AN INTRODUCTION TO THE STOCHASTIC CALCULUS OF VARIATIONS

BY

Jean-Michel BISMUT

Université de Paris-Sud

Département de Mathématique

Bâtiment 425

91405 ORSAY CEDEX

+-+-+-+-+

To be presented at the Conference on Stochastic Systems of Bad-Honnef (June 1982).

The purpose of this paper is to give a general introduction to the methods of the stochastic calculus of variations, and to some of its applications. The subject being in full development, we have not tried to cover the whole thing, but we have essentially focused on the aspects of the calculus which may be of special relevance to control and filtering.

The paper is organised around four essential topics.

In section 1, the theory of stochastic flows is reviewed. It has been known for a long time (see Gihman - Skorohod [19]) that the solutions of a stochastic differential equation depend smoothly on the initial conditions in L_2-sense. Blagoveschenskii and Freidlin [12] announced the stronger result that the solutions of such a stochastic differential equation could be modified so as to obtain an a.s. smoothness with respect to the initial conditions. This result was rediscovered by Malliavin [39] who gave a new impetus to the field. The whole subject has been developped in Ventzell [58], Rozovskii [48], Baxendale [1], Elworthy [17], Bismut [5] - [6], Ikeda - Watanabe [26], Kunita [31] - [32]. The basic idea is that in many aspects a stochastic differential equation behaves like deterministic differential equations so that to such an equation, a flow $\varphi_t(\omega,\cdot)$ of diffeomorphisms of R^d onto itself may be associated, so that $\varphi_t(\omega,\cdot)$ is a continuous stationary independent increment process with values in the group of diffeomorphisms of R^d. A number of standard operations may be performed on $\varphi_.(\omega,\cdot)$ like the lifting of $\varphi_t(\omega,\cdot)$ to tensors, the determination of the equation of $\varphi_t^{-1}(\omega,\cdot)$ etc ... The theory of flows extends to general stochastic differential equations the results of Doss [15], Sussmann [56]. It can be applied to the Malliavin calculus of variations and to study conditional diffusions (Bismut - Michel [10] - [11], Kunita [33], [34], [35]).

In section 2, the results of Haussmann [21] - [22] extending Clark [13] are presented. It has been known for a long time that a connection exists between the differentiability properties of functions and their behaviour as random variables. This is clearly illustrated by the well-known decomposition of functions of the Schwartz space S(R) into linear combinations of weighted Hermite polynomials [47].

Clark [13] and Haussmann [21] - [22] interpreted such relations in the setting of diffusions by relating the representation of certain random variables as martingales in terms of their differentials in the sense of Fréchet. In [7], Bismut exhibited the relation between these results and integration by parts on the Wiener space, and used them to recover the integration by parts obtained by Malliavin [39] using the Ornstein - Uhlenbeck process.

In Section 3, we give a brief exposition of the techniques of the Malliavin calculus. The papers of Malliavin [39] - [40] have been the starting point of the development of the whole field. The basic idea of Malliavin was that it was possible to use the construction of the solution of the stochastic differential equation by means of a Brownian motion as an efficient way of obtaining results on partial differential operators. The technique of Malliavin uses an auxiliary infinite dimensional stochastic process to establish a key integration by parts formula. His ideas were simplified by Shigekawa [49], Kusuoka [36] who used functional analysis techniques and the Ornstein Uhlenbeck operator. These aspects are fully developed and much extended in the papers of Stroock [51] - [52] - [53], to which the reader is refered. We have chosen instead to follow the approach taken by us in [7]. Also note that the key estimates were obtained by Malliavin [40], and Ikeda - Watanabe [26] and have been recently considerably improved by Kusuoka and Stroock [37].

In Section 4, the main results of Bismut - Michel [10] - [11] extending Michel [44] on conditional diffusions are developped. In this application, the theory of flows and the Malliavin calculus are used in combination to obtain results which extend the results of Pardoux [46], Krylov - Rozovskii [28] - [29], Davis [14], Eliott - Kohlmann [16]. These results are apparently difficult to obtain using more classical methods. For related developments, we refer to Kunita [33] - [34] - [35]. Another application of the Malliavin calculus is in Holley - Stroock [23], where the obtained results fall out of the reach of classical methods.

We have excluded of this survey some recent developments of the methods of the calculus of variations to jump processes [8] and boundary processes [9].

AN INTRODUCTION TO THE STOCHASTIC CALSULUS OF VARIATIONS

BY J.M.BISMUT

① Stochastic flows

Let Ω be the set $\mathcal{C}(R^+;R^m)$ whose standard element ω is a trajectory

$$\omega = (w_t^1,\dots,w_t^m)$$

Ω is endowed with its canonical filtration

$$F_t = \mathcal{B}(w_s | s \leq t)$$

which is eventually regularized on the right.

P is the Brownian measure on Ω, with $P[w_0 = 0] = 1$.

$X_0(x),\dots,X_m(x)$ are $m+1$ vector fields defined on R^d with values in R^d, which are bounded, C^∞, with bounded differentials.

For $x_0 \in R^d$, consider the stochastic differential equation written in Stratonovitch form

(1.1)
$$dx = X_0(x)dt + X_i(x).dw^i$$
$$x(0) = x_0$$

where dw^i is the Stratonovitch differential of w^i. (1.1) can be written in Ito's form

(1.1')
$$dx = \left(X_0(x) + \frac{1}{2}\frac{\partial X_i}{\partial x} X_i(x)\right)dt + X_i(x).\delta w^i$$
$$x(0) = x \ .$$

where δw^i is the Ito differential of w^i (the summation sign $\sum_{i=1}^{m}$ will be generally omitted).

a) Ito calculus and Stratonovitch calculus

In the whole paper, we will use as much as we can Stratonovitch integrals. If H and X_t are continuous semi-martingales, it is known [41] that the Stratonovitch integral $\int_0^t H_s \, d X_s$ can be expressed in terms of the Ito integral $\int_0^t H_s \, \delta X_s$ by the relation

$$\int_0^t H_s \, d X_s = \int_0^t H_s \, \delta X_s + \frac{1}{2} < H,X >_t \tag{1.2}$$

where $< H,X >_t$ denotes the quadratic variation of H and X.

In classical stochastic differential calculus, it is known [41] that the formal rules of the usual differential calculus are preserved in the Stratonovitch stochastic calculus.

As a consequence, we see that $X_0, X_1 \dots X_m$ in (1.1) are in fact vector-fields in the sense of differential geometry, i.e. if equation (1.1) is expressed in a new set of coordinates, the new fields $X'_0, \dots, X'_m$ are obtained through the standard rules of transformation of vector-fields. This is of course connected with the fact that the generator $\mathcal{L}$ of the diffusion (1.1) is given by

$$\mathcal{L} = X_0 + \frac{1}{2} \sum_1^m X_i^2 \tag{1.3}$$

where $X_0, X_1, \dots, X_m$ are considered as first-order differential operators. The invariance of $X_0, \dots, X_m$ is also connected with the results of Stroock-Varadhan [55] which show that the probability law of $x.$ is the limit of the probability laws of $x^n_.$, where $x^n_.$ is the solution of the differential equation (1.1) where w^i is replaced by its continuous linear dyadic approximation.

The reasons for using the Stratonovitch calculus are even more stringent here.

Since in fact we will deal with infinite dimensional diffusions, even if the Stratonovitch expressions are still easy to write, the corresponding Ito calculus formulas become hairy.

Ito's calculus will of course be the essential tool through which the analysis of the problems which we consider will be carried over, but once the formulas are proved we will come back to Stratonovitch calculus.

b) Construction of the flow

The idea is to associate to equation (1.1) a flow of diffeomorphisms $\varphi_t(\omega,.)$ of R^d onto R^d such that for any $x_0 \in R^d$, $\varphi_t(\omega,x_0)$ is exactly the solution of (1.1).

The differentiability of the solutions of (1.1) on the variable x_0 in the L_2-sense has been known for a long time (Gihman-Skorokhod [19]) and is in fact sufficient to establish "elementary" properties of the diffusion (1.1) like the fact that the associated semi-group T_t maps $C_b^\infty(R^d)$ in $C_b^\infty(R^d)$.

The a.s. differentiability of the solution of (1.1) in the variable x_0 had been announced in Blagoveschenskii and Freidlin [12]. It has been studied by Malliavin [39], Elworthy [17], Baxendale [1], Ikeda-Watanabe [26].

The complete properties of the flow $\varphi_t(\omega,\cdot)$, namely the fact that a.s. it is one-to-one and onto for every $t \in R^+$ were established in Bismut [5], [6] and Kunita [31] who showed these properties extend when w is replaced by a continuous semi-martingale.

Namely, we have :

Theorem 1.1 : There exists a mapping $\varphi_t(\omega,x)$ defined on $R^+ \times \Omega \times R^d$ with values in R^d such that

a) For any $(t,x) \in R^+ \times R^d$, $\omega \to \varphi_t(\omega,x)$ is measurable.

b) For any $\omega \in \Omega$, $(t,x) \to \varphi_t(\omega,x)$ is continuous.

c) For any $(\omega,t) \in \Omega \times R^+$, $x \to \varphi_t(\omega,x)$ is a C^∞ diffeomorphism of R^d onto R^d.

d) For any $\omega \in \Omega$, and any multi-index m, $(t,x) \to \dfrac{\partial^m \varphi_t}{\partial x^m}(\omega,x)$ is continuous.

e) For any $x_o \in R^d$, $t \to \varphi_t(\omega,x_o)$ is the (essentially) unique solution of (1.1).

Moreover properties a)-e) determine $\varphi(\omega,.)$ uniquely (in the sense of essential uniqueness).

Proof : We only give here a sketch of the proof. The existence of $\varphi_t(\omega,.)$ having properties a), b), d), e) is the "easy" part. In fact, if $x_.^{x_o}$ is the solution of (1.1), it is elementary to show that for any $T > 0$, $p \geq 2$, $x_o, y_o \in R^d$, $s,t \leq T$

(1.4)
$$E\left|x_t^{x_o} - x_s^{y_o}\right|^{2p} \leq C_T\left[\left|x_o - y_o\right|^{2p} + \left|t-s\right|^p\right]$$

Taking $p > d+1$ in (1.4) Kolmogorov's lemma implies the existence of an a.s. continuous mapping $(t,x_o) \to x_t^{x_o}(\omega)$. Using the differentiability of $x_.^{x_o}$ on x_o in the L_2-sense, it is not hard to prove the differentiability of $x_t^{x_o}$ on x_o in the sense of distributions. Using the classical form of the L_2 differential of x^{x_o} on x_o given by the stochastic differential equation

(1.5)
$$dZ^{x_o} = \frac{\partial X_o}{\partial x}(x_t^{x_o})\, Z^{x_o}\, dt + \frac{\partial X_i}{\partial x}(x_t^{x_o})\, Z^{x_o} . dw^i$$
$$Z^{x_o}(0) = I$$

it is not hard to prove (1.4) for Z^{x_o} and to deduce the a.s. differentiability of order 1 on x_o for $x_.^{x_o}$. The same rules give the C^∞ differentiability. Set $x_t^{x_o} = \varphi_t(\omega,x_o)$. Equation (1.5) permits the easy proof that a.s., for any $(t,x_o) \in R^+ \times R^n$, $\dfrac{\partial \varphi_t}{\partial x}(\omega,x_o)$ is invertible. The technical difficulty is now to prove the a.s. injectivity and the a.s. onto property of $\varphi_t(\omega,.)$. Note that if R^d is replaced by a compact connected manifold, since φ_o = identity, a homotopy argument gives precisely the required result (using the invertibility of $\dfrac{\partial \varphi_t}{\partial x}(\omega,x)$). In the case of R^d, the approximation of the flow $\varphi(\omega,.)$ by the flows $\varphi^n(\omega,.)$ of differential equations and a time reversal argument of Malliavin [39] were used in Bismut [5]-[6] to obtain Property c). In Kunita [31] a priori inequalities on (1.1) give this property. □

For $s \in R^+$, let θ_s be the standard translation operator in Ω, i.e. if $\omega = (w_t)$, $\theta_s\omega = (w_{s+t})$. It is then easy to prove that for any stopping time S, on $(S < +\infty)$, a.s., for any $t \geq 0$.

$$(1.6) \qquad \varphi_{S+t}(\omega,.) = \varphi_t(\theta_S\omega) \circ \varphi_S(\omega,.) \ .$$

Let G be the group of C^∞ diffeomorphisms of R^d onto R^d, endowed with the C_K^∞ topology (which is the topology of uniform convergence of a function and its derivatives on compact sets). Theorem 1.1 and (1.6) say precisely that $\varphi_t(\omega,.)$ is a continuous independent increment process with values in G. The lifting of the diffusion (1.1) to an independent increment process with values in G will be essential in the sequel.

Let $\mathfrak{g}$ be the "Lie algebra" of C^∞ vector fields on R^d. $\mathfrak{g}$ may be considered (at least formally) as the Lie algebra of the "Lie Group" G. $\mathfrak{g}$ is then interpreted as the Lie algebra of the tangent vectors to G which are invariant <u>on the right</u>. Equation (1.1) may be formally rewritten as

$$(1.7) \qquad d\varphi = X_o(\varphi)dt + X_i(\varphi).dw^i$$

$$\varphi(0) = id \ .$$

$(X_o(\varphi),\ldots X_m(\varphi)$ is clearly the image of $X_o(e),\ldots,X_m(e)$ through the action of φ on the right).

A case of special interest is the case where the Lie algebra $\mathfrak{g}_o$ generated by $X_o, X_1 \ldots X_m$ is finite dimensional, i.e. there exists a finite number of Lie brackets $Y_1 \ldots Y_r$ of the vector fields $X_o,\ldots,X_m$ such that any Lie bracket of $X_o \ldots X_m$ is a linear combination with <u>constant coefficients</u> of $Y_1 \ldots Y_r$. In this case Theorem 1.1 becomes trivially true, and no proof is even needed. In fact let G_o be a connected Lie group whose Lie algebra is $\mathfrak{g}_0$. The stochastic differential equation (1.7) can be trivially solved on G_0 , which is a finite dimensional manifold (the fact it remains in G_0 follows trivially from Stratonovitch calculus ; infinite life time is trivial by the group structure of G_0 . In fact $\mathfrak{g}_0$ and G_0 may be identified to matrices, and everything is trivial). Now G_0 can be identified to a finite dimensional group of diffeomorphisms of R^d, so that the original equation (1.7) in G has in fact been solved.

The case where $X_0, X_1 \dots X_m$ commute is in fact the easiest. If $\varphi_t^0, \varphi_t^1 \dots \varphi_t^m$ are the one parameter groups of diffeomorphisms of R^d associated to $X_0, X_1 \dots X_m$, the previous argument shows that

(1.8)
$$\varphi_t = \varphi_t^0 \circ \varphi^1_{w_t^1} \circ \dots \varphi^m_{w_t^m} .$$

This is the basis of the papers of Doss [15] and Sussmann [56] (where in fact X_0 is not supposed to commute with $X_1 \dots X_m$).

For other properties of stochastic flows, see Meyer [43].

c) The Ito-Stratonovitch formula on flows

$\varphi_t(\omega,.)$ is now a diffusion with values in G. If z_t is a continuous semi-martingale with values in R^d, a natural question to ask is to know if $\varphi_t(\omega,z_t)$ is a semi-martingale. Of course if φ_t takes its values in a finite dimensional Lie-subgroup G_0, the answer is trivially positive, since $\varphi_t(\omega,z_t)$ is the image of (φ_t,z_t) through the C^∞ mapping $(\varphi,z) \to \varphi(z)$.

In the general case, the problem was given a positive answer (without proof) by Ventzell [58]. The corresponding Ito formula was proved by Rozovskii in [48]. Independently, this result was reproved in Bismut [5], [6] and Kunita [32].

Let z_t be a continuous semi-martingale with values in R^d whose Ito-Meyer decomposition is

(1.9)
$$z_t = z_0 + A_t + \int_0^t H_i \,.\delta w^i$$

where $z_0 \in R^n$, A_t is a continuous adapted bounded variation process, such that $A_0 = 0$ and H_i is an adapted process such that $\int_0^t |H_i|^2 \, ds < +\infty$ a.s. Using the convention in (1.2), we have :

Theorem 1.2 : $\varphi_t(\omega, z_t)$ is a continuous semi-martingale, whose Ito decomposition is given by

$$(1.10) \qquad \varphi_t(\omega,z_t) = z_o + \int_o^t \Big[X_o\big(\varphi_s(\omega,z_s)\big) + \frac{1}{2}\frac{\partial X_i}{\partial x} X_i\big(\varphi_s(\omega,z_s)\big) + \frac{1}{2}\frac{\partial^2\varphi_s}{\partial x^2}(\omega,z_s)\,(H_{i_s},H_{i_s}) + \frac{\partial X_i}{\partial x}\big(\varphi_s(\omega,z_s)\big)\frac{\partial\varphi_s}{\partial x}(\omega,z_s)H_{i,s}\Big]\,ds + \int_o^t X_i\big(\varphi_s(\omega,z_s)\big)\delta w^i + \int_o^t \frac{\partial\varphi_s}{\partial x}(\omega,z_s).\delta z_s$$

In Stratonovitch form, (1.10) can be written as :

$$(1.11) \qquad \varphi_t(\omega,z_t) = z_o + \int_o^t X_o\big(\varphi_s(\omega,z_s)\big)ds + \int_o^t X_i\big(\varphi_s(\omega,z_s)\big).dw^i + \int_o^t \frac{\partial\varphi_s}{\partial x}(\omega,z_s).dz_s \ .$$

Proof : We follow the simple argument of Rozovskii [48]. Let g be a $\geqslant 0$ C^∞ function defined on R^d with compact support such that $\int g(x)dx = 1$. Set $g_\varepsilon(x) = \varepsilon^{-d} g(x/\varepsilon)$. Take $y \in R^n$. By Ito's formula, we have

$$(1.12) \qquad g_\varepsilon(z_t - y)\,\varphi_t(\omega,y) = g_\varepsilon(z_0 - y)\, y + \int_0^t g_\varepsilon(z_s - y)\Big[\Big(X_0 + \frac{1}{2}\frac{\partial X_i}{\partial x}X_i\Big)\big(\varphi_s(\omega,y)\big)ds + X_i\big(\varphi_s(\omega,y)\big).\delta w^i\Big] + \int_0^t \Big[< \frac{\partial g_\varepsilon}{\partial x}(z_s - y), \delta z > + \frac{1}{2}\frac{\partial^2 g_\varepsilon}{\partial x^2}(z_s - y)\,(H_i,H_i)ds\Big]\varphi_s(\omega,y) + \int_0^t < \frac{\partial g_\varepsilon}{\partial x}(z_s - y), H_i > X_i\big(\varphi_s(\omega,y)\big)ds$$

Note that since both sides of (1.10) are continuous, we need to prove (1.10) at a fixed time t. Also note that formula (1.10) can be stopped at adequate stopping times. Integrate (1.12) in y. Integration in y and stochastic integration with respect to δw^i, δz can be trivially interchanged on the R.H.S. of (1.12). Integrate by parts in the integrands of the R.H.S. of (1.12) in the variable y, so that only $g_\varepsilon(z_s - y)$ appears and not its differential : this is possible using the smoothness of $\varphi_s(\omega,y)$ in y. Finally make $\varepsilon \to 0$. Convergence of the bounded variations terms is trivial. For the stochastic integrals, stop (1.12) adequately. The proof of the L_2 convergence of these integrals is then elementary.

d) The equation of the inverse flow

In the case where $\varphi.(\omega,.)$ takes its values in a finite dimensional Lie-group G_0 there is nothing to prove. In fact, in this case, let ${}^gX_0$, ${}^gX_1 \dots {}^gX_m$ be the corresponding left invariant tangent vectors. The equation of $\psi = \varphi^{-1}$ is then trivially

(1.13) $$d\psi = - {}^gX_0(\psi)dt - {}^gX_i(\psi).dw^i$$

$$\psi(0) = e.$$

If G_0 is considered as a Lie transformation group of R^d, note that for $j = 0 \dots m$

$${}^gX_j(\psi) = \frac{\partial\psi}{\partial x} X_j$$

i.e. for any $x \in R^n$

$$\left({}^gX_j(\psi)\right)(x) = \frac{\partial\psi}{\partial x}(x)\; X_j(x) .$$

Now since $\frac{\partial\psi}{\partial x}(x) = \left[\frac{\partial\varphi}{\partial x}\left(\psi(x)\right)\right]^{-1}$, we see that if $y_t = \varphi_t^{-1}(\omega,x)$, then

(1.14) $$dy_t = - \left[\frac{\partial\varphi_t}{\partial x}(\omega,y_t)\right]^{-1} \left[X_0(x)dt + X_i(x).dw^i\right] .$$

Of course, in the general case, this argument does not make really sense.

In fact the following is proved in Bismut [5]-[6](see also Krylov-Rozovskii [30] for a special case).

Theorem 1.3 : Take z_t as in (1.9). Then $y_t = \varphi_t^{-1}(\omega, z_t)$ is a continuous semi-martingale which is the unique solution of the Ito stochastic differential equation

$$(1.15) \qquad dy_t = \left[\frac{\partial\varphi_t}{\partial x}(\omega,y_t)\right]^{-1}\left[\delta z - X_o(z_t)dt - X_i(z_t)\delta w^i\right]$$

$$+ \left[\frac{\partial\varphi_t}{\partial x}(\omega,y_t)\right]^{-1}\left[-\frac{1}{2}\frac{\partial X_i}{\partial x}(z_t)H_i - \frac{1}{2}\frac{\partial X_i}{\partial x}\left(\varphi_t(\omega,y_t)\right)\right.$$

$$\left(H_i - X_i(z_t)\right) - \frac{1}{2}\frac{\partial^2\varphi_t}{\partial x^2}(\omega,y_t)\left(\left[\frac{\partial\varphi_t}{\partial x}(\omega,y_t)\right]^{-1}\left(H_i - X_i(z_t)\right)\cdot\right.$$

$$\left.\left.\left[\frac{\partial\varphi_t}{\partial x}(\omega,y_t)\right]^{-1}\left(H_i - X_i(z_t)\right)\right)\right]dt$$

$$y(0) = z_o$$

In Stratonovitch form, (1.15) is written as

$$(1.16) \qquad dy_t = \left[\frac{\partial\varphi_t}{\partial x}(\omega,y_t)\right]^{-1}\left[dz - X_o(z_t)dt - X_i(z_t).dw^i\right].$$

Proof : Since a.s., for $t \geq 0$ $\varphi_t^{-1}(\omega,.)$ exists, we know a priori that $\varphi_t^{-1}(\omega,z_t)$ is a continuous process. Consider equation (1.15). Since its coefficients are C^∞ in y, by Protter, Emery, Doleans - Dade [18], we know it has a solution on a stochastic interval $[0,T[$, where T is a a.s. >0 stopping time. Now it is a simple exercise to check that (1.15) implies (1.16). Apply Theorem 1.2 using (1.11), and check that $z_t' = \varphi_t(\omega,y_t)$ is the solution of an Ito stochastic differential equation, whose unique solution is z_t . Then on $[0,T[$, $y_t = \varphi_t^{-1}(\omega,z_t)$. Deduce from this that $T = +\infty$ a.s. For the complete proof, see [5]-[6]. □

Remark 2 : Except when $X_1 \ldots X_m$ have compact support, there is no clear-cut argument proving that the solutions of (1.15) do not explode (except Theorem 1.3 itself !).

When $z_t = x \in R^n$, (1.16) is formally equivalent to (1.14), so that in fact $\psi_t = \varphi_t^{-1}$ is a solution of equation (1.13) on G. Note that in general, (1.13) is a "truly" infinite dimensional equation, which is not at all of the type of the equation (1.7) giving φ.

To see better the difference, let us remark that by Ito's formula, for any $f \in C_b^\infty(R^n)$, $x \in R^n$

$$(1.17) \qquad f\left(\varphi_t(\omega,x)\right) - \int_0^t \left[\left(X_o + \frac{1}{2} X_i^2\right)f\right]\left(\varphi_s(\omega,x)\right)ds$$

is a martingale given by

$$(1.18) \qquad \int_0^t (X_i f)\left(\varphi_s(\omega,x)\right).\delta w^i + f(x)$$

while

$$(1.19) \qquad f\left(\psi_t(\omega,x)\right) - \int_0^t \left(-X_o + \frac{1}{2} X_i^2\right)(x)\left[f\left(\psi_s(\omega,x)\right)\right]ds$$

is a local martingale given by

$$(1.20) \qquad -\int_0^t X_i(x)\left[f\left(\psi_s(\omega,x)\right)\right]\delta w^i + f(x)$$

(in (1.19)-(1.20) the vector fields X_j are acting on $f\left(\psi_s(\omega,x)\right)$ as a function of x).

e) The diffusion of tensors

If $g \in G$, we know that g acts on tensors. Namely, if $X \in T_x(R^d)$, we may define the vector $g_*X \in T_{g(x)}(R^d)$ by

$$g_*X = \frac{\partial g}{\partial x}(x)\, X .$$

If k is a 1-form at g(x), the 1-form g^*k at x is defined by

$$g^*k = \widetilde{\frac{\partial g}{\partial x}}(x)\, k .$$

From these rules, it is obvious how to define the action of g on tensors. However g_* sends vectors at x into vectors at g(x) while g^* sends forms at g(x) into forms at x. We will adopt a unified notation. Namely if K(x) is a tensor field, g^{*-1} K(x) is the tensor field which for each x is obtained as the pull back of K(g(x)) through $\frac{\partial g}{\partial x}$ (x). Namely, if X(x) is a vector field

$$\left(g^{*-1} X\right)(x) = \left(\frac{\partial g}{\partial x}\right)^{-1} (x)\ X(g(x))$$

If k(x) is a 1-form

$$\left(g^{*-1} k\right)(x) = \widetilde{\frac{\partial g}{\partial x}} (x)\ k(g(x)) \ .$$

We now recall the definition of the Lie derivative of a tensor field.

Definition 1.4 : If $X(x)$ is a C^∞ vector field, if g_t is the (local) group of diffeomorphisms associated to X, if $K(x)$ is a C^∞ tensor field on R^n, the tensor field $L_X K$ is defined by

$$L_X K(x) = \left[\frac{d}{dt}\left(g_t^{*-1} K\right)(x)\right]_{t=0} \tag{1.21}$$

Of course if $Y(x)$ is a C^∞ vector field, $L_X Y = [X,Y]$.

Theorem 1.5 : Let K be a C^∞ tensor field on R^n. Then for any $x \in R^n$

$$\begin{aligned} \varphi_t^{*-1} K(x) &= K(x) + \int_0^t \left(\varphi_s^{*-1} L_{X_0} K\right)(x) ds + \int_0^t \left(\varphi_s^{*-1} L_{X_i} K\right)(x). dw^i \\ &= K(x) + \int_0^t \left(\varphi_s^{*-1}\left(L_{X_0} + \frac{1}{2} L^2_{X_i}\right)K\right)(x) ds + \\ &+ \int_0^t \left(\varphi_s^{*-1} L_{X_i} K\right)(x). \delta w^i \ . \end{aligned} \tag{1.22}$$

Proof : The proof is elementary using Theorem 1.1 and the equations of $\frac{\partial \varphi_t}{\partial x}$ (ω,x) ,

$\left[\frac{\partial \varphi_t}{\partial x}(\omega,x)\right]^{-1}$. Approximation can also be used (see Bismut [5]). □

Of course if K is a (0,0) tensor f, i.e. a C^∞ function, (1.22) is the classical *Ito-Stratonovitch formula*. (1.22) lifts this *formula* to tensors. Note that the Ito decomposition (1.22) has a geometrical sense, since $\varphi_t^{*-1} K(x)$ is a process valued in the vector space of the considered tensors at point x.

The semi-group T_t associated to the diffusion (1.1) which acts on functions can then be lifted to tensors (see Bismut [5], Kunita [32]). Of course this lifting depends explicitly on $X_o,\ldots,X_m$ and not only on the generator $\mathcal{L}$. The generator of the lifted semi-group is given by

$$\tilde{\mathcal{L}} = L_{X_o} + \frac{1}{2} L^2_{X_i}$$

② Martingales and integration by parts

Let $f \in C_b^\infty(R^d)$, and $T > 0$. Define $h(t,x)$ by :

$$h(t,x) = E\left[f\left(\varphi_{T-t}(\omega,x)\right)\right] \quad t \leqslant T \tag{2.1}$$

It is easy to see that $h(t,x) \in C_b^\infty([0,T]\times R^d)$, and moreover that

$$\begin{cases} \frac{\partial h}{\partial t} + \frac{1}{2} X_o h + \frac{1}{2} X_i^2 h = 0 \\ h(T,.) = f \end{cases}$$

Set

$$k(t,x) = d\,h\,(t,x)$$

where d is the (exterior) differentiation operator in the variable x. k(t,x) is then a C^∞ differential form. Use now the known fact that the operator d and the Lie differential L_{X_j} commute so that

$$\frac{\partial k}{\partial t} + \left(L_{X_o} + \frac{1}{2} L^2_{X_i}\right) k = 0 \tag{2.2}$$

$$k(T,.) = d\,f$$

Using (1.22), (2.2) expresses the fact that $(\varphi_t^{*-1}\,k)\,(x)$ is a martingale valued in $T_x^*(R^n)$. Since $k(T,.) = df$, we have

$$\left(\varphi_t^{*-1}\,k\right)(x) = E^{F_t}\left(\varphi_T^{*-1}\,d\,f\right)\,(x)\ . \tag{2.3}$$

Note that this martingale property is true although the flow $\varphi_t(\omega,.)$ depends on $X_o, X_1 \ldots X_m$ and not only through the generator $\mathcal{L}$.

Of course (2.3) can be obtained directly using (2.1). In fact by (1.6),

$$h\left(t,\varphi_t(\omega,x)\right) = E^{F_t}\left[f\left[\varphi_T(\omega,x)\right]\right]$$

so that by differentiating in x, (2.3) holds trivially.

The representation of k as in (2.3) has been noted by a number of authors like Stroock [50] ; Bismut [2]-[4] used a version of (2.3) in the setting of the general theory of stochastic processes to represent the solution of backward stochastic differential equations appearing in control theory. Haussmann used (2.3) in control theory for the same purpose in [20] to represent the adjoint process in a maximum principle.

In fact using stochastic calculus, we see that

$$h\left(T,\varphi_T(\omega,x)\right) = h(0,x) + \int_0^T (X_i\,h)\left(\varphi_s(\omega,x)\right)\,\delta w^i$$

Using (2.3), we get

$$h\left(T,\varphi_T(\omega,x)\right) = h(0,x) + \int_0^T <(\varphi_s^{*-1}\,X_i)\,(x),\ E^{F_s}(\varphi_T^{*-1}\,d\,f)(x)>.\,\delta w^i \tag{2.4}$$

i.e.

$$f\left(\varphi_T(\omega,x)\right) = h(0,x) + \int_0^T <(\varphi_s^{*-1}\,X_i)\,(x),\ E^{F_s}(\varphi_T^{*-1}\,d\,f)(x)>\,\delta w^i\ . \tag{2.5}$$

The interpretation of (2.5) was that it was related with a sort of "multidimensional Girsanov transformation" (this was clear in [3]) connected with the martingale representation of certain random variables. Haussmann was then led in [21]-[22] to extend the result of Clark [13] expressing the representation of functionals of Brownian motion in terms of their differentials.

Namely let g be a bounded function defined on $\mathcal{C}([0,T];R^d)$ with values in R, which is continuous and differentiable. For every $y \in \mathcal{C}([0,T];R^d)$ $dg(y)$ is an element of the dual of $\mathcal{C}([0,T];R^d)$, i.e. is given by a bounded measure $d\mu^y(t)$ on $[0,T]$ with values in R^d, so that for $z \in \mathcal{C}([0,T];R^d)$

$$(2.6) \qquad < dg(y),z > = \int_{[0,T]} < z_t, d\mu^y(t) > .$$

From the point of view of differential geometry, since z_t is a variation of y_t, z_t is in $T_{y_t} R^n$, so that $d\mu^y(t)$ can be identified to a generalized element of $T^*_{y_t}(R^d)$. If S_t is a continuous function defined on $[0,T]$ with values in $T_x R^d$, it is feasible to set

$$\int_{[0,T]} < S_t , \varphi_t^{*-1} \, d\mu^{\varphi.(\omega,x)}(t) > = \int_{[0,T]} < \varphi_t^*(x) \, S_t , d\mu^{\varphi.(\omega,x)}(t) > .$$

We now have the main result of Haussmann [21]-[22] extending Clark [13].

Theorem 2.1 : If $u = (u^1,\dots,u^m)$ is a bounded predictable process defined on $\Omega \times R^+$ with values in R^m, then for any $x \in R^d$

$$(2.7)\ \mathbf{E}\left[g\big(\varphi.(\omega,x)\big)\int_0^T u^i \, \delta w^i\right] = E\int_0^T u^i \, ds < \varphi_s^{*-1} X_i(x), \int_{[s,T]} \varphi_v^{*-1} \, d\mu^{\varphi.(\omega,x)}(v) > \Big]$$

Proof : The proof in Haussmann [21] is a beautiful example of the use of the Girsanov transformation. We follow our presentation [7] of Haussmann's proof, as simplified by Williams [59].

For $\ell \in R$, consider the martingale

(2.8) $$Z_t^\ell = \exp\left\{\int_0^t - \ell\, u^i\, \delta w^i - \frac{1}{2}\int_0^t |\ell\, u^i|^2\, ds\right\} .$$

Let Q^ℓ be the probability measure on Ω whose density on each F_t relative to P is Z_t^ℓ, i.e.

(2.9) $$\frac{dQ^\ell}{dP}\Big| F_t = Z_t^\ell$$

Now by a classical property of the Girsanov transformation [54], under Q^ℓ, $w_t^\ell = w_t + \int_0^t \ell\, u\, ds$ is a Brownian martingale.

Consider the stochastic differential equation

(2.10) $$dx^\ell = X_0(x^\ell)dt + X_i(x^\ell)\ (dw^i + \ell\, u^i\, ds)$$
$$x^\ell(0) = x .$$

By Theorem 1.3, we know that $x_t^\ell = \varphi_t(\omega, y_t^\ell)$ where y_t^ℓ is the solution of the differential equation

(2.11) $$dy^\ell = \left(\varphi_t^{*-1} X_i\right)(y^\ell)\ \ell\, u^i\, ds$$
$$y^\ell(0) = x .$$

Note that Theorem 1.3 only says that (2.11) has an a.s. non exploding solution, i.e. the a.s. can depend on ℓ. To avoid difficulties, it is feasible to assume that $X_1, \ldots, X_m$ have compact support so that (2.11) is really a "standard" differential equation, and later to remove this assumption. Now under Q^ℓ, $x^\ell_\cdot$ has the same law as $x_\cdot = \varphi_\cdot(\omega, x)$ under P. This implies that

(2.12) $$E\left[\, Z_T^\ell\, g\,(x^\ell)\right] = E\left[g(x)\right]$$

(expectations are taken with respect to P).

The differential of (2.12) in the variable P is then 0. Now

$$(2.13)\qquad \frac{dZ_t^{\ell}}{d\ell}\Big|_{\ell=0} = -\int_0^T u^i\,\delta w^i\ .$$

Moreover using (2.11), y^{ℓ} is seen to be differentiable and moreover

$$(2.14)\qquad \frac{dy_t^{\ell}}{d\ell}\Big|_{\ell=0} = \int_0^t \left(\varphi_s^{*-1} X_i\right)(x)u^i\,ds$$

For any $\omega \in \Omega$ $\ell \to x^{\ell}$ is differentiable as a function from R into $\mathcal{C}([0,T]\,;R^d)$ and then

$$(2.15)\qquad \frac{dx_t^{\ell}}{d\ell}\Big|_{\ell=0} = \varphi_t^{*}\int_0^t \left(\varphi_s^{*-1} X_i\right)(x)u^i\,ds\ .$$

After checking that differentiation under E is feasible in the L.H.S. of (2.12), (2.7) follows. □

We now give the main result of Haussmann [21]-[22].

Corollary : The following equality holds a.s. on Ω *:*

$$(2.16)\qquad g\big(\varphi_{\cdot}(\omega,x)\big) = E\Big[g\big(\varphi_{\cdot}(\omega,x)\big)\Big] + \int_0^T <\varphi_s^{*-1}X_i(x),\ H_s>\ .\ \delta w^i$$

where H *is the predictable projection of* U_s *given by*

$$(2.17)\qquad U_s = \int_{[s,T]} \varphi_v^{*-1}\, d\mu^{\varphi_{\cdot}(\omega,x)}(v)$$

Proof : We know by a result of Ito [41] that any square integrable F_T-measurable random variable M_T can be uniquely represented as

$$M_T = a + \int_0^T u^i\,\delta w^i$$

where $a \in R$, and $u = (u^1 \ldots u^m)$ is adapted and such that $E\int_0^T |u|^2\,ds < +\infty$. Using (2.7), it is not hard to deduce (2.16). □

Observe that the Corollary of Theorem 2.1 is the natural extension of (2.3). In fact Davis in [60] gave a proof of this result using a representation of the type (2.3). Also note that the theory of flows has been used to reduce differentiation in ℓ of (2.10) to a standard calculus of variations on a differential equation. When differentiation of tensors will be later needed, this will help us to keep track of the various computations.

However, a superficial look at formula (2.7) shows that g appears on the L.H.S., and dg on the R.H.S., so that formula (2.7), has all the features of an <u>integration by parts</u> formula. Of course this fact was certainly obscured by the fact that the purpose of control theorists was to obtain a martingale representation result.

In fact let $g(x)\,dx$ be the standard gaussian law on the euclidean space R^k, and $f \in C_c^\infty(R^k)$. If $a \in R^k$, we have

$$(2.18) \qquad \int_{R^k} < df(x),a > g(x)dx = \int_{R^k} f(x) < a,x > g(x)dx$$

Of course (2.18) is trivial because we can use the translation invariance of the Lebesgue measure and standard differential calculus on R^k. An unnatural way of obtaining (2.18) would be to observe that under the probability law $\exp(-<\ell a, x> - \frac{1}{2}|\ell a|^2)$ $g(x)dx$, the law of $x + \ell a$ is equal to $g(x)dx$ so that

$$(2.19) \qquad \int_{R^k} f(x+\ell a) \exp\left[-<\ell a,x> - \frac{1}{2}|\ell a|^2\right] g(x)dx = \int_{R^k} f(x)\; g(x)dx$$

and so (2.18) follows by differentiation in the variable ℓ.

If $Y(x)$ is a C^∞ vector field on R^k with compact support, (2.18) extends to

$$(2.20) \qquad \int_{R^k} (Yf)(x)\; g(x)dx = \int_{R^k} f(x)\left[< Y(x), x > - (\operatorname{div} Y)(x)\right] g(x)dx \; .$$

To obtain (2.20) from (2.18), it suffices to apply (2.18) to $a = e_i$, replacing f par fY^i $(1 \leq i \leq k)$ and to sum the formulas in $1 \leq i \leq k$.

The analogy of (2.7) and (2.20) can now be described. P may be considered as the gaussian cylindrical measure on the Hilbert space $(L_2(R^+ ; R))^m$, which is in fact of 0 measure for P . Now an elementary property of gaussians shows P is quasi-invariant under constant (i.e. non random) translations in $(L_2(R^+ ; R))^m$ so that (2.19), and then (2.18) makes sense. However in (2.7), we have used the full strength of the Girsanov transformation i.e. used the quasi-invariance of P under non anticipating random translations in $(L_2(R^+ ; R))^m$, and so we obtain a formula very similar to (2.20) without having to do a necessarily infinite summation to obtain it.

③ The Malliavin calculus

a) Finite dimensional calculus

Consider the space R^k endowed with a probability law $g(x)dx$ where $g \in C^\infty(R^k)$. Let Φ be a C^∞ mapping from R^k into R^d, and let μ be the probability law of Φ on R^d. A natural question to ask is to know if μ is absolutely continuous with respect to the Lebesgue measure, and if the corresponding density is smooth.

Natural assumptions are that $k \geq d$, and that at each $x \in R^k$, the rank of $\Phi'(x) = \frac{\partial \Phi}{\partial x}(x)$ is maximal i.e. equal to d.

a) If Φ is proper, i.e. if for any bounded set B in R^d, $\Phi^{-1}(B)$ is bounded, then it is trivial to see that μ has a smooth density with respect to the Lebesgue measure using localization and the implicit function theorem.

b) If Φ is not proper, the existence of a density for μ is still trivial, but smoothness is not guaranteed. (take $k = 1$, $g = \frac{1}{\pi(1+x^2)}$, $\Phi(x) = \text{Arctg}\, x$,

$d\mu(x) = 1_{|x| \leq \frac{\pi}{2}} \frac{dx}{\pi}$). In case b), a direct study is necessary.

Take $f \in C_b^\infty(R^d)$. Let $h(x)$ be a C^∞ vector field on R^k with compact support. Now using integration by parts as in (2.20) (where g was supposed to be gaussian) we get

$$\text{(3.1)} \qquad \int_{R^k} < f'(\Phi(x)),\ \Phi'(x)\ h(x) > g(x)dx \; +$$

$$\int_{R^k} f(\Phi(x)) \left(\frac{\operatorname{div} hg}{g}\right)(x)\ g(x)dx = 0 \quad .$$

To prove that μ is given by a C^∞ density, it suffices to show that its differentials in distribution sense are bounded measures. In fact this implies that if $\hat{\mu}(\alpha)$ is the Fourier transform of μ, for any m, $|\alpha|^m\ \hat{\mu}(\alpha)$ is bounded, and so the result follows trivially.

If we want to show that the first order differentials of μ are bounded measures, it suffices to prove that formula (3.1) still makes sense if h can be chosen in (2.21) so that $\Phi'(x)\ h(x) = \frac{\partial}{\partial y^\ell}$ $(1 \leqslant \ell \leqslant d)$ (of course h will not have compact support !).

If R^k is endowed with a natural Hilbert space structure (which is the case when g(x)dx is the gaussian law), a natural choice of h(x) is to take the element in $\left\{H \in R^k\ ;\ \Phi'(x)\ H = \frac{\partial}{\partial y^\ell}\right\}$ (which is not empty) of minimal norm namely

$$\text{(3.2)} \qquad h(x) = \Phi'^*(x) \left[\Phi'\Phi'^*\right]^{-1}(x)\ \frac{\partial}{\partial y^\ell}$$

where $\Phi'^*(x)$ is the transpose of $\Phi'(x)$ (which maps $T^*_{\Phi(x)}R^k$ into the dual space of R^k, identified to R^k).

Of course it must still be proved that this choice of h is feasible, i.e. that the integral in the second term in the L.H.S. of (3.1) makes sense (which is not the case in the one dimensional counterexample given before).

When applying this procedure repeatedly, i.e. by doing as many integrations by parts as needed, we can then prove that the derivatives of μ are bounded measures and then that μ has a C^∞ density.

In his seminal papers [39], [40] Malliavin has shown this procedure could be applied to the solutions of stochastic differential equations.

b) The Malliavin calculus

Namely for $x_0 \in R^d$, consider the solution $\varphi_t(\omega,x_0)$ of (1.1). For a given $T > 0$, consider the mapping

$$\omega \to \varphi_T(\omega,x_0) \ . \tag{3.3}$$

In the previous argument R^k is replaced by Ω, $g(x)dx$ by P and Φ by the mapping (3.3). Of course, to have a complete analogy with the argument of a), P should be regarded as the gaussian cylindrical measure on the Hilbert space $\left[L^2(R^+ ; R)\right]^m$.

In [39], Malliavin has developped a technique of integration by parts based on the use of the Ornstein-Uhlenbeck operator A which is an unbounded self-adjoint operator operating on $L_2(\Omega,P)$. This technique has been completely clarified and extended by Stroock [51]-[53], and Shigekawa [49].

In [40], Malliavin showed that the extension to this situation of the choice of h given by (3.2) was feasible in the sense made precise in a), in an argument later completed by Ikeda-Watanabe [26].

In [7], we suggested a different route to the integration by parts result, which was precisely what has been described in section 2.

Since the approach taken by Malliavin [39]-[40], Stroock [51],[53], Shigekawa [49], Ikeda-Watanabe [26] to the integration by parts is completely described in these papers, we will focus now on our paper [7].

Before going into details, we feel that the essential contribution of Malliavin has been to show that the classical differential and integral calculus can be successfully extended on the probability space of Brownian motion, by using the classical Ito calculus on diffusions to express the various quantities appearing in the integration by parts procedure.

We now develop as in [7] the first application of the Malliavin calculus.

c) An application : Hörmander's theorem

In his celebrated paper [24], Hörmander proved that if O is an open set in R^d such that at each $x \in O$, the vector space spanned by $X_0(x),\dots,X_m(x)$ and their Lie brackets at x is equal to R^d, the operator $\mathcal{L}$ is hypoelliptic, i.e. if u is a distribution such that $\mathcal{L}\,u \in C^\infty(O)$, then $u \in C^\infty(O)$.

If we apply this result to the differential operator $\bar{\mathcal{L}}$ operating on $C^\infty(R \times R^d)$

$$\bar{\mathcal{L}} = \frac{\partial}{\partial t} + \mathcal{L}$$

we see that if O is an open set in R^d such that at any $x \in O$, the vector space T'_x in R^d spanned by $X_1(x),\dots,X_m(x)$ and all the brackets of length ≥ 2 of $X_0, X_1 \dots X_m$ at x span the whole space R^d, $\bar{\mathcal{L}}$ is hypoelliptic on $R \times O$.

Now if $p_t(x,dy)$ are the transition probabilities of the diffusion (1.1), we know that if $\mathcal{L}^*$ is the adjoint operator of $\mathcal{L}$ (with respect to the Lebesgue measure), then for any $x \in R^n$,

$$\text{(3.4)} \qquad \left(\frac{\partial}{\partial t} - \mathcal{L}^*_y\right) p_t(x,y) = 0 \qquad \text{on }]0,+\infty[\times R^d .$$

By applying Hörmander's theorem, we see that under the previous conditions, for any $x \in R^d$, $p_t(x,dy) = p_t(x,y)\,dy$, where $p_t(x,y)$ is C^∞ in the variables $(t,y) \in]0,+\infty[\times O$. In fact Hörmander's theorem has been used by Ichihara and Kunita [25] to prove the corresponding results on the transition probabilities.

We will now try to show how the Malliavin calculus can be used to prove directly that the $p_t(x,dy)$ are smooth in (t,y). Note that using pseudo-differential operators, Kohn [27] has simplified Hörmander's original proof (for a systematic exposition, see Trêves [57]).

We will not show how the hypoellipticity of $\bar{\mathcal{L}}$ can be recovered from the smoothness of $p_t(x,dy)$ but concentrate on a probabilistic proof of the smoothness of $p_t(x,dy)$.

Let H be a continuous function defined on $\mathcal{C}([0,T];R^d)$ with values in $T^*_x(R^d)$ which is bounded, differentiable, with a uniformly bounded differential. For $y \in \mathcal{C}([0,T];R^d)$, dH(y) can be identified to a finite measure $dv^y(t)$ on [0,T] with values in $R^d \otimes R^d$, so that

$$(3.5) \qquad z \in \mathcal{C}([0,T];R^d) \to \langle dH(y),z \rangle = \int_{[0,T]} dv^y(t)\, z_t$$

$dv^y(t)$ can be identified to a generalized linear mapping from $T_{y_t}(R^d)$ into $T^*_x(R^d)$. If $\ell \in T_x(R^d)$ we define the action of $\int_0^T \varphi_t^{*-1}\, dv^y(t)$ on ℓ by

$$(3.6) \qquad \int_0^T \varphi_t^{*-1}\, dv^y(t)\ (\ell) = \int dv^y(t)\ \varphi_t^*(\ell) \ .$$

We then have the following result in Bismut [7].

Theorem 3.1 : Let $f \in C_b^\infty(R^d)$. Then the following equality holds

$$(3.7) \qquad E\Big[f\big(\varphi_T(\omega,x)\big) \langle H\big(\varphi_\cdot(\omega,x)\big), \int_0^T \big(\varphi_s^{*-1} X_i\big)(x)\delta w^i \rangle \Big] =$$

$$E \langle H\big(\varphi_\cdot(\omega,x)\big), \int_0^T \big(\varphi_s^{*-1} X_i\big)(x) \rangle \langle \big(\varphi_s^{*-1} X_i\big)(x)ds, \big(\varphi_T^{*-1}\, df\big)(x) \rangle +$$

$$+ E\Big[f\big(\varphi_T(\omega,x)\big) \int_0^T \langle \big(\varphi_s^{*-1} X_i\big)(x), \int_{[s,T]} [\varphi_h^{*-1}(\omega,x) dv^{\varphi_\cdot(\omega,x)}(h)]$$

$$\big(\varphi_s^{*-1} X_i\big)(x) \rangle ds \Big].$$

Proof : Write $h = \sum_1^d h_k e^k$. Use then formula (2.7), where g is $h_k(x)\ f(x_T)$, and $u^i = \big(\varphi_s^{*-1} X_i\big)^k(x)$ (which is adapted !). Summing the corresponding formulas in k, we get (3.7). □

Of course we have obtained (3.7) by analogy with (3.1). In fact using (2.15),

$\Phi'(\omega)$ can be interpreted as the mapping.

$$u \in L_2([0,T] ; R^m) \to \varphi_T^* \int_0^T \varphi_s^{*-1} X_i \; u_i \; ds \in T_{\varphi_T(\omega,x)} \; R^d$$

The adjoint mapping $\Phi'^*(\omega)$ is clearly

$$p \in T^*_{\varphi_T(\omega,x)} \; R^d \to < \varphi_T^* \, \varphi_s^{*-1} X_i \, (x), p > \in L_2([0,T] ; R^m)$$

so that $\Phi' \; \Phi'^*(\omega)$ is the mapping

$$(3.8) \qquad p \in T^*_{\varphi_T(\omega,x)} \; R^d \to \varphi_T^* \int_0^T \varphi_s^{*-1} X_i \, (x) < \varphi_s^{*-1} X_i \, , \, \varphi_T^{*-1} \, p > ds$$

Formula (3.7) is then the analogous of formula (3.1) with

$$(3.9) \qquad h = \Phi'^*(\omega) \; \varphi_T^* H$$

The choice of h in (3.8) is of course justified by the argument leading to (3.2).

Consider now the following assumption H1 : $x \in R^d$ is such that the vector space spanned by $X_1(x), \ldots X_m(x)$, and the brackets of $X_0, X_1 \ldots X_m$ of length ≥ 2 at x generate R^d.

__Definition 3.2__ : __For__ $T > 0$, $C_T(\omega)$ __is the linear mapping from__ $T_x^* R^d$ __into__ $T_x R^d$ __given by__

$$(3.10) \qquad p \to \sum_{i=1}^{m} \int_0^T < \varphi_s^{*-1} \; X_i \, (x), p > (\varphi_s^{*-1} \; X_i \,)(x) ds \; .$$

$C_T(\omega)$ defines a symmetric positive form on $T_x^* \; (R^d)$, and by (3.8) is clearly related to $\Phi' \; \Phi'^*(\omega)$.

Recall that in a), Φ' was supposed to be of maximal rank d (here $k = +\infty$!).

The first spectacular result of Malliavin [39] was that precisely H1 implies that $\Phi' \Phi'^*$ is a.s. invertible. Namely

Proposition 3.3 : Under H1, a.s., for any $T > 0$, $C_T(\omega)$ is invertible.

Proof : We sketch the proof in [39]-[7]. If $f \neq 0$ is such that $C_T(\omega)\, f = 0$, clearly $< \varphi_s^{*-1} X_i(x), f > = 0$ for $s \leqslant T$. Using (1.22), it follows that $< \varphi_s^{*-1}[X_j,X_i](x),f >=0$ $(1 \leqslant i,j \leqslant m\ldots, s \leqslant T)$ and so $< \varphi_s^{*-1} [X_j,[X_j,X_i]](x),f > = 0$ $(s \leqslant T)$, so that canceling the drift term in $< \varphi_s^{*-1} X_i (x),f >$,$< (\varphi_s^{*-1} X_0)(x),f > = 0$ $(s \leqslant T)$. Iterating the procedure, we find that in particular f is orthogonal to $X_1(x)\ldots X_m(x)$ and to the brackets considered in H1, so that $f = 0$. Of course negligible sets must be worked out [7]. □

As noted in Malliavin [39] (see Stroock [53]) Proposition 3.3 is enough to show that the law of $\varphi_T(\omega,x)$ is given by a density $p_t(x,y)dy$.

Recall that in a), it was underlined that a feasibility assumption was to be checked on the choice (3.2).

Malliavin [40] and Ikeda-Watanabe [26] were able to prove that when H1 is still verified excluding X_0 - i.e. by considering only $X_1\ldots X_m$ and their brackets - this is the case. More recently, Kusuoka and Stroock [37]-[52] were able to prove that this is still the case under H1.

Namely

Theorem 3.4 : If $x \in R^d$ is such that H1 is verified, then for any $t > 0$, $\| C_t^{-1} \|$ belongs to all the $L_p (1 \leqslant p < +\infty)$.

Proof : It is out of question to give here the whole proof of Malliavin [40], Ikeda-Watanabe [26], Kusuoka and Stroock [37]-[52], so that we only give a brief sketch.

Take $u \in R$, and let b_t be a one-dimensional Brownian motion such that $b_0 = 0$. We have the easy equality

$$(3.11) \qquad E\left[\exp - \frac{\alpha^2}{2}\int_0^T |x + b_s|^2\, ds\right] = (\operatorname{ch} \alpha T)^{-1/2} \exp\left[- \frac{\alpha x^2}{2}\, t\, h(\alpha T)\right].$$

Using (3.11) and Čebyšev inequality, it can be shown that if X_t is a continuous semi-martingale on a filtered probability space $(\Omega, \{F_t\}_{t \geq 0}, P)$ whose Ito-Meyer decomposition is

$$X_t = X_o + \int_o^t A_s \, ds + N_t \tag{3.12}$$

where

- X_o is F_o-measurable
- A is predictable and such that $|A| \leq M$
- N_t is a Brownian martingale such that $N_o = 0$

then for any $\varepsilon > 0$, $T > 0$

$$P\left[\int_o^T |X_s|^2 ds < \varepsilon\right] \leq \sqrt{2} \exp - \left[16\left(\frac{\varepsilon}{T^2} + \frac{M^2 T}{3}\right)\right]^{-1} \tag{3.13}$$

Now under the restricted assumptions of Malliavin [40], Ikeda-Watanabe [26], for $f \in R^d$, $\|f\| = 1$, there is one bracket of $X_1 \ldots X_m$, which will be written $X_{[I]}$ such that $|<f, X_I(x)>| \geq \eta > 0$. If $I = i$ $(1 \leq i \leq m)$, $< C_t f, f >$ will be large enough. If $X_{[I]} = [X_i, X_{[J]}]$, by using (1.22), we see that $< f, \varphi_s^{*-1} X_{[J]} >$ is a semi-martingale such that its quadratic variation is given by

$$\sum_{k=1}^{m} \int_o^t < f, \varphi_s^{*-1} [X_k, X_{[J]}] >^2 ds \tag{3.14}$$

By doing a time change on $< f, \varphi_s^{*-1} X_{[J]} >$, we can then go back to the situation studied in (3.12)-(3.13) and find that the probability that $\int_o^t < f, \varphi_s^{*-1} X_{[J]} >^2 ds$ is "small" is itself small enough. By induction, we can prove that for any $k \in N$

$$P\left[\int_o^t < f, \varphi_s^{*-1} X_i >^2 ds < \varepsilon\right] = o(\varepsilon^k) \quad \varepsilon \to 0 \tag{3.15}$$

The theorem follows easily from (3.15) as in [40]-[26]. In fact in [40]-[26], instead of (3.11), an estimate on the variance of b_s on [0,1] was obtained using the Fourier series representation of the Brownian motion.

In the case where X_0 is necessary to fulfill H1, Kusuoka and Stroock [37]-[52] use the fact that the drift term in the Ito-Meyer decomposition of (1.22) is a.s. a Hölder function. Combining this fact with the previous estimates, they obtain the Theorem in full generality. □

From Theorem 3.5, it is then possible to obtain :

Theorem 3.5 : If $x \in R^d$ is such that H1 is verified at x, for any $t > 0$, the law of $\varphi_t(\omega,x)$ is given by $p_t(x,y)dy$, where $p_t(x,.) \in C_b^\infty(R^d)$.

Remark : Using the techniques of Stroock [51], the previous results can be localized, so that the smoothness of $p_t(x,dy)$ depends only on the behaviour of $X_0, X_1 \dots X_m$ on a neighbourhood of y.

Moreover the techniques of Bismut [7] used in combination with Kusuoka-Stroock [37]-[52] give the existence of a smooth resolvent operator $V^\lambda(x,y)$ ($\lambda > 0$, $y \neq x$) associated to $\mathcal{L}$, when $\mathcal{L}$ verifies Hörmander's conditions (the resolvent can be smooth while the semi-group is not !).

④ Application to filtering

The Malliavin calculus has been applied to situations where standard analytic techniques do not work.

In [23], Holley and Stroock have obtained results concerning the existence and smoothness of the finite-dimensional distributions of an infinite system of interacting diffusions. In this case, while infinite dimensional analogues of Partial Differential equations appear to be difficult to use, the Malliavin calculus is an efficient tool to prove the indicated smoothness results.

Here, we will concentrate on the results of Bismut-Michel [11] extending Michel [44] on conditional diffusions.

a) A finite dimensional analog

We start again with the situation studied in Section 3 a). Besides we assume that $x = (x',x'')$ where $x' \in R^{k'}$, $x'' \in R^{k''}$ (of course $k' + k'' = k$).

We want to study the smoothness of the conditional law of $\Phi(x)$ given x''. A natural assumption is that $k' \geq d$, and the partial differential $\Phi'_{x'}(x',x'')$ is of maximal rank d.

Following the ideas of 3 a), it is natural to try to obtain a conditional formula of integration by parts. Namely let h(x) be a C^∞ function defined on R^k with valus in $R^{k'}$ which has compact support. Let U(x") be a measurable bounded function on $R^{k''}$. For $f \in C_b^\infty(R^d)$, we have

$$(4.1) \qquad \int_{R^k} U(x'') < f'(\Phi(x)), \Phi'_{x'}(x)\, h(x) > g(x)dx$$

$$+ \int_{R^k} U(x'')\, f(\Phi(x))\, \frac{\mathrm{div}_{x'}(h\, g)}{g}\, (x)\, g(x)dx = 0$$

(in (4.1) the divergence is taken with respect to the variable x'). Of course (4.1) is a consequence of (3.1), where instead of taking a general h, we chose h with values in $R^{k'}$ (the extension of (4.1) to a measurable U is trivial).

Assuming that the choice

$$h(x) = \Phi'^{*}_{x'}\, (\Phi'_{x'}\, \Phi'^{*}_{x'})^{-1}\, (x)\, \frac{\partial}{\partial y^\ell}$$

is feasible in (4.1), we see that

$$(4.2) \qquad \int_{R^k} U(x'')\, (\frac{\partial}{\partial y^\ell} f)\, (\Phi(x))\, g(x)dx + \int_{R^k} U(x'')\, f(\Phi(x))\, K_\ell(x)\, g(x)dx = 0$$

If $P_{x''}(dx)$ is the conditional law of x given x", we get from (4.2)

$$(4.3) \qquad \int_{R^k} \frac{\partial}{\partial y^\ell} f(\Phi(x))\, P_{x''}(dx) = -\int_{R^k} f(\Phi(x))\, K_\ell(x)\, P_{x''}(dx) \quad \text{a.s.}$$

By eliminating negligible sets adequately, it is clear that a.s., the differential of the law of $\Phi(x)$ given x'' is a bounded measure.

By iterating the procedure, we may obtain the smoothness of the conditional law of $\Phi(x)$ given x''.

b) Application to diffusions

From (2.7), we may obtain the analog of (4.1). Assume that $w = (w', w'')$ where $w' = (w^1 \dots w^{m'})$, $w'' = (w^{m'+1} \dots w^m)$. Assume that in (2.7), g is replaced by $U(w'')\, g(\varphi_\cdot(\omega, x))$, where we suppose (at the beginning !) that U verifies the same type of assumptions as g.

If, in (2.7), we assume that $(u^{m'+1}, \dots, u^m)$ are all 0, it is clear that on the r.h.s. of (2.7), no differential of U will ever appear, so that (2.7) can be extended by assuming that U is only bounded and measurable.

Conditional expectations with respect to w'' may then be taken on both sides of (2.7).

c) The main results on conditional diffusions

Consider the system of stochastic differential equations

$$(4.4) \quad \begin{aligned} dx &= X_0(x,z)dt + \sum_{i=1}^m X_i(x,z).dw^i + \sum_{j=1}^d \tilde{X}_j(x,z)(d\tilde{w}^j + \ell^j(x,z)dt) \\ x(0) &= x_0 \\ dz &= Z_0(z)dt + \sum_{j=1}^d Z_j(z)\,(d\tilde{w}^j + \ell^j(x,z)dt) \\ z(0) &= z_0 \end{aligned}$$

where $w = (w^1 \dots w^m)$, $\tilde{w} = (\tilde{w}^1 \dots \tilde{w}^d)$ are independent Brownian motions. Of course the vector fields and the functions appearing in (4.4) are supposed to be smooth with bounded differentials.

$x_t \in R^n$ is supposed to be the state of the system and $z_t \in R^p$ the observation on the system.

Set

$$F_t^z = \mathcal{B}(z_s | s \leqslant t)$$

Using the results of Schwartz in the theory of prediction [42], it is easy to see that there exists a continuous process τ_t^z with values in the set Π of probability measures on $\mathcal{C}(R^+ ; R^n)$ such that for every $\{F_t^z\}_{t \geqslant 0}$-stopping time T, on $(T < +\infty)$, τ_T^z is the conditional law of the process x given F_T^z .

In Bismut-Michel [11], the techniques of the calculus of variations have been applied to study in detail the process $\{\tau_T^z\}$.

Note that the technique briefly summarized in b) allows us not to make any distinction between smoothing, filtering, and prediction, since the process τ_T^z can be analyzed globally.

The main results in Bismut-Michel [11] are

Theorem 4.1 : Assume that $y_o = (x_o, z_o)$ is such that the vector subspace of $R^n \times \{0\}$ spanned by $((X_1(y_o), 0) \ldots ((X_m(y_o), 0))$ and all the Lie brackets at y_o of the vectors $((X_1, 0), \ldots, (X_m, 0),\ (\tilde{X}_1, Z_1), \ldots, (\tilde{X}_m, Z_m))$ in which $(X_1, 0), \ldots, (X_m, 0)$ appear at least once is equal to $R^n \times \{0\}$. Then a.s., for every $T \geqslant 0$, $t > 0$, the law of x_t for τ_T^z is equal to $q_{t,T}^z(y)dy$, where $q_{t,T}^z(y) \in C_b^\infty(R^n)$.

This result extends the results of Pardoux [45] where x_0 was assumed to have itself a probability law given by a density - while here we assume that x_0 is fixed - and moreover in [45] (4.4) was supposed to be partially elliptic in the sense that $X_1 \ldots X_m$ had to span R^n . Theorem 4.1 is an extension of Hörmander's theorem, but is still a probabilistic result (i.e. there is a a.s. in it).

Theorem 4.2 : Assume that $p = d$ and that $Z_1 \dots Z_d$ span R^d. Then if $X_1 \dots X_m$, $\ell_1 \dots \ell_d$ have compact support, a.s., for every $T \geqslant 0$, τ_T^z is the law of a non homogeneous Feller process.

The assumption on $Z_1 \dots Z_d$ is standard. The compactness assumption is technical, and can be weakened.

Since in general for τ_T^z , x_t is not a semi-martingale, a change of variables is done in [11] with the help of a flow depending only on z so as to get a standard diffusion process (under τ_T^z).

Under the same assumptions as Theorem 4.2, the unnormalized filtering equation is reduced to a standard - i.e. with no diffusion term - partial differential equation, with coefficients very irregular in time. This is done in general by using the theory of stochastic flows inside the Girsanov density so as to perform an integration by parts inside the Girsanov density (which is trivial in the standard filtering problem [38]).The results of Eliott and Kohlmann [16] and Davis [14] are then extended in full generality.

Note that except in the partially elliptic case where the filtering equation can be solved "pointwise", under the assumptions of Theorem 4.1, it is not clear at all how could classical analysis techniques be applied, because of the non differentiability of the coefficients of the equation in time, in order to obtain analytically the same result. In the partially elliptic case, the results of Pardoux [45] (on filtering) can be reobtained using the filtering equation of [11].

REFERENCES

[1] P. Baxendale :

Wiener processes on manifolds of maps ,

Proceedings Royal Soc. Edinburgh,87A,127-152(1980).

[2] J.M. Bismut :

Linear-quadratic optimal stochastic control with random coefficients ,

SIAM J. of Control 14, 419-444 (1976)

[3] J.M. Bismut :

An introductory approach to duality in optimal stochastic control ,

SIAM Review 20, 62-78 (1978).

[4] J.M. Bismut :

Contrôle des systèmes linéaires-quadratiques ,

Séminaire de Probabilités n° XII, p. 180-264, Lecture Notes in Math. 649
Berlin-Springer 1978.

[5] J.M. Bismut :

Mécanique aléatoire ,

Lecture Notes in Math. n° 866, Berlin Springer 1981.

[6] J.M. Bismut :

A generalized formula of Ito and some other properties of stochastic flows ,

Z. Wahrsch, 55, 331-350 (1981).

[7] J.M. Bismut :

Martingales, the Malliavin calculus and hypoellipticity under general Hörmander's conditions ,

Z. Wahrsch, 56, 469-505 (1981).

[8] J.M. Bismut :

Calcul des variations stochastique et processus de sauts ,

A paraître.

[9] J.M. Bismut :

The calculus of boundary processes ,

To be presented at the Kyoto Conference in Probability (1982).

[10] J.M. Bismut - D. Michel :
Structure des diffusions conditionnelles et calcul des variations ,
C.R.A.S. 292, Série I, 731-734 (1981).

[11] J.M. Bismut - D. Michel :
Diffusions conditionnelles ,
J. of Funct. Anal. Part I 44, 174-211 (1981), Part II 45, 274-292 (1982).

[12] Y.N. Blagoveschenskii - M.I. Freidlin :
Certain properties of processes depending on parameters ,
Sov. Math. Dokl. 2, 633-636 (1961).

[13] J.M.C. Clark :
The representation of functionals of Brownian motion by stochastic integrals ,
Ann. Math. Stat 41, 1282-1295 (1970), 42, 1778 (1971).

[14] M.H.A. Davis :
Pathwise non linear filtering ,
"Stochastic systems"(M. Hazewinkel ed.) Dordrecht : Reidel. To appear.

[15] H. Doss :
Liens entre équations différentielles stochastiques et ordinaires ,
Ann. Inst. H. Poincaré 13, 99-125 (1977).

[16] R.J. Eliott - M. Kohlmann :
Robust filtering for correlated multidimensional observation ,
Preprint 1981.

[17] K.D. Elworthy :
Stochastic dynamical systems and their flows ,
In "Stochastic Analysis" A. Friedman and M. Pinsky ed. p 79-95 ,
New-York : Acad. Press 1978.

[18] M. Emery :
Equations différentielles stochastiques lipschitziennes : étude de la stabilité ,
Séminaire de Probabilités n° 13 , p. 281-293, Lecture Notes in Math 721,
Berlin : Springer 1979.

[19] I.I. Gihman - A.V. Skorohod :
Stochastic differential equations ,
Erg. der Math. und ihrer Grenzgeb, Band 72, Berlin : Springer 1972.

[20] U. Haussmann :
On the stochastic maximum principle ,
SIAM J. of Control and Opt. 16, 236-251 (1978).

[21] U. Haussmann :
Functionals of Ito processes as stochastic integrals ,
SIAM J. of Control and Opt. 16, 252-269 (1978).

[22] U. Haussmann :
On the integral representation of Ito processes ,
Stochastics 3, 17-27 (1979).

[23] R. Holley - D. Stroock :
Diffusions on an infinite dimensional torus ,
J. of Funct. Anal. 42, 29-63 (1981).

[24] L. Hörmander :
Hypoelliptic second order differential equations ,
Acta Math. 119, 147-171 (1967).

[25] K. Ichihara - H. Kunita :
A classification of second order degenerate elliptic operators and its probabilistic characterization ,
Z. Wahrsch, 30, 235-254 (1974).

[26] N. Ikeda - S. Watanabe :
Stochastic differential equations and diffusion processes ,
Amsterdam, North-Holland 1981.

[27] J.J. Kohn :
Pseudo-differential operators and hypoellipticity ,
Proc. Symp. Pure Math. 23, 61-69 (1973).

[28] N.V. Krylov - B.L. Rozovskii :
On the Cauchy problem for linear stochastic partial differential equations ,
Math. USSR Izv. 11, 1267-1284 (1977).

[29] N.V. Krylov - B.L. Rozovskii :
On conditional distributions of diffusion processes ,
Math. USSR Izv. 12, 336-356 (1978).

[30] N.V. Krylov - B.L. Rozovskii :
On the first integral and Liouville equations for diffusion processes ,
In "Stochastic differential systems", M. Arato and D. Vermes ed.
p. 117-125,Lecture Notes in Control and Inf. Sci. 36, Berlin, Springer
1981.

[31] H. Kunita :
On the decomposition of solutions of stochastic differential equations ,
In "Stochastic Integrals", D. Williams ed, p. 213-255, Lecture Notes in
Math. n° 851, Berlin, Springer 1981.

[32] H. Kunita :
Some extensions of Ito's formula ,
Séminaire de Probabilités n° XV, p. 118-141, Lecture Notes in Math.
n° 850, Berlin, Springer 1981.

[33] H. Kunita :
Stochastic partial differential equations connected with non linear filtering ,
Proceedings of CIME ; To appear.

[34] H. Kunita :
Cauchy problems for stochastic partial differential equations arising in non linear filtering theory ,
Systems and Control Letters 1, 37-41 (1981).

[35] H. Kunita :
Densities of a measure valued process governed by a stochastic differential equation ,
Systems and Control Letters 1, 100-104 (1981).

[36] S. Kusuoka :
Analytic functionals of Wiener processes and absolute continuity ,
In "Functional Analysis in Markov processes" p. 1-46, Lecture Notes in
Math. n° 923, Berlin, Springer 1982.

[37] S. Kusuoka - D. Stroock :
To appear.

[38] R.S. Liptser - A.N. Shiryayev :
Statistics of random processes ,
Vol. I, II, Berlin : Springer 1977-1978.

[39] P. Malliavin :
Stochastic calculus of variations and hypoelliptic operators ,
Proceedings of the Conference on Stochastic differential equations of Kyoto (1976), p. 195-263, Tokyo : Kinokuniya and New-York : Wiley 1978.

[40] P. Malliavin :
C^k *hypoellipticity with degeneracy ,*
In "Stochastic Analysis", A. Friedman and M. Pinsky ed, p. 199-214, New-York : Acad. Press 1978.

[41] P.A. Meyer :
Un cours sur les intégrales stochastiques ,
Séminaire de Probabilités n° X, p. 245-400, Lecture Notes in Math n° 511, Berlin : Springer 1973.

[42] P.A. Meyer :
Sur les désintégrations régulières de L. Schwartz ,
Séminaire de Probabilités n° 7, p. 217-222, Lecture Notes in Math n° 321, Berlin : Springer 1973.

[43] P.A. Meyer :
Flot d'une équation différentielle stochastique ,
Séminaire de Probabilités n° 15, p. 103-117, Lecture Notes in Math n° 850, Berlin : Springer 1981.

[44] D. Michel :
Régularité des lois conditionnelles en théorie du filtrage non linéaire et calcul des variations stochastique ,
J. of Funct. Anal. 41, 8-36 (1981).

[45] E. Pardoux :
Stochastic partial differential equations and filtering of diffusion processes ,
Stochastics 3, 127-167 (1979).

[46] E. Pardoux :
Equations du filtrage non linéaire, de la prédiction et du lissage ,
Preprint Université de Provence 1980.

[47] M. Reed - B. Simon :

Methods of modern mathematical Physics ,

Vol I, New-York : Acad. Press 1972.

[48] B.L. Rozovskii :

A formula of Ito-Ventcell,

Vestnik Moskou Univ. Ser 1 Math Meh 26-32 (1973).

[49] I. Shigekawa :

Derivatives of Wiener functionals and absolute continuity of induced measures ,

J. Math. Kyoto Univ. 20, 263-289 (1980).

[50] D. Stroock :

On certain systems of parabolic equations ,

Comm. Pure and Applied Math. 23, 447-457 (1970).

[51] D. Stroock :

The Malliavin calculus and its applications to second order parabolic differential equations ,

Math Systems Theory, Part I 14, 25-65 (1981), Part II 141-171 (1981).

[52] D. Stroock :

Some applications of stochastic calculus to partial differential equations ,

Ecole d'été de Probabilités de Saint Flour. A paraître.

[53] D. Stroock :

The Malliavin calculus : a functional analytic approach ,

J. Funct. Anal. 44, 212-257 (1981).

[54] D. Stroock - S.R.S. Varadhan :

Multidimensional diffusion processes ,

Grundlehren Math Wissenschaften 233, Berlin : Springer 1979.

[55] D. Stroock - S.R.S. Varadhan :

On the support of diffusions with applications to the strong maximum principle ,

6 th Berkeley Symposium in Probability and Statistics, Vol III, 333-359, Berkeley : University of California Press 1972.

[56] H. Sussmann :

On the gap between deterministic and stochastic ordinary differential equations ,

Annals of Prob. 6, 19-41 (1978).

[57] F. Trêves :

Introduction to Pseudodifferential and Fourier Integral operators ,

Vol I, New-York : Plenum Press 1981.

[58] A.D. Ventzell :

On equations of the theory of conditional processes ,

Theory Prob. and Appl. 10, 357-361 (1965).

[59] D. Williams :

" To begin at the beginning " ,

In "Stochastic Integrals", D. Williams ed. p. 1-55, Lecture Notes in Math. n° 851, Berlin : Springer 1981.

[60] M.H.A. Davis :

Functionals of Ito processes as stochastic integrals ,

Proc. Cambridge Philos. Soc. 87, 157-166 (1980).