t\Phi \; u_t dt\right).$$

This can be rewritten as:

$$E(\Phi \delta u) = E((D\Phi, u))$$

and $u \longmapsto \delta u = \int_0^1 u_s dw_s$ appears as the adjoint of the gradient operator. Actually, we can extend the domain of δ to every process u such that the map:

$$\Phi \longrightarrow E((D\Phi, u))$$

is continuous on $L^2(\Omega)$.

It can be shown that this is the domain of definition of the Skorohod integral of u. Here, we shall take δ as a definition for the Skorohod integral.

3.6.1.1 Definition

(i) An element $u \in L^2((0,1) \times \Omega, \mathrm{R})$ is said to be Skorohod integrable if:

$$\exists C > 0, \forall \Phi \in H^{2,1}, |E((D\Phi, u))| \leq C \|\Phi\|_2$$

(ii) Let $\delta : L^2((0,1) \times \Omega, \mathrm{R}) \to \mathrm{L}^2(\Omega)$ be the unbounded operator with domain:

$$\text{Dom } \delta = \{\text{Skorohod integrable processes}\}$$

and defined by: if $u \in$ Dom δ, $\forall \Phi \in H^{2,1}, E(\Phi\, \delta\, u) = E((D\, \Phi, u))$

3.6.1.2 Basic properties of the Skorohod integral

Notation:

$$F_t = \sigma(w_s, s \leq t)$$
$$F^t = \sigma(w_s - w_t,\ t \leq s \leq 1)$$

Theorem.

a. If $u \in L^2((0,1) \times \Omega, \mathrm{R})$ and is F_t-adapted, then $u \in$ Dom δ and $\delta u = \int_0^1 u_s dw_s$ is the Itô integral.

b. If $u \in L^2((0,1) \times \Omega, H^{2,1})$, then $u \in$ Dom δ and $\delta u = \int_0^1 u_s dw_s$ is the backward Itô integral.

c. If $u \in L^{2,1} = L^2((0,1), H^{2,1})$, then $u \in$ Dom δ.

d. If $u \in$ Dom δ and $\Phi \in H^{2,1}$, then:

$$\delta(\Phi\, u) = \Phi\, \delta u - \int_0^1 D_t\Phi \; u_t dt$$

in the sense that $\Phi u \in$ Dom δ if and only if the right hand side is square integrable and then the equality holds.

Proof: (a) is a consequence of Theorem 3.2.3, and (b) is proved in the same way. (d) follows from the definition of δ. Let us indicate a proof of (c), let $u \in L^{2,1}$. Define $t_k^n = k2^{-n}$,

$$u_{n,k} = \frac{1}{t_{k+1}^n - t_k^n} \int_{t_k^n}^{t_{k+1}^n} u_s ds$$

$$u_n(t) = \sum_{k=0}^{2^n-1} u_{n,k} 1_{[t_k^n, t_{k+1}^n L]}(t)$$

Consider the random variable:

$$\xi_n = \sum_{k=0}^{2^n-1} u_{n,k}(w_{t_{k+1}^n} - w_{t_k^n}) - \\ - \frac{1}{t_{k+1}^n - t_k^n} \int_{t_k^n}^{t_{k+1}^n} \int_{t_k^n}^{t_{k+1}^n} D_r u_s ds\, dr$$

It can be shown that $\xi_n \in L^2(\Omega)$; it then follows from (d) that $u_n \in \mathrm{Dom}\,\delta$ and $\delta u_n = \xi_n$. One can moreover show that $E(\xi_n \times \xi_m)$ converges as $n, m \to \infty$. Then $\{\xi_n, n \in rmN\}$ is Cauchy in $L^2(\Omega)$. But $u_n \to u$ in $L^2(\Omega \times (0,1))$, and δ is a closed operator (as the adjoint of the operator D which has a dense domain). Consequently, $u \in \mathrm{Dom}\,\delta$, and

$$\delta u = L^2 - \lim_{n \to +\infty} \delta u_n$$

3.6.1.3 Further properties

(i) In cases (a, b, c):

$$\delta u = L^2 - \lim_{n \to +\infty} \sum_{k=0}^{2^n-1} \bar{u}_{n,k}(w_{t_{k+1}^n} - w_{t_k^n})$$

where

$$\bar{u}_{n,k} = E\left(u_{n,k} | F_{t_k^n} \vee F^{t_{k+1}^n}\right)$$

(ii) In cases (a, b, c):

$$E(\delta\, u) = 0$$

and in case c:

$$E(\delta\, u)^2 = E\left(\int_0^1 u_t^2 dt\right) + E\left(\int_0^1 \int_0^1 D_s u_t D_t u_s ds\, dt\right)$$

this last term vanishes in case a because if $s > t$, $D_s u_t = 0$, and in case b since if $s < t$, $D_s u_t = 0$.

3.6.2 Skorohod's integral as a process

Define: $\int_0^t u_s dw_s = \delta(u\, 1_{[0,t]})$ for $t \in (0,1)$.

3.6.2.1 Theorem.

The process $\{\int_0^t u_s dw_s,\ t \in [0,1]\}$ has a continuous modification if $u \in L^{2,1}$ and one of the following properties is satisfied:

- $\exists p > 1, \sup_{t \in [0,1]} E\left(\left(\int_0^1 |D_s u_t|^2 ds\right)^p\right) < +\infty$
- $\exists p > 2, E\left(\int_0^1 \left(\int_0^1 |D_s u_t|^2 ds\right)^p dt\right) < +\infty$.

3.6.2.2 Theorem.

Under the assumptions of Theorem 3.6.2.1,

$$\sum_{k=0}^{2^n-1} \left(\int_{t_k^n}^{t_{k+1}^n} u_s dw_s\right)^2 \longrightarrow \int_0^1 u_s^2\, ds$$

in probability, as $n \to \infty$.

3.6.3 The generalized Itô formula

Notation:

$H^{p,1} = \{\Phi \in L^p(\Omega), D\Phi \in L^p(\Omega \times (0,1))\}$

$L^{p,1} = L^p((0,1), H^{p,1})$

$L^{p,2} = \{u \in L^{p,1}, D_t u \in H^{p,1}$ a.e., $E\left(\int_0^1 \int_0^1 |D_s D_t u|^p ds\, dt\right) < +\infty\}$

Theorem. Consider:

$$X_t = X_0 + \int_0^t A_s\, ds + \int_0^t B_s\, dw_s$$

where: $X_0 \in H^{4,1}$, $A \in L^{4,1}$, $B \in L^{p,2}$ for some $p > 4$.
and let $F \in C^2(\mathrm{R})$. Then:

$$F(X_t) = F(X_0) + \int_0^t F'(X_s)\, A_s\, ds + \int_0^t F'(X_s) B_s\, dw_s \\ + \frac{1}{2}\int_0^t F''(X_s)\, (\nabla X)_s\, B_s\, ds.$$

where:

$$(\nabla X)_t = \lim_{s \to t, s > t} D_t X_s + \lim_{s \to t, s < t} D_t X_s = D_t^+ X_t + D_t^- X_t$$

$$= B_t + 2\left(D_t X_0 + \int_0^t D_t X_r \, dr + \int_0^t D_t B_r \, dw_r \right)$$

Remark. In the adapted case, $(\nabla X)_t = B_t$ and this formula reduces to the usual Itô formula.

Proof: The three ingredients of the proof are the continuity of $\{X_t\}$ which follows from Theorem 6.2.1. Property 6.1.2. d, and Theorem 6.2.2. Write t_k for t_k^n.

$$F(X_t) - F(X_0) = \lim_n \sum_{k=0}^{[t2^n]-1} [F(X_{t_{k+1}}) - F(X_{t_k})]$$

$$= \lim_n \sum_k F'(X_{t_k}) \int_{t_k}^{t_{k+1}} A_s \, ds + \lim_n \sum_k F'(X_{t_k}) \int_{t_k}^{t_{k+1}} B_s \, dW_s +$$

$$+ \frac{1}{2} \lim_n \sum_k F''(X_k) \left(\int_{t_k}^{t_{k+1}} B_s \, dW_s \right)^2$$

clearly,

$$\sum_k F'(X_{t_k}) \int_{t_k}^{t_{k+1}} A_s \, ds \to \int_0^t F'(X_s) A_s \, ds$$

The convergence:

$$\frac{1}{2} \sum_k F''(X_k) \left(\int_{t_k}^{t_{k+1}} B_s \, dW_s \right)^2 \to \frac{1}{2} \int_0^t F''(X_s) B_s^2 ds$$

is an easy consequence of Theorem 6.2.2.

$$\sum_k F'(X_{t_k}) \int_{t_k}^{t_{k+1}} B_s dW_s = \sum_k \int_{t_k}^{t_{k+1}} F'(X_{t_k}) B_s dW_s +$$

$$+ \sum_k \int_{t_k}^{t_{k+1}} F''(X_{t_k}) D_s X_{t_k} B_s ds$$

$$\to \int_0^t F'(X_s) B_s dW_s + \frac{1}{2} \int_0^t F''(X_s) [(\nabla X)_s - B_s] B_s \, ds$$

3.6.4 The extended Stratonovitch integral

Definition. We say that $\{u_t, t \in (0,1)\}$ is Stratonovitch integrable if:

$$\sum_{k=0}^{2^n-1} u_{n,k} \left(w_{t_{k+1}^n \wedge t} - w_{t_k^n \wedge t} \right)$$

converges in probability, $\forall t \in [0,1]$. In that case the limit, denoted $\int_0^t u_s o dw_s$, is called the Stratonovitch integral of u.

Notation:

$L_c^{2,1} = \{u \in L^{2,1} /$ the set of functions $\{s \to D_t u_s;\ s \in [0,1] - \{t\}\}_{t \in [0,1]}$ is equi-continuous, and $\sup_{s,t} E(|D_t u_s|^2) < +\infty\}$.

If $u \in L_c^{2,1}$, one can define $D_t^+ u_t, D_t^- u_t$ and so $(\nabla u)_t$.

Theorem. If $u \in L_c^{2,1}$, u is Stratonovitch integrable and:

$$\int_0^t u_s \circ dw_s = \int_0^t u_s dw_s + \frac{1}{2} \int_0^t (\nabla u)_s \, ds.$$

Remark. If u_t is a F_t-semimartingale:

$$\int_0^1 u_t \circ dw_t = \int_0^1 u_t dw_t + \frac{1}{2} \langle u, w \rangle_1$$

If u_t is a F^t "backward semi-martingale"

$$\int_0^1 u_t \circ dw_t = \int_0^1 u_t dw_t - \frac{1}{2} \langle u, w \rangle_1.$$

More generally, if $u \in L_c^{2,1}$, $\langle u, w \rangle_1$ exists and is given by:

$$\langle u, w \rangle_1 = \int_0^1 (D_t^+ u_t - D_t^- u_t) dt$$

In the case u F_t adapted, D_t^- $u_t = 0$. So:

$$\frac{1}{2} \langle u, w \rangle_1 = \frac{1}{2} \int_0^1 D_t^+ \, u_t \, dt$$

$$= \frac{1}{2} \int_0^1 (\nabla u)_t \, dt$$

In the case u F^t adapted,

$D_t^+ \mu_t = 0$. So:

$$\frac{1}{2} \langle u, w \rangle_1 = \frac{1}{2} \int_0^1 D_t^- u_t \, dt$$

$$= \frac{1}{2} \int_0^t (\nabla u)_t \, dt$$

3.6.5 The generalized Stratonovich formula

Let the assumptions of Itô's formula be in force and assume in addition:

$$B \in L_c^{2,1} \quad \text{and} \quad \nabla B \in L^{4,1}$$

then, if: $X_t = X_0 + \int_0^t A_s ds + \int_0^t B_s \circ dw_s$,

$$F(X_t) = F(X_0) + \int_0^t F'(X_s) A_s ds + \int_0^t F'(X_s) B_s \circ dw_s$$

3.6.6 Bibliographical comments

We have mainly followed the presentation and results of Nualart-Pardoux [C 4]. Earlier results on the Skorohod integral include the work of Skorohod [C 14], Gaveau-Trauber [C 1], Krée [C 2] and Nualart-Zakai [C 5]. Note that our notion of "Skorohod integrable processes" is slightly more general than that in the original paper by Skorohod. Various versions of the generalized Itô formula have been given by Sevljakov [C 12], Sekiguchi-Shiota [C 11], Ustunel [C 15] and Nualart-Pardoux [C 4]. Ogawa [C 8] has constructed a generalized stochastic integral which essentially coincides with what we call the "extended Stratonovich integral." For an overwiew and comparison of all the existing results, we refer the reader to the expository paper of Nualart [C 3]. For applications to stochastic differential equations, we refer to Ogawa [C 9], Shiota [C 13], Ocone-Pardoux [C 6], [C 7] and Pardoux-Protter [C 10].

REFERENCES

A. *Stochastic calculus of variations and its application to solutions of stochastic differential equations.*

[A1] BISMUT, J. M., (1981). *Martingales, The Malliavin calculus and hypoellipticity under general Hörmander's conditions.* Z. Wahrs, V. Geb. 56, 469-505.

[A2] BISMUT, J. M., (1982). *An introduction to the stochastic calculus of variations,* in Stochastic Differential Systems, M. Kohlmann & N. Christopeit eds, p.33-72, Lecture Notes in Control and Inf. Sci. 43, Springer-Verlag.

[A3] IKEDA, N. and WATANABE, S., (1981). *Stochastic differential equations and diffusion processes.* North Holland, Kodansha, Amsterdam/Tokyo.

[A4] KUSUOKA, S., and STROOCK, D. W., (1982). *Applications of the Malliavin calculus.* Part I. Taniguchi Symposium, Takata. K. Itô ed. Kinokuniya-Tokyo (1984) 277-306. Part II, J. Fac. Sci. Tokyo, Sec IA, 32 (1985) 1-76.

[A5] MALLIAVIN, P., (1976). *Stochastic calculus of variation and hypoelliptic operators* in Proc. Int. Symp. on SDE Kyoto K. Itô ed. p. 195-263, Kinokuniya-Tokyo (1978).

[A6] MALLIAVIN, P, (1978). *C^k-hypoellipticity with degeneracy* in Stochastic Analysi, A. Friedman and M. Pinsky eds. p. 199-214 and p. 327-340. Academic Press.

[A7] NORRIS, J, (1986). *Simplified Malliavin Calculus,* in Séminaire de Probabilités XX, J. Azéma, M. Yor Eds, Lecture Notes in Mathematics 1204, Springer Verlag, 101-130.

[A8] OCONE, D. (1988). *A guide to the stochastic calculus of variations,* Proc. Silivri Conf, Lecture Notes in Math. Springer-Verlag .

[A9] SHIGEKAWA, I., (1978). *Absolute continuity of probability laws of Wiener functionals.* Proc. Jap. Acad. 54, A, p.230-233.

[A10] SHIGEKAWA, I., (1980). *Derivatives of Wiener functionals and absolute continuity of induced measures.* J. Math. Kyoto Univ. 20. 263-289.

[A11] STROOCK, D. W., (1981). *The Malliavin calculus and its applications to second order parabolic differential operators.* Parts I and II. Math. Systems Theory, 14, 25-65 and 141-171.

[A12] STROOCK, D. W., (1981). *The Malliavin calculus. A functional analytic approach.* J. Funct. Analysis 44, 212-257.

[A13] STROOCK, D. W., (1983). *Some applications of stochastic calculus to partial differential equations.* Ecole d'été de Probabilités de St Flour XI. Lecture Notes in Mathematics 976, Springer-Verlag.

[A14] STROOCK, D. W., (1986). *Malliavin calculus and applications to PDEs.* IMA Preprint 278, IMA Minneapolis, MN.

[A15] STROOCK, D. W. and VARADHAN, S.R.S., *On the support of diffusion processes with applications to the maximum principle.* Proceedings of sixth Berkeley Symposium on Mathematical Statistics and Probability, Vol. III, 333-360.

[A16] WATANABE, S. *Lectures on stochastic differential equations and Malliavin calculus.* Tata Institute of Fundamental Research, Springer Verlag (1984).

[A17] ZAKAI, M. (1985). *The Malliavin calculus.* Acta Applicandae Mat. 3, 175-207.

B. Applications of the stochastic calculus of variations to nonlinear filtering.

[B1] BISMUT, J. M. and MICHEL, D., (1981). *Diffusions conditionnelles.* J. Funct. Anal. Part I: 44, 174-211, Part II: 45 (1982) 274-282.

[B2] CHALEYAT-MAUREL, M., (1986). *Robustesse en théorie du filtrage non linéaire et calcul des variations stochastique.* J. Funct-Anal.

[B3] CHALEYAT-MAUREL, M., - MICHEL, D., (1984) *Des résultats de non existence de filtre de dimension finie.* Stochastics 13.

[B4] FERREYRA, G., *Smoothness of the unnormalized conditional measure in stochastic nonlinear filtering.* Preprint Math. Dep. LSU Baton Rouge, LA, USA.

[B5] HAZEWINKEL, M. and MARCUS, S.,(1982). *On Lie Algebras and finite dimensional filtering.* Stochastics 7.

[B6] KUSUOKA, S. and STROOCK, D. W., (1984). *The partial Malliavin calculus and its application to nonlinear filtering.* Stochastics, 12, 83-142.

[B7] MICHEL, D., (1981). *Régularité des lois conditionelles en théorie du filtrage non linéaire et calcul des variations stochastique.* J. Funct. Anal. 41, 8-36.

[B8] MICHEL, D., (1984). *Conditional laws and Hörmander condition.* in Stochastic Analysis, K. Itô ed., North Holland, 387-408.

[B9] OCONE, D., *Stochastic calculus of variations for stochastic partial differential equations.* J. Funct, Anal., to appear.

[B10] OCONE, D., (1988). *Probability densities for conditional statistics in the cubic sensor problem.* Math. of Systems, Signals and Control.

[B11] SUSSMANN, H.,(1981). *Rigorous results on the cubic sensor problem.* in Stochastic Systems, Hazewinkel M. and Willems S. eds, Reidel.

C. Skorohod's integral and non adapted stochastic calculus

[C1] GAVEAU, B. and TRAUBER, P., (1982) *"L'intégrale Stochastique comme opérateur de divergence dans l'espace fonctionnel.* J. Funct. Anal. 46, 230-238.

[C2] KREE, M.,(1977). *Propriétés de trace en dimension infinie d'espaces de type Sobolev.* Bull. Soc. Math. France 105, 141-163.

[C3] NUALART, D., *Non-causal stochastic integrals and calculus.* Proc. Silivri Conf, Lecture Notes in Mathematics, Springer-Verlag (to appear).

[C4] NUALART, D. and PARDOUX, E., *Stochastic calculus with anticipating integrands.* Prob. Theory and Rel. Fields, to appear.

[C5] NUALART, D. and ZAKAÏ, M., (1986). *Generalized stochastic integrals and the Malliavin calculus.* Prob. Theory and Rel. Fields 73, 255-280.

[C6] OCONE, D. and PARDOUX, E. *A generalized Itô-Ventzell formula. Application to a class of anticipating stochastic differential equations.* Preprint.

[C7] OCONE, D. and PARDOUX, E. *Bilinear stochastic differential equations with boundary conditions.* Preprint.

[C8] OGAWA S., (1984). *Quelques propriétés de l'intégrale stochastique de type non-causal.* Japan J. of Applied Math. 1, 405-416.

[C9] OGAWA, S., (1984). *Sur la question d'existence de solutions d'une équation différentielle stochastique de type non-causal.* J. Math. Kyoto Univ. 24, 699-704.

[C10] PARDOUX, E. and PROTTER, P., *Stochastic Volterra equation with anticipating coefficients.* Preprint.

[C11] SEKIGUCHI, T. and SHIOTA, Y.(1985). *L^2 theory of non-causal stochastic integrals.* Math. Rep. Toyama Univ. 8, 119-195.

[C12] SEVLJAKOV, A., (1981). *The Itô formula for the extended stochastic integral.* Theory Prob. and Math. Stat. 22, 163-174.

[C13] SHIOTA, Y., (1986). *A linear stochastic integral equation containing the extended Itô integral.* Math. Rep. Toyama Univ. 9, 43-65.

[C14] SKOROHOD, A. V., (1975). *On a generalization of a stochastic integral.* Theory Prob. and Applic. 20, 219-233.

[C15] USTUNEL, A. S., (1986). *La formule de changement de variable pour l'intégrale anticipante de Skorohod.* CRAS.