An Introduction to Malliavin Calculus

Courant Institute of Mathematical Sciences

New York University

Peter K. Friz

August 3, 2002

These notes available on
www.math.nyu.edu/phd_students/frizpete
Please send corrections to
Peter.Friz@cims.nyu.edu

These lecture-notes are based on a couple of seminar-talks I gave at Courant in Spring 2001. I am extremely grateful to Prof. S.R.S. Varadhan for supporting and stimulating my interest in Malliavin Calculus. I am also indepted to Nicolas Victoir, Enrique Loubet for their careful reading of this text.
-Peter F.

Notations:

Ω ... Wienerspace $C[0,1]$ resp. $C([0,1],\mathbb{R}^m)$
$\mathcal{F}$ ... natural filtration
H ... $L^2[0,1]$ resp. $L^2([0,1],\mathbb{R}^m)$
$H^{\otimes k}$ ... tensorproduct $\cong L^2([0,1]^k)$, $\quad H^{\hat{\otimes} k}$ ... symmetric tensorproduct
$\tilde{H}$ ... Cameron-Martin-space $\subset \Omega$, elements are paths with derivative in H
$W : \mathcal{F} \to \mathbb{R}$ ... Wiener-measure on Ω
$\beta_t = \beta(t)$ ... Brownian Motion (= coordinate process on $(\Omega, \mathcal{F}, W)$)
$W : H \to L^2(\Omega)$ ... defined by $W(h) = \int_0^1 h d\beta$
$\mathcal{S}_2$ ... Wiener polynomials, functionals of form polynomial$(W(h_1), ..., W(H_n)$
$\mathcal{S}_1$ ... cylindrical functionals, $\subset \mathcal{S}_2$
$\mathcal{D}^{k,p}$ ... $\subset L^p(\Omega)$ containing k-times Malliavin differentiable functionals
$\mathcal{D}^\infty$ ... $\cap_{k,p}\mathcal{D}^{k,p}$, smooth Wiener functionals
λ, λ^m ... (m-dimensional) Lebesgue-measure
ν, ν^n ... (n-dimensional) standard Gaussian measure
∇ ... gradient-operator on $\mathbb{R}^n$
$L^p(\Omega, H)$ ... H-valued random-variables s.t. $\int_\Omega \|\cdot\|_H dW < \infty$
D ... Malliavin derivative, operator $L^p(\Omega) \to L^p(\Omega, H)$
δ ... $= D^*$ the adjoint operator, also: divergence, Skorohod Integral
L ... $= \delta \circ D$, Ornstein-Uhlenbeck operator $L^p(\Omega) \to L^p(\Omega)$
$W^{k,p}$ ... Sobolev-spaces built on $\mathbb{R}^n$ $\quad H^k ... W^{k,2}$
∂ ... (for functions $f : \mathbb{R} \to \mathbb{R}$) simple differentiation
∂^* ... adjoint of ∂ on $L^2(\mathbb{R}, \nu)$
$\mathcal{L}$ ... $= \partial^* \partial$, one-dimensional OU-operator
$\partial_i, \partial_{ij}$ ... partial derivatices w.r.t. x_i, x_j etc
$\mathcal{L}$ ... generator of m-dimensional diffusion process, for instance $\mathcal{L} = E^{ij}\partial_{ij} + B^i \partial_i$
H_n ... Hermite-polynomials
$\Delta_n(t)$ ... n-dimensional simplex $\{0 < t_1 < ... < t_n < t\} \subset [0,1]^n$
$J(\cdot)$ ... Iterated Wiener-Ito integral, operator $L^2[\Delta_n]\ to \mathcal{C}_n \subset L^2(\Omega)$
$\mathcal{C}_n$ ... n^{th} Wiener Chaos
α ... multiinex (finite-dimensional)
X ... m-dimensional diffusion process given by SDE, driven by d BMs
$\Lambda = \Lambda(X)$ ... $< DX, DX >_H$, Malliavin covariance matrix
V, W ... vectorfields on $\mathbb{R}^m$, seen as map $\mathbb{R}^m \to \mathbb{R}^m$ or as first order differential operator

B, A_0 ... vectorfields on $\mathbb{R}^m$, appearing as drift term in Ito (resp. Stratonovich) SDE
$A_1, \ldots, A_d$ ...vectorfields on $\mathbb{R}^m$, appearig in diffusion term of the SDE
$\circ d\beta$ ... Stratonovich differential = Ito differential + (...)dt
X ... diffusion given by SDE, $X(0) = x$
Y, Z ... $\mathbb{R}^{m \times m}$-valued processes, derivative of X w.r.t. $X(0)$ resp. the inverse
∂ ... ∂V is short for the matrix $\partial_j V^i$, $\quad \nabla_W V$ ... connection, $= (\partial V)W$
$[V, W]$ ... Lie-bracket, yields another vectorfield
Lie $\{...\}$ ... the smallest vectorspace closed under Lie-brackets, containing $\{...\}$
$\mathcal{D}$ ... $= C_c^\infty$, test-functions
$\mathcal{D}'$ ... Schwartz-distributions = cont. functionals on $\mathcal{D}$

Chapter 1

Analysis on the Wiener Space

1.1 Wiener Space

Ω will denote the *Wiener Space* $C([0,1])$. As usual, we put the *Wiener measure* W on Ω therefore getting a probability space

$$(\Omega, \mathcal{F}, W)$$

where $\mathcal{F}$ is generated by the coordinate maps. On the other hand we can furnish Ω with the $\|\cdot\|_\infty$ - norm making it a (separable) Banach-space. $\mathcal{F}$ coincides with the σ-field generated by the open sets of this Banach-space. Random-variables on Ω are called *Wiener functionals*. The coordinate process $\omega(t)$ is a Brownian motion under W, with natural filtration $\sigma(\{\omega(s) : s \le t\}) \equiv \mathcal{F}_t$. Often we will write this Brownian motion as $\beta(t) = \beta(t, \omega) = \omega(t)$, in particular in the context of stochastic Wiener-Itô integrals.

1.2 Two simple classes of Wiener functionals

Let f be a polynomial, $h_1, \ldots, h_n \in H \equiv L^2[0,1]$. Define first a class of *cylindrical* functionals

$$\mathcal{S}_1 = \{F : F = f(\beta_{t_1}, \ldots, \beta_{t_n}))\},$$

then the larger class of *Wiener polynomials*

$$\mathcal{S}_2 = \{F : F = f(W(h_1), \ldots, W(h_n))\}$$

where $W(h) \equiv \int_0^1 h d\beta$.
Remarks: - Both $\mathcal{S}_i$ are algebras. In particular $\mathcal{S}_2$ is what [Malliavin2] p13 calls the *fundamental algebra*.
- A $\mathcal{S}_2$-type function with all h_i's deterministic step functions is in $\mathcal{S}_1$.
- In both cases, we are dealing with r.v. of the type

$$F = f(n\text{-dimensional gaussian}) = \tilde{f}(n \text{ indep. std. gaussians}).$$

Constructing $\tilde{f}$ boils down to a Gram-Schmidt-orthonormalization for the h_i's. When restricting discussion to $\mathcal{S}_2$-functionals one can actually forget Ω and simply work with $(\mathbb{R}^n, \nu^n)$, that is, $\mathbb{R}^n$ with n-dimensional standard Gaussian measure $d\nu^n(x) = (2\pi)^{-n/2}\exp(-|x|^2/2)dx$.
This remark looks harmless here but will prove useful during the whole setup of the theory.
- $\mathcal{S}_1 \subset \mathcal{S}_2 \subset L^p(\Omega)$ for all $p \geq 1$ as t he polynomial growth of f assures the existence of all moments. From this point of view, one could weaken the assumptions on f, for instance smooth and of maximal polyniomal growth or exponential-martingale-type functionals.

1.3 Directional derivatives on the Wiener Space

Recall that $[W(h)](\omega) = \int_0^1 hd\beta(\omega)$ is constructed as L^2-limits and hence, as element in $L^2(\Omega, W)$, only W-a.s. defined. Hence, any $\mathcal{S}_2$- or more general Wiener functional is only W-a.s. defined.
In which directions can we shift the argument ω of a functional while keeping it a.s. well-defined? By Girsanov's theorem, the *Cameron-Martin*-directions

$$\tilde{h}(\cdot) := \int_0^{\cdot} h(t)dt \in \Omega \quad \text{with } h \in H$$

are fine, as the shifted Wiener-measure $(\tau_{\tilde{h}})W$ is equivalent to W. The set of all $\tilde{h}$ is the *Cameron-Martin-space* $\tilde{H}$. It is known that for a direction $k \in \Omega - \tilde{H}$ the shifted measure is singular wrt to W, see [RY], Ch. VIII/2. Hence, $F(\omega + k)$ does not make sense, when F is an a.s. defined functional, and neither does a directional derivative in direction k.

Remarks: - The paths $\tilde{h}$ are sometimes called *finite energy paths*
- The set $\tilde{H} \subset \Omega$ has zero W-measure, since every $\tilde{h}$ is of bounded variation while W-a.s. Brownian paths are not.
- The map $h \longmapsto \tilde{h}$ is a continuous linear injection from H into $(\Omega, \|\cdot\|_\infty)$.
- Also, $h \longmapsto \tilde{h}$ is a bijection from $H \to \tilde{H}$ with inverse $\frac{d}{dt}\tilde{h}(t) = h(t)$. This derivative exists dt-a.s. since $\tilde{h}$ is absolutly continuous, moreover $h \in H$ i.e. square-integrable.
In particular, we can use this transfer the Hilbert-structure from H to $\tilde{H}$. For $g, k \in \tilde{H}$ let $\dot{g}, \dot{k}$ denote their square-integrable derivatives. Then

$$< g, k >_{\tilde{H}} \equiv < \dot{g}, \dot{k} >_H = \int_0^1 \dot{g}\dot{k}d\lambda$$

- In a more general context $\tilde{H}$ (or indeed H) are known as *reproducing kernel space* for the Gaussian measure W on the Banach space Ω (terminology from [DaPrato], p40).

1.4 The Malliavin derivative D in special cases

Take $F \in \mathcal{S}_1$, with slightly different notation

$$F(\omega) = f(\omega(t_1), \ldots, \omega(t_n)) = f(W(1_{[0,t_1]}), \ldots, W(1_{[0,t_n]})$$

Then, at $\epsilon = 0$

$$\frac{d}{d\epsilon} F(\omega + \epsilon \tilde{h})$$

equals

$$\sum_{i=1}^{n} \partial_i f(\omega(t_1), \ldots, \omega(t_n)) \int_0^{t_i} h d\lambda =< DF, h >_H$$

where we define

$$DF = \sum_i \partial_i f(W(1_{[0,t_1]}), \ldots, W(1_{[0,t_n]}) 1_{[0,t_i]}.$$

This extends naturally to $\mathcal{S}_2$ functionals,

$$DF = \sum_i \partial_i f(W(h_1), \ldots, W(h_n)) h_i,$$

and this should be regarded as an H-valued r.v.

Remarks: - D is well-dened. In particular for $F = W(h) = \int_0^1 h d\beta$ this is a consequence of the Ito-isometry.
- Sometimes it is convinient to write

$$D_t F(\omega) = \sum_i \partial_i f(W(h_1)(\omega), \ldots) h_i(t)$$

which, of course, is only $\lambda \times W$-as well-defined.
- Since $D(\int_0^1 h d\beta) = D(W(h)) = h$,

$$DF = \sum_i \partial_i f(W(h_1), \ldots, W(h_n)) D(W(h_i)),$$

which is the germ of a chain-rule-formula.
- Here is a product rule, for $F, G \in \mathcal{S}_2$

$$D(FG) = FDG + GDF. \tag{1.1}$$

(Just check it for monomials, $F = W(H)^n, G = W(g)^m$.) See [Nualart], p34 for an extension.

- As f has only polynomial growth, we have $DF \in L^p(\Omega, H)$ i.e. $\int_\Omega \|DF(\omega)\|_H^p dW < \infty$. For $p = 2$, this can be expressed simpler, $DF \in L^2([0,1] \times \Omega)$, (after fixing a version) $DF = DF(t, \omega)$ can be thought of a stochastic process.

1.5 Extending the Malliavin Derivative D

So far we have

$$D: \ L^p(\Omega) \supset \mathcal{S}_2 \to L^p(\Omega, H).$$

It is instructive to compare this to the following well-known situation in (Sobolev-)analysis. Take $f \in L^p(U)$, some domain $U \subset \mathbb{R}^n$. Then the *gradient*-operator $\nabla = (\partial_i)_{i=1,\ldots,n}$ maps an appropriate subset of $L^p(U)$ into $L^p(U, \mathbb{R}^n)$. The $\mathbb{R}^n$ comes clearly into play as it is (isomorphic to) the tangent space at any

point of U.
Going back to the Wiener-space we see that H (or, equivalently, $\tilde{H}$) plays the role of the tangent space to the Wiener spaces. [1]

Again, ∇ on $L^P(U)$. What is its natural domain? The best you can do is $(\nabla, W^{1,p})$, which is a closed operator, while (∇, C_c^1) (for instance) is a *closable operator*. This closability (see [RR]) is exactly what you need to extend to operator to the closure of C_c^1 with respect to $\|\cdot\|_{W^{1,p}}$ where

$$\|f\|_{W^{1,p}}^p = \int_U |f|^p d\lambda^n + \sum_{i=1}^n \int_U |\partial_i f|^p d\lambda^n$$

or equivalently

$$\int_U |f|^p d\lambda^n + \int_U \|\nabla f\|_{\mathbb{R}^n}^p d\lambda^n.$$

Using an Integration-by-Parts formula (see the following section on IBP), $(D, \mathcal{S}_2)$ is easily seen to be closable (details found in [Nualart] p26 or [Uestuenel]). The extended domain is denoted with $\mathcal{D}^{1,p}$ and is exactly the closure of $\mathcal{S}_2$ with respect to $\|\cdot\|_{1,p}$ where

$$\begin{aligned}\|F\|_{1,p}^p &= \int_\Omega |F|^p dW + \int_\Omega \|DF\|_H^p dW \\ &= \mathbb{E}|F|^p + \mathbb{E}\|DF\|_H^p.\end{aligned}$$

Remarks: - Look up the definition of closed operator and compare Bass' way to introduce the Malliavin derivative ([**?**, **?**, Bass] p193) with the classical result in [Stein] p122.

- For simplicity take $p = 2$ and consider $F = f(W(h_1), \ldots, W(h_n))$ with h_i's in H. As mentioned in section 1.2, there is n.l.o.g. by assuming the h_i's to be orthonormal.

Then

$$\|DF\|_H^2 = \sum_{i=1}^n (\partial_i f(n \text{ iid std gaussians}))^2.$$

and $\|F\|_{1,2}^2$ simply becomes

$$\int_{\mathbb{R}^n} f^2 d\nu^n + \int_{\mathbb{R}^n} \|\nabla f\|_{\mathbb{R}^n}^2 d\nu^n$$

which is just the norm on the *weighted Sobolev-space* $W^{1,p}(\nu^n)$. More on this link between $\mathcal{D}^{1,p}$ and finite-dimensional Sobolev-spaces is to be found in [Malliavin1] and [Nualart].

- A frequent characterization of Sobolev-spaces on $\mathbb{R}^n$ is via Fourier transform (see, for instance, [Evans] p 282). Let $f \in L^2 = L^2(\mathbb{R}^n)$, then

$$f \in H^k \text{ iff } (1 + |x|^k)\hat{f} \in L^2.$$

[1] We follow Malliavin himself and also Nualart by defining DF as H-valued r.v.. This seems the simplest choice in view of the calculus to come. Oksendal, Uestuenel and Hsu define it as $\tilde{H}$-valued r.v. As commented in Section 1.3 the difference is purely notational since there is a natural isomorphism between H and $\tilde{H}$. For instance, we can write $D(\int_0^1 h d\beta) = h$ while the $\tilde{H}$ choice leads to ($\tilde{H}$-derivative of)$(\int_0^1 h d\beta = \int_0^\cdot h d\lambda$.

Moreover,
$$\|f\|_{H^k} \sim \|(1+|x|^k)\hat{f}\|_{L^2}.$$
In particular, this allows a natural definition of $H^s(\mathbb{R}^n)$ for all $s \in \mathbb{R}$. For later reference, we consider the case $k=1$. Furthermore, for simplicity $n=1$. Recall that $i\partial$ is a self-adjoint operator on $(L^2(\mathbb{R}), <\cdot,\cdot>)$

$$\begin{aligned} <(1+x)\hat{f},(1+x)\hat{f}> &= <(1+i\partial)f,(1+i\partial)f> \\ &= <f,f>+<i\partial f, i\partial f> \\ &= <f,f>+<f,-\partial^2 f> \\ &= <f,(1+A)f>, \end{aligned}$$

where A denotes the negative second derivative. In Section 1.10 this will be linked to the usual definition of Sobolev-spaces (as seen at the beginning of this section), both on $(\mathbb{R}^n, \lambda^n)$ as on (Ω, W).
- The preceeding discussion about how to obtain the optimal domain for the gradient on $(\mathbb{R}^n, \lambda^n)$ is rarely an issue in practical exposures of Sobolev Theory on $\mathbb{R}^n$. The reason is, of course, that we can take weak derivatives resp. distributional derivatives. As well known, Sobolev-spaces can then be defined as those L^p-functions whose weak derivatives are again in L^p. A priori, this can't be done on the Wiener-spaces (at this stage, what are the smooth test-functions there?).

1.6 Integration by Parts

As motivation, we look at $(\mathbb{R}, \lambda)$ first. Take f smooth with compact support (for instance), then, by the translation invariance of Lebesgue-measure,
$$\int f(x+h)d\lambda = \int f(x)d\lambda$$
and hence, after dividing by h and $h \to 0$,
$$\int f' d\lambda = 0.$$
Replacing f by $f \cdot g$ this reads
$$\int f' g d\lambda = -\int f g' d\lambda.$$
The point is that IBP is the infinitesimal expression of a measure-invariance. Things are simple here because λ^n is translation invariant, $(\tau_h)_*\lambda^n = \lambda^n$. Let's look at $(\mathbb{R}^n, \nu^n)$. It is elementary to check that for any $h \in \mathbb{R}^n$
$$\frac{d(\tau_h)_*\nu^n}{d\nu^n}(x) = \exp(\sum_{i=1}^n h_i x_i - \frac{1}{2}\sum_{i=1}^n h_i^2).$$
The corresponding fact on the Wiener space (Ω, W) is the Cameron-Martin theorem. For $\tilde{h} \in \tilde{H} \subset \Omega$ and with $\tau_{\tilde{h}}(\omega) = \omega + \int_0^\cdot h = \omega + \tilde{h}$
$$\frac{d(\tau_{\tilde{h}})_*W}{dW}(\omega) = \exp(\int_0^1 h d\beta(\omega) - \frac{1}{2}\int_0^1 h^2 d\lambda).$$

Theorem 1 (IBP on the Wiener Space) *Let $h \in H$, $F \in S_2$. Then*

$$\mathbb{E}(< DF, h >_H) = \mathbb{E}(F \int_0^1 h d\beta).$$

Proof: (1st variant following [Nualart]) By homogeneity, w.l.o.g. $\|h\| = 1$. Furthermore, we can find f such that $F = f(W(h_1), \ldots, W(h_n))$, with (h_i) orthonormal in H and $h = h_1$. Then, using classical IBP

$$\begin{aligned} \mathbb{E} < DF, h > &= \mathbb{E} \sum_i \partial_i f < h_i, h > \\ &= \int_{\mathbb{R}^n} \partial_1 f(x) (2\pi)^{-n/2} e^{-|x|^2/2} dx \\ &= -\int_{\mathbb{R}^n} f(x) (2\pi)^{-n/2} e^{-|x|^2/2} (-x_1) dx \\ &= \int_{\mathbb{R}^n} f(x) \cdot x_1 d\nu^n \\ &= \mathbb{E}(F \cdot W(h_1)) \\ &= \mathbb{E}(F \cdot W(h)) \end{aligned}$$

(2nd variant following an idea of Bismut, see [Bass]) We already saw in section 1.4 that for $F \in \mathcal{S}_1$ the directional derivative in direction $\tilde{h}$ exists and coincides with $< DF, h >$. For such F

$$\begin{aligned} \int_\Omega F(\omega) dW(\omega) &= \int_\Omega F(\tau_{-\tilde{h}}(\omega) + \tilde{h}) dW(\omega) \\ &= \int_\Omega F(\omega + \tilde{h}) d(\tau_{-\tilde{h}})_* W(\omega) \\ &= \int_\Omega F(\omega + \int_0^\cdot h d\lambda) \exp(-\int_0^1 h d\beta + \frac{1}{2} \int_0^1 h^2 d\lambda) dW(\omega), \end{aligned}$$

using Girsanov's theorem. Replace h by ϵh and observe that the l.h.s. is independent of ϵ. At least formally, when exchanging integration over Ω and $\frac{d}{d\epsilon}$ at $\epsilon = 0$, we find

$$\int_\Omega \big(< DF, h > - F(\omega) \int_0^1 h d\beta \big) dW(\omega) = 0$$

as required. To make this rigorous, approximate F by F's which are, together with $\|DF\|_H$ bounded on Ω. Another approximation leads to $\mathcal{S}_2$-type functionals. □

Remarks: - IBP on the Wiener spaces is one of the cornerstones of the Malliavin Calculus. The second variant of the proof inspired the name *Stochastic Calculus of Variations*: Wiener-paths ω are perturbed by paths $\tilde{h}(.)$. Stochastic Calculus of Variations has (well, a priori) nothing to do with classical calculus of variations.

- As before we can apply this result to a product FG, where both $F, G \in \mathcal{S}_2$. This yields

$$\mathbb{E}(G < DF, h >) = \mathbb{E}(-F < DG, h > + FGW(h)). \tag{1.2}$$

1.7 Itô representation formula / Clark-Ocone-Haussmann formula

As already mentioned $DF \in L^2([0,1]\times\Omega)$ can be thought of a stochastic process. Is it adapted? Let's see. Set

$$F(s) := \mathcal{E}_s(h) := \exp(\int_0^s h d\beta - \frac{1}{2}\int_0^s h^2 d\lambda),$$

an exponential martingale. $F := \mathcal{E}(h) := F(1)$ is not quite in $\mathcal{S}_2$ but easily seen to been in $\mathcal{D}^{1,p}$ and, at least formally and $d\lambda(t)$ -a.s.

$$\begin{aligned} D_t F &= e^{-\frac{1}{2}\int_0^1 h^2 d\lambda} D_t(\exp \int_0^1 h d\beta) \\ &= e^{-\frac{1}{2}\int_0^1 h^2 d\lambda} \exp\big(\int_0^1 h d\beta\big) h(t) \\ &= F h(t) \end{aligned}$$

The used chain-rule is made rigorous by approximation in $\mathcal{S}_2$ using the partial sums of the exponential.
As F contains information up to time 1, $D_t F$ is not adapted to $\mathcal{F}_t$ but we can always project down

$$\begin{aligned} \mathbb{E}(D_t F|\mathcal{F}_t) &= \mathbb{E}(F(1)h(t)|\mathcal{F}_t) \\ &= h(t)\mathbb{E}(F(1)|\mathcal{F}_t) \\ &= h(t)F(t), \end{aligned}$$

using the martingale-property of $F(t)$. On the other hand, F solves the SDE

$$dF(t) = h(t)F(t)d\beta(t)$$

with $F(0) = 1 = \mathbb{E}(F)$. Hence

$$\begin{aligned} F &= \mathbb{E}(F) + \int_0^1 h(t)F(t)d\beta(t) \\ &= \mathbb{E}(F) + \int_0^1 \mathbb{E}(D_t F|\mathcal{F}_t)d\beta(t) \end{aligned} \quad (1.3)$$

By throwing away some information this reads

$$F = \mathbb{E}(F) + \int_0^1 \phi(t,\omega)d\beta \quad (1.4)$$

for some adapted process ϕ in $L^2([0,1]\times\Omega)$. We proved (1.4) for F of the form $\mathcal{E}(h)$, sometimes called *Wick-exponentials*, call $\mathcal{E}$ the set of all such Fs. Obviously this extends the the linear span $(\mathcal{E})$ and by a density argument, see **??** for instance, to any $F \in L^2(\Omega, W)$. This is the **Itô representation theorem**. Looking back to (1.3), we can't expect this to hold for any $F \in L^2(\Omega, W)$ since D is only defined on the proper subset $\mathcal{D}^{1,2}$. However, it is true for $F \in \mathcal{D}^{1,2}$, this is the **Clark-Ocone-Haussmann formula**.

Remarks: - In most books, for instance [Nualart], the proof uses the Wiener-Ito-Chaos-decomposition, although approximation via the span($\mathcal{E}$) should work.
- A similar type of computations allows to compute, at least for $F \in \text{span}(\mathcal{E})$, [2] and $d\lambda(t) \times dW(\omega)$ a.s.

$$D_t \mathbb{E}(F|\mathcal{F}_s) = \mathbb{E}(D_t F|\mathcal{F}_s) 1_{[0,s]}(t).$$

In particular,

$$F \text{ is } \mathcal{F}_s - adapted \quad \Rightarrow \quad D_t F = 0 \text{ for Lebesgue-a.e.} t > s. \tag{1.5}$$

The intuition here is very clear: if F only depends on the early parts of the paths up to time s, i.e. on $\{\omega(s') : s' \leq s\}$, perturbing the pathes later on (i.e. on $t > s$) shouldn't change a thing. Now recall the interpretation of $< DF, h >= \int D_t F h(t) dt$ as directional derivatives in direction of the perturbation $\tilde{h} = \int_0^{\cdot} h d\lambda$.

- Comparing (1.4) and (1.3), the question arises what really happens for $F \in L^2 - \mathcal{D}^{1,2}$. There is an extension of D to $\mathcal{D}'$, the space of *Meyer-Watanabe-distribution* built on the space $\mathcal{D}^\infty$ (introduced a little bit later in this text), and $L^2 \subset \mathcal{D}'$. In this context, (1.3) makes sense for all $F \in L^2$, see [Uestuenel], p42.

1.8 Higher derivatives

When $f : \mathbb{R}^n \supset U \to \mathbb{R}$ then $\nabla f = (\partial_i f)$ is a vectorfield on U, meaning that at each point $\nabla f(x) \in T_x U \cong \mathbb{R}^n$ with standard differential-geometry notation. Then $(\partial_{ij}) f$ is a (symmetric) 2-tensorfield, i.e. at each point an element of $T_{.}U \otimes T_{.}U \cong \mathbb{R}^n \otimes \mathbb{R}^n$. As seen in section 1.5 the tangent space of Ω corresponds to H, therefore $D^2 F$ (still to be defined!) should be a $H \otimes H$-valued r.v. (or $H \hat{\otimes} H$ to indicate symmetry). No need to worry about tensor-calculus in infinite dimension since $H \otimes H \cong L^2([0,1]^2)$. For $F \in \mathcal{S}_2$ (for instance), randomness fixed

$$D^2_{s,t} F := D_s(D_t F)$$

is $d\lambda^2(s,t)$-a.s. well-defined, i.e. good enough to define an element of $L^2([0,1]^2)$. Again, there is closability of the operator $D^2 : L^p(W) \to L^p(W, H \hat{\otimes} H)$ to check, leading to a maximal domain $\mathcal{D}^{2,p}$ with associated norm $\|\cdot\|_{2,p}$ and the same is done for higher derivatives. Details are in [Nualart],p26.
Remarks: - $\mathcal{D}^{k,p}$ is not an algebra but

$$\mathcal{D}^\infty := \cap_{k,p} \mathcal{D}^{k,p}$$

is. As with the class of rapidly decreasing functions, underlying the tempered distributions, $\mathcal{D}^\infty$ can be given a metric and then serve to introduce continuous functionals on it, the *Meyer-Watanabe-distributions*. This is quite a central point in many exposures including [IW], [Oksendal2], [Ocone] and [Uestuenel].
- Standard Sobolev imbedding theorems, as for instance [RR], p215 tell us that for $U = \mathbb{R}^n$

$$W^{k,p}(U) \subset C_b(U)$$

[2] An extension to $\mathcal{D}^{1,2}$ is proved in [Nualart], p32, via WICD.

whenever $kp > \dim U = n$. Now, *very* formally, when $n = \infty$ one could have a function in the intersection of all these Sobolev-spaces without achieving any continuity. And this is what happens on Ω!!! For instance, taking $F = W(h), h \in H$ gives $DF = h, D^2F = 0$, therefore $F \in \mathcal{D}^\infty$. On the other hand, [Nualart] has classified those h for which a continuous choice of $W(h)$ exists, as those L^2-functions that have a representative of bounded variation, see [Nualart] p32 and the references therein.

1.9 The Skorohod Integral / Divergence

For simplicity consider $p = 2$, then

$$D: \ L^2(\Omega) \supset \mathcal{D}^{1,2} \to L^2(\Omega, H),$$

a densely defined unbounded operator. Let δ denote the adjoint operator, i.e. for $u \in \mathrm{Dom}\delta \subset L^2(\Omega, H) \cong L^2([0,1] \times \Omega)$ we require

$$\mathbb{E}(< DF, u >_H) = \mathbb{E}(F\delta(u)).$$

Remark: On $(\mathbb{R}^n, \lambda^n)$,

$$\int_{\mathbb{R}^n} < \nabla f, u >_{\mathbb{R}^n} d\lambda^n = \int_{\mathbb{R}^n} f(-\mathrm{div} u) d\lambda^n, \tag{1.6}$$

this explains (up to a minus-sign) why δ is called *divergence.*

Take $F, G \in \mathcal{S}_2, h \in H$. Then $\delta(Fh)$ is easily computed using the IBP-formula (1.2)

$$\begin{aligned}\mathbb{E}(\delta(Fh)G) &= \mathbb{E}(< Fh, DG >)\\ &= \mathbb{E}(F < h, DG >)\\ &= \mathbb{E}(-G < h, DF >) + \mathbb{E}(FGW(h))\end{aligned}$$

which implies

$$\delta(Fh) = FW(h) - < h, DF > \tag{1.7}$$

Taking $F \equiv 1$ we immediatly get that δ coincides with the Itô-integral on (deterministic) L^2-functions. But we can see much more: take F $\mathcal{F}_r$-measurable, $h = 1_{(r,s]}$. We know from (1.5) that $D_t F = 0$ for a.e. $t > r$. Therefore

$$< h, DF >= \int_0^1 1_{[r,s]}(t) D_t F dt = 0,$$

i.e.

$$\delta(Fh) = FW(h) = F(\beta_s - \beta_r) = \int_0^1 F h d\beta$$

by the very definition of the Itô-integral on adapted step-functions. [3]

By an approximation, for $u \in L^2_a$, the closed subspace of $L^2([0,1] \times \Omega)$ formed by the adapted processes, it still holds that

$$\delta(u) = \int_0^1 u(t) d\beta(t),$$

[3] Also called *simple processes.* See [KS] for definitions and density results.

see [Nualart] p41 or [Uestuenel] p15.
The "divergence" δ is therefore a generalization of the Ito-integral (to non-adapted integrands) and - in this context - called *Skorohod-integral*.

Remark: For $u \in H$, $W(u) = \delta(u)$ (1.7) also reads

$$\delta(Fu) = F\delta(u) - < u, DF > \tag{1.8}$$

and this relation stays true for $u \in \mathrm{Dom}(\delta)$, $F \in \mathcal{D}^{1,2}$ and some integrability condition, see [Nualart], p40. The formal proof is simple, using the product rule

$$\begin{aligned}
\mathbb{E} < Fu, DG > &= \mathbb{E} < u, FDG > \\
&= \mathbb{E} < u, D(FG) - G(DF) > \\
&= \mathbb{E}[(\delta u)FG - < u, DF > G] \\
&= \mathbb{E}[(F\delta u - < u, DF >)G].
\end{aligned}$$

1.10 The OU-operator

We found gradient and divergence on Ω. On $\mathbb{R}^n$ plugging them together yields a positive operator (the negative *Laplacian*)

$$A = -\Delta = -\mathrm{div} \circ \nabla.$$

Here is an application. Again, we are on $(\mathbb{R}^n, \lambda^n)$, $< ., . >$ denotes the inner product on $L^2(\mathbb{R}^n)$.

$$\begin{aligned}
\|f\|^2_{W^{1,2}} = \|f\|^2_{H^1} &= \int |f|^2 d\lambda^n + \int \|\nabla f\|^2_{\mathbb{R}^n} d\lambda^n \\
&= \int |f|^2 d\lambda^n + \int fAf d\lambda^n \quad \text{using (1.6)} \\
&= < f, f > + < Af, f > \\
&= < (1 + A)f, f > \\
&= < (1 + A)^{1/2} f, (1 + A)^{1/2} f > = \|(1 + A)^{1/2} f\|^2,
\end{aligned}$$

using the sqareroot of the positive operator $(1 + A)$ as defined for instance by spectal calculus. For $p \neq 2$ there is no equality but one still has

$$\| \cdot \|_{W^{1,p}} \sim \|(1 + A)^{1/2} \cdot \|_{L^p},$$

when $p > 1$, see [Stein] p135.

Let's do the same on (Ω, W), first define the *Ornstein-Uhlenbeck* operator

$$L := \delta \circ D.$$

Then the same is true, i.e. for $1 < p < \infty$.

$$\| \cdot \|_{1,p} \sim \|(1 + L)^{1/2} \cdot \|_{L^p(\Omega)},$$

with equality for $p = 2$, the latter case is seen as before. This result is a corrolary from the *Meyer Inequalities*. The proof is not easy and found in [Uestuenel] p19, [Nualart] p61 or [Sugita] p37.

How does L act on a $\mathcal{S}_2$-type functional $F = f(W(h_1), \dots, W(h_n))$ where we take w.l.o.g. the h_i's orthonormal? Using $DF = \sum(\partial_i f)h_i$ and formula (1.7) we get

$$\begin{aligned} LF &= \sum_i \partial_i f W(h_i) - \sum_i < D\partial_i f, h_i > \\ &= \sum_i \partial_i f W(h_i) - \sum_{i,j} \partial_{ij} f < h_j, h_i > \\ &= (L^{(n)} f)(W(h_1), \dots, W(h_n)) \end{aligned}$$

where $L^{(n)}$ is defined as the operator for functions on $\mathbb{R}^n$

$$\begin{aligned} L^{(n)} &:= \sum_{i=1}^n [x_i \partial_i - \partial_{ii}] \\ &= x \cdot \nabla - \Delta. \end{aligned}$$

Remarks: - Minus $L^{(n)}$ is the generator of the n-dimensional OU-process given by the SDE

$$dx = \sqrt{2} d\beta - x dt.$$

with explicit solution

$$x(t) = x_0 e^{-t} + \sqrt{2} e^{-t} \int_0^t e^s d\beta(s) \tag{1.9}$$

and for t fixed $x(t)$ has law $\mathcal{N}(x_0 e^{-t}, (1 - e^{-2t})\mathrm{Id})$.

- L plays the role the same role on (Ω, W) as $L^{(n)}$ on $(\mathbb{R}^n, \nu^n)$ or $A = -\Delta$ on $(\mathbb{R}^n, \lambda^n)$.

- Here is some "OU-calculus", (at least for) $F, G \in \mathcal{S}_2$

$$L(FG) = F\ LG + G\ LF - 2 < DF, DG >, \tag{1.10}$$

as immediatly seen by (1.1) and (1.8).

- Some more of that kind,

$$\delta(FDG) = F\ LG - < DF, DG > . \tag{1.11}$$

1.11 The OU-semigroup

We first give a result from semigroup-theory.

Theorem 2 *Let $\mathcal{H}$ be a Hilbert-space, $B : \mathcal{H} \to \mathcal{H}$ be a (possibly unbounded, densely defined) positive, self-adjoint operator. Then $-B$ is the infinitesimal generator of a strongly continuous semigroup of contractions on $\mathcal{H}$.*

Proof: $(-B) = (-B)^*$ and positivity implies that $-B$ is *dissipative* in Pazy's terminology. Now use Corollary 4.4 in [Pazy], page 15 (which is derived from the *Lumer-Phillips* theorem, which is, itself, based on the *Hille-Yoshida* theorem). □

Applying this to A yields the heat-semigroup on $L^2(\mathbb{R}^n, \lambda^n)$, applying it to $L^{(n)}$ yields the OU-semigroup on $L^2(\mathbb{R}^n, \nu^n)$, and for L we get the OU-semigroup on $L^2(\Omega, W)$.
Let's look at the OU-semigroup $P_t^{(n)}$ with generator $L^{(n)}$. Take $f : \mathbb{R}^n \to \mathbb{R}$, say, smooth with compact support. Then it is well-known that, using (1.9),

$$\begin{aligned}(P_t^{(n)} f)(x) &= \mathbb{E}^x f(x(t)) = \\ &= \int_{\mathbb{R}^n} f(e^{-t}x + \sqrt{1-e^{-2t}}y) d\nu^n(y).\end{aligned}$$

is again a continuous function in x. (This property is summarized by saying that $P_t^{(n)}$ is a *Feller-semigroup.*) Similarly, whenever $F : \Omega \to \mathbb{R}$ is nice enough we can set

$$\begin{aligned}(P_t F)(x) &= \int_\Omega F(e^{-t}x + \sqrt{1-e^{-2t}}y) dW(y) \qquad (1.12)\\ &= \int_\Omega F(x\cos\phi + y\sin\phi) dW(y)\end{aligned}$$

A priori, this is not well-defined for $F \in L^p(\Omega)$ since two W-a.s. identical F's could lead to different results. However, this does not happen:

Proposition 3 *Let $1 \le p < \infty$. Then P_t is a well-defined (bounded) operator from $L^p(\Omega) \to L^p(\Omega)$ (with norm ≤ 1).*

Proof: Using Jensen and the rotational invariance of Wiener-measure, with $R(x, y) = (x\cos\phi + y\sin\phi, -x\sin\phi + y\cos\phi)$, we have

$$\begin{aligned}\|P_t F\|^p_{L^p(\Omega)} &= \int_\Omega \big[\int_\Omega F(x\cos\phi + y\sin\phi) dW(y)\big]^p dW(x)\\ &\le \int_{\Omega\times\Omega} \big[|F \otimes 1|(R(x, y))\big]^p d(W \otimes W)(x, y)\\ &= \int_{\Omega\times\Omega} \big[|F \otimes 1|(x, y)\big]^p d(W \otimes W)(x, y)\\ &= \int_{\Omega\times\Omega} |F(x)|^p d(W \otimes W)(x, y) = \int_\Omega |F(x)|^p dW(x) = \|F\|^p_{L^p(\Omega)}.\end{aligned}$$

□

It can be checked that P_t as defined via (1.12) and considered as operator on $L^2(\Omega)$, coincides with the abstract semigroup provided by the theorem at the beginning of this section. It suffices to check that P_t is a semigroup with infinitesimal generator L, the OU-operator, see [Uestuenel] p17.
Remark: P_t is actually more than just a contraction on L^p, it is *hypercontractive* meaning that it increases the degree of integrability, see also [Uestuenel].

1.12 Some calculus on $(\mathbb{R}, \nu)$

From section 1.10,

$$(\mathcal{L}f)(x) := (L^{(1)})f(x) = xf'(x) - f''(x).$$

Following in notation [Malliavin1], [Malliavin2], denote by $\partial f = f'$ the differentiation-operator and by ∂^* the adjoint operator on $L^2(\nu)$. By standard IBP

$$(\partial^* f)(x) = -f'(x) + xf(x).$$

Note that $\mathcal{L} = \partial^* \partial$. Define the *Hermite polynomials* by

$$H_0(x) = 1,\ H_n = \partial^* H_{n-1} = (\partial^*)^n 1.$$

Using the commutation relation $\partial\partial^* - \partial^*\partial = \mathrm{Id}$, an induction (one-line) proof yields $\partial H_n = nH_{n-1}$. An immediate consequence is

$$\mathcal{L} H_n = nH_n.$$

Since H_n is a polynomial of degree n, $\partial^m H_n = 0$ when $m > n$, therefore

$$\begin{aligned} < H_n, H_m >_{L^2(\nu)} &= < H_n, (\partial^*)^m 1 > \\ &= < (\partial)^m H_n, 1 > \\ &= 0 \end{aligned}$$

On the other hand, since $\partial^n H_n = n!$

$$< H_n, H_n >_{L^2(\nu)} = n!,$$

hence $\left\{ \frac{1}{(n!)^{1/2}} H_n \right\}$ is a orthonormal system which is known to be complete, see [Malliavin2] p7. Hence, given $f \in L^2(\nu)$ we have

$$f = \sum c_n H_n \quad \text{with} \quad c_n = \frac{1}{n!} < f, H_n > .$$

Assume that all derivatives are in L^2, too. Then

$$< f, H_n > = < f, \partial^* H_{n-1} > = < \partial f, H_{n-1} > = \ldots = < \partial^n f, 1 > .$$

Denote this projection on 1 by $E(\partial^n f)$ and observe that it equals $\mathbb{E}((\partial^n f)(X))$ for a std. gaussian X. We have

$$f = \sum_{n=0}^{\infty} \frac{1}{n!} E(\partial^n f) H_n. \tag{1.13}$$

Apply this to $f_t(x) = \exp(tx - t^2/2)$ where t is a fixed parameter. Noting $\partial^n f_t = t^n f_t$ and $E(\partial^n f_t) = t^n$ we get

$$\exp(tx - t^2/2) = \sum_{n=0}^{\infty} \frac{t^n}{n!} H_n(x).$$

Remark: [Malliavin1], [Malliavin2] extend ∂, ∂^*, $\mathcal{L}$ in a straightforward-manner to $(\mathbb{R}^{\mathbb{N}}, \nu^{\mathbb{N}})$ which is, in some sense, (Ω, W) with a fixed ONB in H.

1.13 Iterated Wiener-Ito integrals

There is a close link between Hermite polynomials and iterated Wiener-Ito integrals of the form

$$J_n(f) := \int_{\Delta_n} f d\beta^{\otimes n} := \int_0^1 \dots \int_0^{t_1} f(t_1, \dots t_n) d\beta_{t_1} ... d\beta_{t_n},$$

(well-defined) for $f \in L^2(\Delta_n)$ where $\Delta_n := \Delta_n(1) := \{0 < t_1 < \dots... < t_n < 1\} \subset [0,1]^n$. Note that only integration over such a simplex makes sure that every Ito-integration has an adapted integrand. Note that $J_n(f) \in L^2(\Omega)$. A straight-forward computation using the Ito-isometry shows that for $n \neq m$

$$\mathbb{E}(J_n(f)J_m(g)) = 0$$

while

$$\mathbb{E}(J_n(f)J_n(g)) = < f, g >_{L^2(\Delta_n)} .$$

Proposition 4 *Let $h \in H$ with $\|h\|_H = 1$. Let $h^{\otimes n}$ be the n-fold product, a (symmetric) element of $L^2([0,1]^n)$ and restrict it to Δ_n. Then*

$$n! J_n(h^{\otimes n}) = H_n(W(h)). \tag{1.14}$$

Proof: Set

$$M_t := \mathcal{E}_t(g) \quad \text{and} \quad N_t := 1 + \sum_{n=1}^{\infty} \int_{\Delta_n(t)} g^{\otimes n} d\beta^{\otimes n}$$

where $g \in H$. By the above orthonormality relations N_t is seen to be in L^2. Moreover, both $Y = M$ resp. N solve the integral equation,

$$Y_t = 1 + \int_0^t Y_s g(s) d\beta_s.$$

By a unicity result for SDEs (OK, it's just Gronwall's Lemma for the L^2-norm of $M_t - N_t$) we see that W-a.s. $M_t = N_t$ Now take $f \in H$ with norm one. Use the above result with $g = \tau f$, $t = 1$

$$\exp(\tau \int_0^1 f d\beta - \frac{1}{2}\tau^2) = 1 + \sum_{n=1}^{\infty} \tau^n J_n(f^{\otimes n}), \tag{1.15}$$

and using the generating function for the Hermite polynomials finishes the proof. □

A simple "geometric" corollary of the preceeding is that for h, g both norm one elements in H,

$$\mathbb{E}(H_n(W(h))H_m(W(g)) = 0$$

if $n \neq m$ and

$$\mathbb{E}(H_n(W(h))H_n(W(g)) = n!(< h, g >_H)^n.$$

Remark: If it were just for this corollary, an elementary and simple proof is contained in [Nualart].

1.14 The Wiener-Ito Chaos Decomposition

Set

$$\mathcal{C}_n := \{J_n(f) : f \in L^2(\Delta_n)\} \quad (n^{th} \text{ Wiener Chaos})$$

a family of closed, orthogonal subspaces in $L^2(\Omega)$.

For $F = \mathcal{E}(h) \in L^2(\Omega)$ we know from the proof of proposition 4 that

$$F = 1 + \sum_{n=1}^{\infty} J_n(h^{\otimes n}) \quad \text{(orthogonal sum).}$$

Less explicitly this is an orthogonal decompostion of the form

$$F = f_0 + \sum_{n=1}^{\infty} J_n(f_n)$$

for some sequence of $f_n \in L^2(\Delta_n)$. Clearly, this extends to span $(\mathcal{E})$, and since this span is dense in $L^2(\Omega)$ this further extends to any $F \in L^2(\Omega)$ which is the same as saying that

$$L^2(\Omega) = \bigoplus_{n=0}^{\infty} \mathcal{C}_n \quad \text{(orthogonal).}$$

when setting $\mathcal{C}_0$ the subspace of the constants. Indeed, assume that is a non-zero element $G \in (\bigoplus \mathcal{C}_n)^{\perp}$, wlog of norm one. But there is a $F \in \text{span}(\mathcal{E}) \subset (\bigoplus \mathcal{C}_n)$ arbitrarily close - contradiction. This result is called the **Wiener-Ito Chaos Decomposition**.

Remarks: - A slightly different description of of the Wiener-Chaos,

$$\begin{aligned} \mathcal{C}_n &= \text{closure of span}\{J_n(h^{\otimes n}) : \|h\|_H = 1\} \\ &= \text{closure of span}\{H_n(W(h)) : \|h\|_H = 1\}. \end{aligned} \tag{1.16}$$

The second equality is clear by (1.14). Denote by $\mathcal{B}_n$ the r.h.s., clearly $\mathcal{B}_n \subset \mathcal{C}_n$. But since span $(\mathcal{E}) \subset \bigoplus \mathcal{B}_n$, taking the closure yields $\bigoplus \mathcal{B}_n = L^2$, hence $\mathcal{B}_n = \mathcal{C}_n$.

We now turn to the spectral decomposition of the OU-operator L

Theorem 5 *Let Π_n denote the orthogonal projection on $\mathcal{C}_n$, then*

$$L = \sum_{n=1}^{\infty} n\Pi_n.$$

Proof: Set $X = W(h), Y = W(k)$ for two norm one elements in H, $a =< h, k >$, $F = H_n(X)$. Then

$$\begin{aligned} \mathbb{E}(LF, H_m(Y) &= \mathbb{E} < DH_n(X), DH_m(Y) > \\ &= \mathbb{E} < nH_{n-1}(X)h, mH_{m-1}(Y)k > \qquad \text{using } H_n' = \partial H_n = nH_n \\ &= nma\mathbb{E}(H_{n-1}(X), H_{m-1}(Y)) \end{aligned}$$

which, see end of last section, is 0 when $n \neq m$ and

$$nma(n-1)!a^{n-1} = nn!a^n = n\mathbb{E}(H_n(X), H_m(Y))$$

otherwise i.e. when $n = m$. By density of the linear span of such $H_n(X)$'s the result follows. $\Box$

Another application of Hermite polynomials is the fine-structure of $\mathcal{C}_n$. Let $\mathbf{p} : \mathbb{N} \to \mathbb{N}_0$ such that $|\mathbf{p}| = \sum_n \mathbf{p}(n) < \infty$. Fix a ONB e_i for H and set

$$H_{\mathbf{p}} := \prod_n H_{\mathbf{p}(n)}(W(e_n)) \tag{1.17}$$

well-defined since $H_0 = 1$ and $p(n) = 0$ but finitely often. Set $\mathbf{p}! = \prod p(n)!$

Proposition 6 *The set*

$$\{\frac{1}{(\mathbf{p}!)^{1/2}} H_{\mathbf{p}} : |\mathbf{p}| = n\}$$

forms a complete orthonormal set for the n^{th} Wiener-chaos $\mathcal{C}_n$.

Note that this proposition is true for any ONB-choice in H.
Proof: Orthonormality is quickly checked with the $\perp$-properties of $H_n(W(h))$ seen before. Next we show that $H_{\mathbf{p}} \in \mathcal{C}_n$. We do induction by N,the number of non-trivial factors in (1.17). for $N = 1$ this is a consequence of (1.14). For $N > 1$, $H_{\mathbf{p}}$ splits up in

$$H_{\mathbf{p}} = H_{\mathbf{q}} \times H_i \quad \text{with} \quad H_i = H_i(W(e_j))$$

some i, j where $H_q \in \mathcal{S}_2$ is a Wiener-polynomial in which $W(e_j)$ does not appear as argument. Randomness fixed, it follows by the orthonormality of the e_i's that

$$DH_{\mathbf{q}} \in e_j^{\perp} \quad \text{hence} \quad DH_{\mathbf{q}} \perp DH_i.$$

By induction hypothesis, $H_{\mathbf{q}} \in \mathcal{C}_{|\mathbf{q}|} = \mathcal{C}_{n-i}$. Hence

$$LH_{\mathbf{q}} = (n-i)H_{\mathbf{q}}$$

using the the spectral decomposition of the OU-operator. By (1.10),

$$\begin{aligned} L(H_{\mathbf{p}}) &= L(H_{\mathbf{q}}H_i) \\ &= H_{\mathbf{q}}LH_i + H_iLH_{\mathbf{q}} - 2 < DH_q, DH_i > \\ &= H_{\mathbf{q}}(iH_i) + H_i(n-i)H_{\mathbf{q}} \\ &= nH_{\mathbf{p}}, \end{aligned}$$

hence $H_{\mathbf{p}} \in \mathcal{C}_n$. Introduce $\tilde{C}_n$, the closure of the span of all $H_{\mathbf{p}}$'s with $|\mathbf{p}| = n$. We saw that $\tilde{C}_n \subset \mathcal{C}_n$ and we want to show equality. To this end, take any $F \in L^2(\Omega, W)$ and set

$$f_k := \mathbb{E}[F|\sigma(W(e_1), \ldots, W(e_k))].$$

By martingale convergence, $f_k \to F$ in L^2. Furthermore

$$f_k = g_k(W(e_1), \ldots, W(e_k))$$

for some $g_k \in L^2(\mathbb{R}^k, \nu^k) = (L^2(\mathbb{R}, \nu))^{\otimes k}$. Since the (simple) Hermite polynomials form an ONB for $L^2(\mathbb{R}, \nu)$ its k-fold tensor product has the ONB

$$\{\frac{1}{(\mathbf{q}!)^{1/2}} \prod_{i=1}^{k} H_{\mathbf{q}(i)}(x_i) : \text{ all multiindices } \mathbf{q} : \{1, \ldots, k\} \to \mathbb{N}\}.$$

Hence

$$f_k \in \bigoplus_{i=0}^{\infty} \tilde{C}_i,$$

Set $f_k^n := \Pi_n f_k$, then we still have $\lim_{k\to\infty} f_k^n = F$, while $f_k^n \in \mathcal{C}_n$ for all k. Therefore $\tilde{C}_n = \mathcal{C}_n$ as claimed. □

Remarks: - Compare this ONB for $\mathcal{C}_n$ with (1.16). Choosing $h = e_1, e_2, \ldots$ in that line will *not* span $\mathcal{C}_n$. The reason is that $(e_i^{\otimes n})_i$ is not a basis for $H^{\hat{\otimes} n}$, the symmetric tensor-product space, whereas $h^{\otimes n}$ for all unit elements is a basis. For instance, look at $n = 2$. A basis is $(e_i^{\otimes 2})_i$ *and* $(e_i \hat{\otimes} e_j)_{i,j}$ and

$$(e_i + e_j)^{\otimes 2} - e_i^{\otimes 2} - e_j^{\otimes 2} = e_i \otimes e_j + e_j \otimes e_i,$$

the last expression equals (up to a constant) $e_i \hat{\otimes} e_j$.
- The link between Hermite-polynomials and iterated Wiener-Ito integrals, can be extended to this setting. For instance,

$$H_{\mathbf{p}} = H_2(W(e_1)) \times H_1(W(e_2)) = (\text{some constant}) \times J_3(e_1 \hat{\otimes} e_1 \hat{\otimes} e_2).$$

There is surprisingly little found in books about this. Of course, it's contained in Ito's original paper [Ito], but even [Oksendal2] p3.4. refers to that paper when it comes down to it.

1.15 The Stroock-Taylor formula

Going back to the WICD, most authors prove it by an iterated application of the Ito-representation theorem, see section 1.7. For instance, [Oksendal2], p1.4 writes this down in detail. Let's do the first step

$$\begin{aligned}
F &= \mathbb{E}F + \int_0^1 \phi_t d\beta_t \\
&= \mathbb{E}F + \int_0^1 \big(\mathbb{E}(\phi_t) + \int_0^t \phi_{s,t} d\beta_s\big) d\beta_t \\
&= \mathbb{E}F + \int_0^1 \mathbb{E}(\phi_t) d\beta_t + \int_{\Delta_2} \phi(s,t,\omega) d\beta_s d\beta_t \\
&= f_0 + J_1(f_1) + \int_{\Delta_2} \phi(s,t,\omega) d\beta_s d\beta_t
\end{aligned}$$

when setting $f_0 = \mathbb{E}(F), f_1 = E(\phi)$. It's not hard to see that $\int_{\Delta_2} \phi(s,t,\omega) d\beta_s d\beta_t$ is orthogonal to $\mathcal{C}_0$ and $\mathcal{C}_1$ (the same proof as for deterministic integrands - it always boils down to the fact that an Ito-integral has mean zero), hence we found the first two f's of the WICD. But we also saw in section 1.7 that $\phi_t = \mathbb{E}(D_t F|\mathcal{F}_t)$, hence

$$f_1(t) = \mathbb{E}(D_t F),$$

$d\lambda(t)$-a.s. and for $F \in \mathcal{D}^{1,2}$. Similarly,

$$f_2(s,t) = \mathbb{E}(D^2_{s,t} F)$$

$d\lambda^2(s,t)$-a.s. and so for "higher" f_n's, provided all necessary Malliavin-derivatives of F exist. We have

Theorem 7 (Stroock-Taylor) *Let $F \in \cap_k \mathcal{D}^{k,2}$, then the following refined WICD holds,*

$$\begin{aligned} F &= \mathbb{E}F + \sum_{n=1}^{\infty} J_n(\mathbb{E}(D^n F)) \\ &= \mathbb{E}F + \sum_{n=1}^{\infty} \frac{1}{n!} I_n(\mathbb{E}(D^n F)) \end{aligned}$$

where

$$I_n(f) := \int_{[0,1]^n} f d\beta^{\otimes n} := n! J_n(f)$$

for any $f \in L^2(\Delta^n)$ (or symmetric $f \in L^2[0,1]^n$), this notation only introduced here because of its current use in other texts.
Example: Consider $F = f(W(h))$ with $\|h\|_H = 1$ a smooth function f which is together with all its derivatives in $L^2(\nu)$. By iteration,

$$D^n F = (\partial^n f)(W(h)) h^{\otimes n},$$

hence

$$\begin{aligned} \mathbb{E}(D^n F) &= h^{\otimes n} \mathbb{E}((\partial^n f)(W(h)) \\ &= h^{\otimes n} E(\partial^n f) \end{aligned}$$

where we use the notation from 1.12,

$$E(f) = \int f d\nu.$$

Then

$$\begin{aligned} J_n(\mathbb{E}(D^n F)) &= E(\partial^n f) J_n(h^{\otimes n}) \\ &= E(\partial^n f) \frac{1}{n!} H_n(W(h)), \end{aligned}$$

and Stroock-Taylor just says

$$f(W(h)) = E(f) + \sum_{n=1}^{\infty} \frac{1}{n!} E(\partial^n f) H_n(W(h))$$

which is, unsurprisingly, just (1.13) evaluated at $W(h)$.

Chapter 2

Smoothness of laws

2.1

Proposition 8 *Let $F = (F_1, ..., F_m)$ be an m-dimensional r.v. Suppose that for all k and all multiindices α with $|\alpha| = k$ there is a constant c_k such that for all $g \in C^k(\mathbb{R}^m)$*

$$|\mathbb{E}[\partial^\alpha g(F)]| \leq c_k \|g\|_\infty. \tag{2.1}$$

Then the law of F has a C^∞ density.

Proof: Let $\mu(dx) = \mathbb{P}(F \in dx)$ and $\hat{\mu}$ its Fourier-transform. Fix $u \in \mathbb{R}^m$ and take $g = \exp(i < u, \cdot >)$. Then, when $|\alpha| = k$,

$$|u^\alpha||\hat{\mu}(u)| = |\mathbb{E}[\partial^\alpha g(F)]| \leq c_k.$$

For any integer l, by choosing the right α's of order l and maximising the l.h.s we see that

$$(\max_{i=1,\ldots,m} |u_i|)^l \|\hat{\mu}(u)| \leq c_l$$

Hence, at infinity, $\hat{\mu}(u)$ decays faster than any polynomial in $|u|$. On the other hand, as $\mathcal{F}$-transform $\hat{\mu}$ is bounded (by one), therefore $\hat{\mu} \in L^1(\mathbb{R}^m)$. By standard Fourier-transform-results we have

$$\mathcal{F}^{-1}(\hat{\mu}) =: f \in C_0(\mathbb{R}^m)$$

and since $\hat{f} = \hat{\mu}$, by uniqueness, $d\mu = f d\lambda^m$. Replacing α by $\alpha + (0, \ldots, 0, l, 0, \ldots, 0)$ we have

$$|u_i|^l |u^\alpha||\hat{f}(u)| \leq c_{k+l}$$

But since $|u^\alpha||\hat{f}(u)| = |\hat{\partial^\alpha f}|$ we conclude as before that $\partial^\alpha f \in C_0$. □

Remark: - Having (2.1) only for $k \leq m+1$ you can still conclude that $\hat{\mu}(u) = O(\frac{1}{|u|^{m+1}})$ and hence in L^1, therefore $d\mu = f d\lambda$ for continuous f. However, as shown in [Malliavin1], having (2.1) only for $k = 1$, i.e. only involving first derivatives, one still has $d\mu = f d\lambda^m$ for some $f \in L^1(\mathbb{R}^m)$.

Now one way to proceed is as follows: for all $i = 1, \ldots, m$ let $F_i \in \mathcal{D}^{1,2}$ (for the moment) and take $g : \mathbb{R}^m \to \mathbb{R}$ as above. By an application of the chain-rule, j fixed,

$$\begin{aligned} < Dg(F), DF_j > &= < \partial_i g(F) DF_i, DF_j > \\ &= \partial_i g(F) < DF_i, DF_j > \end{aligned}$$

Introducing the *Malliavin covariance matrix*

$$\Lambda_{ij} =< DF_i, DF_j > \tag{2.2}$$

and assuming that

$$\Lambda^{-1} \text{ exists } W - a.s. \tag{2.3}$$

this yields a.s.

$$\begin{aligned} \partial_i g(F) &= (\Lambda^{-1})_{ij} < Dg(F), DF_j > \\ &= < Dg(F), (\Lambda^{-1})_{ij} DF_j > \end{aligned}$$

and hence

$$\begin{aligned} \mathbb{E}[\partial_i g(F)] &= \mathbb{E} < Dg(F), (\Lambda^{-1})_{ij} DF_j > \\ &= \mathbb{E}[g(F), \delta((\Lambda^{-1})_{ij} DF_j)] \end{aligned}$$

by definition of the divergence δ while *hoping* that $(\Lambda^{-1})_{ij} DF_j \in \text{Dom}\delta$. In this case we have

$$\mathbb{E}[\partial_i g(F)] \leq \|g\|_\infty \mathbb{E}[\delta((\Lambda^{-1})_{ij} DF_j)]$$

and we can conclude that F has a density w.r.t. Lebesgue measure λ^m. With some additional assumptions this outline is made rigorous: [1]

Theorem 9 *Suppose* $F = (F_1, \ldots, F_m)$, $F_i \in \mathcal{D}^{2,4}$ *and* Λ^{-1} *exists a.s. Then* F *has a density w.r.t. to* λ^m.

Under much stronger assumptions we have the following result.

Theorem 10 *Suppose* $F = (F_1, \ldots, F_m) \in \mathcal{D}^\infty$ *and* $\Lambda^{-1} \in L^p$ *for all* p *then* F *has a* C^∞*-density.*

For reference in the following proof,

$$\begin{aligned} D(g(F)) &= \partial_i g(F) DF^i & (2.4) \\ L(g(F) &= \partial_i g(F) LF^i - \partial_{ij} g(F) \Lambda_{ij} & (2.5) \\ L(FG) &= FLG + GLF - 2 < DF, DG > & (2.6) \end{aligned}$$

the last equation was already seen in (1.10). The middle equation is a simple consequence of the chain-rule (2.4) and (1.8).
Also, D, L and $\mathbb{E}$ are extended compontentwise to vector- or matrix-valued r.v., for instance $< DF, DF >= \Lambda$.
Proof: Since $0 = D(\Lambda\Lambda^{-1}) = L(\Lambda\Lambda^{-1})$ we have

$$D(\Lambda^{-1}) = -\Lambda^{-1}(D\Lambda)\Lambda^{-1}$$

and

$$L(\Lambda^{-1}) = -\Lambda^{-1}(L\Lambda)\Lambda^{-1} - 2 < \Lambda^{-1} D\Lambda, \Lambda^{-1}(D\Lambda)\Lambda^{-1} >$$

Take a (scalar-valued) $Q \in \mathcal{D}^\infty$ (at first reading take $Q = 1$) and a smooth function $g : \mathbb{R}^m \to \mathbb{R}$. Then

$$\begin{aligned} \mathbb{E}[\Lambda^{-1} < DF, D(g \circ F) > Q] &= \mathbb{E}[\Lambda^{-1} < DF, DF > (\nabla g \circ F) Q] \\ &= \mathbb{E}[(\nabla g \circ F) Q]. & (2.7) \end{aligned}$$

[1] [Nualart], p81

We also have

$$L(F(g\circ F)) = F(L(g\circ F)) + (LF)(g\circ F) - 2 < DF, D(g\circ F) > .$$

This and the self-adjointness of L yields

$$\begin{aligned}
\mathbb{E}[\Lambda^{-1} < DF, D(g\circ F) > Q] &= \frac{1}{2}\mathbb{E}[\Lambda^{-1}\{-L(F(g\circ F)) + F(L(g\circ F)) + (LF)(g\circ F)\}Q] \\
&= \frac{1}{2}\mathbb{E}[-F(g\circ F)L(\Lambda^{-1}Q) + (g\circ F)L(\Lambda^{-1}FQ) + (g\circ F)\Lambda^{-1}(LF)Q] \\
&= \mathbb{E}[(g\circ F)R(Q)] \qquad (2.8)
\end{aligned}$$

with a random vector

$$R(Q) = \frac{1}{2}[-FL(\Lambda^{-1}Q) + L(\Lambda^{-1}FQ) + \Lambda^{-1}(LF)Q].$$

From the vector-equality (2.7) = (2.8)

$$\mathbb{E}[(\partial_i g\circ F)Q] = \mathbb{E}[(g\circ F)\{e_i\cdot R(Q)\}],$$

with i^{th} unit-vector e_i. Now the idea is that together with the other assumptions $Q\in\mathcal{D}^\infty$ implies (componentwise) $R(Q)\in\mathcal{D}^\infty$. To see this you start with proposition 3 but then some more information about L and its action on $\mathcal{D}^\infty$ is required. We don't go into details here, but see [Bass] and [IW].
The rest is easy, taking $Q = 1$ yields

$$|\mathbb{E}[\partial_i g\circ F]| \le c_1\|g\|_\infty$$

and the nice thing is that we can simply iterate: taking $Q = e_j\cdot R(1)$ we get

$$\mathbb{E}[\partial_{ji} g\circ F] = \mathbb{E}[(\partial_i g\circ F)(e_j\cdot R(1))] = \mathbb{E}[(g\circ F)e_i\cdot R(e_j\cdot R(1))]$$

and you conclude as before. Obviously we can continue by induction. Hence, by the first proposition of this section we get the desired result. □

Chapter 3

Degenerated Diffusions

3.1 Malliavin Calculus on the d-dimensional Wiener Space

Generalizing the setup of Chapter 1, we call

$$\Omega = C([0,1], \mathbb{R}^d)$$

the d-dimensional Wiener Space. Under the d-dimensional Wiener measure on Ω the coordinate process becomes a d-dimensional Brownian motion, $(\beta^1, \ldots, \beta^d)$. The reproducing kernel space is now

$$H = L^2([0,1], \mathbb{R}^d) = L^2[0,1] \times \ldots \times L^2[0,1] \quad (d \text{ copies}).$$

As in Chapter 1 the Malliavin derivate of a real-valued r.v. X can be considered as a H-valued r.v. Hence we can write

$$DX = (D^1 X, \ldots, D^d X).$$

For a m-dimensional random variable $X = (X^i)$ set

$$DX = (D^j X^i)_{ij},$$

which appears as a $(m \times d)$-matrix of $L^2[0,1]$-valued r.v. The Malliavin covariance matrix, as introduced in Chapter 2, reads

$$\Lambda_{ij} =< DX^i, DX^j >_H = \sum_{k=1}^{d} < D^k X^i, D^k X^j >_{L^2[0,1]},$$

or simply

$$\Lambda =< DX, (DX)^T >_{L^2[0,1]} . \tag{3.1}$$

3.2 The problem

Given vector-fields $A_1, \ldots, A_d, B$ on $\mathbb{R}^m$ consider the SDE

$$dX_t = A_j(X_t) d\beta_t^j + B(X_t) dt \tag{3.2}$$

For some fixed $t > 0$ (and actually $t \leq 1$ due to our choice of Ω) we want to investigate the regularity of the law of $X(t)$, i.e. existence and smoothness of a density with repect to λ^m on $\mathbb{R}^m$. We assume all the coefficients as nice as we need (smooth, bounded, bounded derivatives etc). Indeed, the degeneration we are interested in lies somewhere else: taking all coefficients zero, the law of $X(t)$ is just the Dirac-measure at $X(0) = x$, in particular there doesn't exist a density.

3.3 SDEs and Malliavin Calculus, the 1-dimensional case

For simplicity take $m = d = 1$ and consider

$$X_t = x + \int_0^t a(X_s) d\beta_s + \int_0^t b(X_s) ds. \tag{3.3}$$

Our try is to *assume* [1] that all X_s are in the domain of D and then to bring D under the integrals. To this end recall from section 1.7 that for fixed s and a $\mathcal{F}_s$-measureable r.v. F one has $D_r F = 0$ for λ-a.e. $r > s$.
Let $u(s, \omega)$ be some $\mathcal{F}_s$-adapted process, and let $r \leq t$. Then

$$D_r \int_0^1 u(s) ds = \int_0^t D_r u(s) ds = \int_r^t D_r u(s) ds,$$

the first step can be justified by a R-sum approximation and the closedness of the operator D. The stochastic integral is more interesting, we restrict ourself to a simple adapted process [2] of the form

$$u(t, \omega) = F(\omega)\ h(t)$$

with $h(t) = 1_{(s_1, s_2]}(t)$ and $\mathcal{F}_{s_1}$-measurable F. Again, let $r \leq t$. Then

$$\begin{aligned}
D_r \int_0^t F h(s) d\beta(s) &= D_r \Big[\int_{[0,r)} F h(s) d\beta(s) + \int_{[r,t]} F h(s) d\beta(s) \Big] \\
&= 0 + D_r \int_0^1 F h(s) 1_{[r,t]}(s) d\beta(s) \\
&= D_r \big[F W(h 1_{[r,t]}) \big] \\
&= (D_r F) W(h 1_{[r,t]}) + F h(r) \\
&= \int_0^1 D_r F h 1_{[r,t]}(s) d\beta(s) + u(r) \\
&= u(r) + \int_r^t D_r u(s) d\beta(s) \quad (*)
\end{aligned}$$

Let us comment on this result. First, if it makes you uncomfortable that our only a.s.-welldefined little r pops up in intervals, rewrite the preceding compution in integrated form, i.e. multiply everything with some arbitrary deterministic $L^2[0,1]$-function $k = k(r)$ and integrate r over $[0,1]$. (Hint: interchange

[1] For a proof see [IW], p393.

[2] We already proceeded like this in section 1.9 when computing $\delta(u)$.

integration w.r.t $d\beta_s$ and dr).
Secondly, a few words about $(*)$. The reduction from $\int_0^t$ on the l.h.s. to $\int_r^t$ at the end is easy to understand - see the recall above. Next, taking $t = r + \epsilon$ we can, at least formally, reduce $(*)$ to $u(r)$ alone. Also, the l.h.s. is easily seen to equal $D_r \int_{r-\epsilon}^t$. That is, when operating D_r on $\int_{r-\epsilon}^{r+\epsilon} u d\beta$ we create somehow a Dirac point-mass $\delta_r(s)$. But that is not surprising! Formally, $D_r Y = < Y, \delta_r >$ corresponding to a (non-admissible!) perturbation of ω by a Heaviside-function with jump at r, say $H(\cdot - r)$ with derivative δ_r. Now, *very* formally, we interpret β as Brownian path perturbed in direction $H(\cdot - r.)$ Taking differentials for use in the stochastic integral we find the Dirac mass δ_r appearing.

(A detailed proof is found in [Oksendal2], corollary 5.13.)

Back to our SDE, applying these results to (3.3) we get

$$\begin{aligned} D_r X_t &= a(X_r) + \int_r^t D_r a(X_s) d\beta_s + \int_r^t D_r b(X_s) ds \\ &= a(X_r) + \int_r^t a'(X_s) D_r X(s) d\beta_s + \int_r^t b'(X_s) D_r X(s) ds \end{aligned}$$

Fix r and set $\tilde{X} := D_r X$. We found the (linear!) SDE

$$d\tilde{X}_t = a'(X_t)\tilde{X}_t d\beta_t + b'(X_t)\tilde{X}_t dt, \quad t > r \tag{3.4}$$

with initial condition $\tilde{X}_r = a(X_r)$.

3.4 Stochastic Flow, the 1-dimensional case

A similar situation occurs when investigating the sensitivity of (3.3) w.r.t. the initial condition $X(0) = x$. Set

$$Y(t) = \frac{\partial}{\partial x} X(t).$$

(A nice version of) $X(t, x)$ is called *stochastic flow.*

A formal computions (see [Bass], p30 for a rigorous proof) gives the same SDE

$$dY_t = a'(X_t)Y_t d\beta_t + b'(X_t)Y_t dt, \quad t > 0$$

and clearly $Y(0) = 1$. Matching this with (3.4) yields

$$D_r X(t) = Y(t)Y^{-1}(r)a(X(r)). \tag{3.5}$$

Remark: In the multidimensional setting note that for ω fixed

$$D_r X(t) \in \mathbb{R}^{m \times d}$$

while

$$Y(t) \in \mathbb{R}^{m \times m}.$$

([Bass] actually makes the choice $m = d$ for a simpler exposure.)

3.5 SDE/flows in multidimensional setting

Rewrite (3.2) in coordinates

$$dX^i = A^i_k(X)d\beta^k + B^i(X)dt, \quad i = 1, \dots, m \tag{3.6}$$

with initial condition $X(0) = x = (x^j) \in \mathbb{R}^m$. Set

$$(Y)_{ij} = \partial_j X^i \equiv \frac{\partial}{\partial x^j} X^i.$$

As before (formally)

$$d\partial_j X^i = \partial_l A^i_k \partial_j X^l d\beta^k + \partial_l B^i \partial_j X^l dt$$

To simplify notation, for any vector-field V on $\mathbb{R}^m$, considered as map $\mathbb{R}^m \to \mathbb{R}^m$, we set [3]

$$(\partial V)_{ij} = \partial_j V^i. \tag{3.7}$$

This yields the following $(m \times m)$-matrix SDE

$$\begin{aligned} dY &= \partial A_k(X) Y d\beta^k + \partial B(X) Y dt \\ Y(0) &= I \end{aligned}$$

and there is no ambiguity in this notation. Note that this is (as before) a linear SDE. We will be interested in the inverse $Z := Y^{-1}$. As a motivation, consider the following 1-dimensional ODE

$$dy = f(t)ydt$$

Clearly $z = 1/y$ satisfies

$$dz = -f(t)zdt.$$

We can recover the same simplicity in the multidimensional SDE case by using Strantonovich Calculus, a first-order stochastic calculus.

3.6 Stratonovich Integrals

3.6.1

Let M, N be continuous semimartingales, define [4]

$$\int_0^t M_s \circ dN_s = \int_0^t M_s dN_s + \frac{1}{2} < N, M >_t$$

resp.

$$M_t \circ dN_t = M_t dN_t + \frac{1}{2} d < N, M >_t .$$

The Ito-formula becomes

$$f(M_t) = f(M_0) + \int_0^t f'(M_s) \circ dM_s. \tag{3.8}$$

[3] If you know classical tensor-calculus it is clear that $\partial_j V^i$ corresponds to a matrix where i represent lines and j the columns.

[4] Do not mix up the bracket with the inner product on Hilbert Spaces.

See [Bass]p27 or any modern account on semimartingales for these results. A special case occurs, when M is given by the SDE

$$dM_t = u_t d\beta_t + v_t dt \quad \text{or} \quad dM_t = u_t \circ d\beta_t + \tilde{v}_t dt$$

Then

$$M_t \circ d\beta_t = M_t d\beta_t + \frac{1}{2} u_t dt. \tag{3.9}$$

One could take this as a definition ([Nualart], p21 does this).

3.6.2

Of course there is a multidimensional version of (3.8) (write it down!). For instance, let $V : \mathbb{R}^m \to \mathbb{R}^m$ and X some m-dimensional process, then

$$dV(X) = (\partial V)(X) \circ dX. \tag{3.10}$$

It also implies a first order product rule

$$d(MN) = N \circ dM + M \circ dN$$

where M, N are (real-valued) semi-martingales.

For later use we discuss a slight generalization. Let Y, Z be two matrix-valued semi-martingales (with dimensions such that $Y \cdot Z$ makes sense). Define $d(YZ)$ component-wise. Then

$$d(ZY) = (\circ dZ)Y + Z \circ dY. \tag{3.11}$$

This might look confusing at first glance, but it simply means

$$Z^i_k(t) Y^k_j(t) = Z^i_k(0) Y^k_j(0) + \int_0^t Y^k_j \circ dZ^i_k + \int_0^t Z^i_k \circ dY^k_j.$$

3.6.3

Let M, N, O, P be semimartingales and

$$dP = N dO.$$

Then it is well-known that [5]

$$M dP = M N dO. \tag{3.12}$$

A similar formula, less well-known, holds for Stratonovich differentials. Let

$$dP = N \circ dO$$

then

$$M \circ dP = MN \circ dO. \tag{3.13}$$

[5] [KS], p145.

Proof: The equals $MdP + \frac{1}{2}d < M, P >= M(NdO + \frac{1}{2}d < N, O >) + \frac{1}{2}d < M, P >$ so the only thing to show is

$$Md < N, O > +d < M, P >= d < MN, O > .$$

Now $d < M, P >= N < M, O >$ ([KS], p143). On ther other hand

$$d(MN) = MdN + NdM + d(\text{bounded variation})$$

shows $d < MN, O >= Md < N, O > +N < M, O >$ (since the bracket kills the bounded variation parts) and we are done. □

3.7 Some differential geometry jargon

3.7.1 Covariante derivatives

Given two smooth vectorfields V, W (on $\mathbb{R}^m$) and using (3.7)

$$(\partial V)W = W^j \partial_j V^i \partial_i,$$

where we follow the differential-geometry usage to denote the basis by $(\partial_1, \ldots, \partial_m)$. This simply means that $(\partial V)W$ is a vector whose ith component is $W^j \partial_j V^i$. Also, we recognize directional derivatives (in direction W) on the r.h.s. In Riemannian geometry this is known as the *covariante derivative* [6] of V in direction W. A standard notation is

$$\nabla_W V = W^j \partial_j V^i \partial_i.$$

∇ is called *connection.*

3.7.2 The Lie Bracket

Let V, W be as before. It is common in Differential Geometry that a vector $V(x) = (V^i(x))$ is identified with the first order differential operator

$$V(x) = V^i(x)\partial_i \mid_x .$$

Consider the ODEs on $\mathbb{R}^m$ given by

$$dX = V(X)dt.$$

It is known [7] that there exists (at least locally) an intergral curve. More precisely, for every $x \in \mathbb{R}^n$ there exists some open (time) interval I_x around 0 and a smooth curve $X_x : I(x) \to \mathbb{R}^m$ which satisfies the ODE and the initial condition $X_x(0) = x$. By setting

$$V_t(x) = X_x(t)$$

we obtain a so-called *local 1-parameter group.* For t fixed $V_t(\cdot)$ is a diffeomorphism between appropriate open sets. [Warner] p37 proves all this, including existence, on a general manifold.

[6] On a general Riemannian manifold there is an additional term due to curvature. Cleary, curvature is zero on $\mathbb{R}^m$.

[7] A simple consequence of the standard existence/uniqueness result for ODEs.

Consider a second ODE, say $dY = W(Y)dt$ with local one-parameter group $W_t(\cdot)$. Then, for t small enough everyhting exists and a second order expansion yields

$$W_{-t} \circ V_{-t} \circ W_t \circ V_t(x) \sim t^2.$$

Dividing the l.h.s. by t^2 and letting $t \to 0$ one obtains a limit in $\mathbb{R}^m$ depending on x, say $[V, W](x)$, the so-called *Lie Bracket*. We see that it is measures how two flows lack to commute (infinitesimaly).

Considering $[V, W]$ as first order operator one actually finds

$$[V, W] = V \circ W - W \circ V,$$

where the r.h.s. is to be understood as composition of differential operators. Note that the r.h.s. is indeed a 1^{st} order operator, since $\partial_{ij} = \partial_{ji}$ (when operating on smooth functions as here). We see that the Lie bracket measures how much two flows lack to commute.
It is immediate to check that [8]

$$\begin{aligned} [V, W] &= \nabla_V W - \nabla_W V \\ &= (\partial W)V - (\partial V)W \end{aligned}$$

Generally speaking, whenever there are two vectorfields "mixed together" the Lie bracket is likely to appear.
Example: Let A be a vectorfield. Inspired by section 3.5 consider

$$dX = A(X)dt, \quad X(0) = x$$

and

$$dY = \partial A(X)Y dt, \quad Y(0) = I$$

Consider the matrix-ODE

$$dZ = -Z\partial A(X)dt, \quad Z(0) = I. \tag{3.14}$$

By computing $d(ZY) = (-Z\partial A(X)dt)Y + Z(\partial A(X)Y dt = 0$ we see that Y^{-1} exists for all times and $Z = Y^{-1}$. Without special motivation, but for later use we compute

$$\begin{aligned} d[Z_t V(X_t)] &= (dZ_t)V(X_t) + Z_t dV(X_t) \\ &= \big[- Z\partial A(X_t)V(X_t) + Z_t \partial V(X_t)A(X_t)\big] dt \\ &= Z\big[\partial V(X_t)A(X_t) - \partial A(X_t)V(X_t)\big] dt \\ &= Z[A, V](X_t)dt \end{aligned} \tag{3.15}$$

3.8 Our SDEs in Stratonovich form

Recall

$$\begin{aligned} dX &= A_k(X)d\beta^k + B(X)dt \\ &= A_k(X) \circ d\beta^k + A_0(X)dt \end{aligned} \tag{3.16}$$

[8] In Riemannian geometry, the first equation is known as the *torsion-free*-property of a Riemanniann connection ∇.

with $X(0) = x$. It is easy to check that

$$A_0^i = B^i - \frac{1}{2} A_k^j \partial_j A_k^i, \quad i = 1, \ldots, m.$$

In the notations introduced in the last 2 sections,

$$\begin{aligned} A_0 &= B - \frac{1}{2}(\partial A_k) A_k \\ &= B - \frac{1}{2}\nabla_{A_k} A_k. \end{aligned}$$

With Y defined as as in section 3.5 we obtain

$$\begin{aligned} dY &= \partial A_k(X) Y \circ d\beta^k + \partial A_0(X) Y dt \\ Y(0) &= I \end{aligned}$$

and $Z = Y^{-1}$ exists for all times and satisfies a generalized version of (3.14),

$$\begin{aligned} dZ &= -Z\partial A_k(X) \circ d\beta^k - Z\partial A_0(X) Z dt \\ Z(0) &= I. \end{aligned} \tag{3.17}$$

The proof goes along (3.14) using (3.11): Since we already discussed the deterministic version we restrict to the case where $A_0 \equiv 0$. Then

$$\begin{aligned} d(ZY) &= (\circ dZ) Y + Z \circ dY \\ &= -Z\partial A_k(X) Y \circ d\beta^k + Z\partial A_k(X) Y \circ d\beta^k \\ &= 0 \end{aligned}$$

(References for this and the next section are [Bass] p199-201, [Nualart] p109-p113 and [IW] p393.)

3.9 The Malliavin Covariance Matrix

Define the $(m \times d)$ matrix

$$\sigma = (A_1 | \ldots\ldots | A_d).$$

Then a generalization of (3.5) holds (see [Nualart] p109 for details)

$$\begin{aligned} D_r X(t) &= Y(t) Y^{-1}(r) \sigma(X_r) \\ &= Y(t) Z(r) \sigma(X_r), \end{aligned}$$

and $D_r X(t)$ appears at a (random) $\mathbb{R}^{m \times d}$-matrix as already remarked at the end of section 3.4. Fix t and write $X = X(t)$. From (3.1), the Malliavin covariance matrix equals

$$\begin{aligned} \Lambda = \Lambda_t &= \int_0^1 D_r X (D_r X)^T dr \\ &= Y(t) \Big[\int_0^t Z(r) \sigma(X_r) \sigma^T(X_r) Z^T(r) dr \Big] Y^T(t). \end{aligned} \tag{3.18}$$

3.10 Absolute continuity under Hörmander's condition

We need a generalization of (3.15).

Lemma 11 *Let V be a smooth vector field on $\mathbb{R}^m$. Let X and Z be processes given by the Stratonovich SDEs (3.16) and (3.17). Then*

$$\begin{aligned} d(Z_t V(X_t)) &= Z_t[A_k, V](X_t) \circ d\beta^k + Z_t[A_0, V](X_t)dt \\ &= Z_t[A_k, V](X_t)d\beta^k \qquad (3.19) \\ &\quad + Z_t\Big(\frac{1}{2}[A_k, [A_k, V]] + [A_0, V]]\Big)(X_t)dt. \qquad (3.20) \end{aligned}$$

First observe that the second equality is a simply application of (3.9) and (3.19) with V replaced by $[A_k, V]$. To see the first equality one could just point at (3.15) and argue with "1st order Stratonovich Calculus" . Here is a rigorous **Proof:** Since the deterministic case was already condidered in (3.15) we take w.l.o.g $A_0 \equiv 0$. Using (3.10) and (3.11) we find

$$\begin{aligned} d(ZV(X)) &= (\circ dZ)V(X) + Z \circ dV \\ &= (-Z\partial A_k V) \circ d\beta^k + (Z\partial V A_k) \circ d\beta^k \\ &= Z\big([A_k, V] \mid_X \big) \circ d\beta^k \end{aligned}$$

□

If you don't like the Strantonovich differentials, a (straight forward but quite tedious) compution via standard Ito calculus is given in [Nualart], p113.

Corollary 12 [9] *Let τ be a stopping time and $y \in \mathbb{R}^m$ such that*

$$< Z_t V(X_t), y >_{\mathbb{R}^m} \equiv 0 \quad \textit{for } t \in [0, \tau].$$

Then for $i = 0, 1, \ldots, d$

$$< Z_t[A_i, V](X_t), y >_{\mathbb{R}^m} \equiv 0 \quad \textit{for } t \in [0, \tau].$$

Proof: Let's prove

$$Z_t V(X_t) \equiv 0 \quad \Rightarrow \quad Z_t[A_i, V](X_t) \equiv 0.$$

(The proof of the actual statement goes along the same lines.) First, the assumption implies that

$$Z_t[A_k, V](X_t)d\beta^k + Z_t\Big(\frac{1}{2}[A_k, [A_k, V]] + [A_0, V]]\Big)(X_t)dt \equiv 0 \quad \text{for } t \in [0, \tau].$$

By uniqueness of semimartingale-decomposition into (local) martingale and and bounded variation part we get (always for $t \in [0, \tau]$)

$$Z_t[A_k, V](X_t) \equiv 0 \text{ for } k = 1, \ldots, d$$

and

$$Z_t\Big(\frac{1}{2}[A_k, [A_k, V]] + [A_0, V]]\Big)(X_t)dt \equiv 0$$

[9] Compare [Bell], 75

By iterating this argument on the first relation

$$Z_t[A_k,[A_j,V]](X_t) \equiv 0 \text{ for } k,j = 1,\ldots,d$$

and together with the second relation we find

$$Z_t[A_0,V](X_t) \equiv 0$$

and we are done. $\Box$

In the following the *range* denotes the image $\Lambda(\mathbb{R}^m) \subset \mathbb{R}^m$ of some (random, time-dependent) $m \times m$ - matrix Λ.

Theorem 13 *Recalling $X(0) = x$, for any $t > 0$ it holds that*

$$\text{span}\{A_1|_x,\ldots,A_d|_x,[A_j,A_k]|_x,[[A_j,A_k],A_l]|_x \ ; \quad j,k,l = 0,\ldots,d\}$$
$$\subset \text{range } \Lambda_t \ \text{a.s.}$$

Proof: For all $s \le t$ define

$$R_s = \text{ span } \{Z(r)A_i(X_r) : r \in [0,s], i = 1,\ldots,d\}$$

and

$$R = R(\omega) = \bigcap_{s>0} R_s.$$

We claim that $R_t = \text{range } \Lambda_t$. From (3.18) it follows that

$$\text{range } \Lambda_t = \text{ range } \int_0^t Z(r)\sigma(X_r)\sigma^T(X_r)Z^T(r)ds \tag{3.21}$$

and since, any $r \le t$ fixed, span $\{Z(r)A_i : i = 1,\ldots,d\}$ = range $Z(r)\sigma(X_r) \supseteq$ range $Z(r)\sigma(X_r)\sigma^T(X_r)Z^T(r)$ the inclusion $R_t \supseteq \text{range}\Lambda_t$ is clear. On the other hand take some $v \in \mathbb{R}^m$ orthgonal to range Λ_t. Clearly

$$v^T\Lambda_t v = 0.$$

Using (3.21) we actually have

$$\int_0^t |v^T Z_s\sigma(X_s)|^2_{\mathbb{R}^m} = \sum_{k=1}^d \int_0^t |v^T Z_r A(X_k)|^2 = 0$$

Since every diffusion path $X_k(\omega)$ is continuous we see that the whole integrand is continuous and we deduce that, for all k and $r \le t$,

$$v \perp Z_r A_k(X_r).$$

We showed (Range $\Lambda_t)^\perp \subseteq R_t^\perp$ and hence the claim is proved.

Now, by Blumenthal's 0-1 law there exists a (deterministic) set $\tilde{R}$ such that $\tilde{R} = R(\omega)$ a.s. Suppose that $y \in \tilde{R}^\perp$. Then a.s. there exists a stopping time τ such that $R_s = \tilde{R}$ for $s \in [0,\tau]$. This means that for all $i = 1,\ldots,d$ and for all $s \in [0,\tau]$

$$< Z_s A_i(X_s), y >_{\mathbb{R}^m} = 0.$$

or simply $y \perp Z_s A_i(X_s)$. Moreover, by iterating Corollary 12 we get

$$y \perp Z_s[A_j, A_k],\ Z_s[[A_j, A_k], A_l],\ \ldots$$

for all $s \in [0, \tau]$. Calling S the set appearing in the l.h.s. of (3.21) and using the last result at $s = 0$ shows that $y \in S^{\perp}$. So we showed $S \subseteq \tilde{R}$. On the other hand, it is clear that a.s. $\tilde{R} \subseteq R_t = \text{Range } \Lambda_t$ as we saw earlier. The proof is finished. □

Combing this with Theorem (9) we conclude

Theorem 14 *Let $A_0, \ldots, A_d$ be smooth vector fields (satisfying certain boundedness conditions [10]) on $\mathbb{R}^m$ which satisfy "Hörmander's condition" (H1) that is that [11]*

$$A_1 \mid_x, \ldots, A_d \mid_x, [A_j, A_k] \mid_x, [[A_j, A_k], A_l] \mid_x \ \ldots; \quad j, k, l = 0, \ldots \tag{3.22}$$

span the whole space $\mathbb{R}^m$. Equivalently we can write

$$\text{Lie } \{A_1|_x, \ldots, A_d|_x, [A_1, A_0]|_x, \ldots, [A_d, A_0]|_x\} = \mathbb{R}^m. \tag{3.23}$$

Fix $t > 0$ and let X_t be the solution of the SDE

$$dX_t = \sum_{k=1}^{d} A_k(X_t) \circ d\beta_t + A_0(X_t)dt. \quad X(0) = x.$$

Then the law of $X(t)$, i.e. the measure $\mathbb{P}[X(t) \in dy]$, has a density w.r.t. to Lebesgue-measure on $\mathbb{R}^m$.

3.11 Smoothness under Hoermander's condition

Under essentially the same hypothesis as in the last theorem [12] one actually has a smooth density of $X(t)$, i.e. $\in C^{\infty}(\mathbb{R}^m)$. The idea is clearly to use Theorem 10, but there is some work to do. We refer to [Norris] and [Nualart], [Bass] and [Bell].

3.12 The generator

It is well-know that the generator of a (Markov) process given by the SDE

$$\begin{aligned} dX &= A_k(X) \circ d\beta^k + A_0(X)dt \\ &= A_k(X)d\beta^k + B(X)dt \\ &= \sigma d\beta + B(X)dt. \end{aligned} \tag{3.24}$$

is the second order differential operator

$$\mathcal{L} = \frac{1}{2} E^{ij} \partial_{ij} + B^i \partial_i \tag{3.25}$$

[10] Bounded and bounded derivatives will do - we have to guarantee existence and uniqueness of X and Y as solution of the corresponding SDEs.

[11] Note that $A_0 \mid_x$ *alone* is not contained in the following list while it does appear in all brackets.

[12] One requires that the vectorfields have bounded derivatives of all orders, since higher-order analogues to Y come into play.

with $(m \times m)$ matrix $E = \sigma\sigma^T$ (or $E^{ij} \equiv \sum_{k=1}^d A_k^i A_k^j$. in coordinates). Identifying a vector field, say V, with a first order differntial operator, the expression $V^2 = V \circ V$ makes sense as a second order differential operator. In coordinates,

$$V^i \partial_i (V^j \partial_j) = V^i V^j \partial_{ij} + V^j (\partial_j V^i) \partial_i.$$

Note the last term on the r.h.s is the vector $V\partial V = \nabla_V V$. Replacing V by A_k and summing over all we see that

$$E^{ij}\partial_{ij} = \sum_{k=1}^d A_k^2 - \sum_k \nabla_{A_k} A_k.$$

We recall (see chapter 3.8) that $A_0 = B - \frac{1}{2}\sum_k \nabla_{A_k} A_k$. Hence

$$\mathcal{L} = \frac{1}{2}\sum_{k=1}^d A_k^2 + A_0. \tag{3.26}$$

Besides giving another justification of the Stratonovich calculus, it is important to notice that this "sum-of-square"-form is invariant under coordinate-transformation, hence a suited operator for analysis on manifolds.

3.13

Example 1 (bad): Given two vectorfields on $\mathbb{R}^2$ (in 1st order diff. operator notation)

$$A_1 = x_1\partial_1 + \partial_2, \quad A_2 = \partial_2$$

set

$$\mathcal{L} = \frac{1}{2}(A_1^2 + A_2^2).$$

Expanding,

$$(x_1\partial_1 + \partial_2)^2 = x_1^2\partial_{11} + x_1\partial_1 + 2x_1\partial_{12} + 2\partial_{22},$$

yields

$$\mathcal{L} = \frac{1}{2}E^{ij}\partial_{ij} + b_i\partial_i$$

with

$$E = \begin{pmatrix} x_1^2 & x_1 \\ x_1 & 2 \end{pmatrix}$$

and $B = (x_1, 0)^T$. Now $E = \sigma\sigma^T$ with

$$\sigma = \begin{pmatrix} x_1 & 0 \\ 1 & 1 \end{pmatrix}$$

and the associated diffusion process is

$$\begin{aligned} dX_t &= \sigma(X_t)d\beta_t + B(X_t)dt \\ &= A_1(X_t)d\beta_t^1 + A_2(X_t)d\beta_t^2 + B(X_t)dt. \end{aligned}$$

We see that when we start from the x_2-axis i.e. on $\{x_1 = 0\}$ there is no drift, $B \equiv 0$, and both brownian motions push us around along direction x_2, therefore no chance of ever leaving this axis again. Clearly, in such a situation, the law of $X(t)$ is singular with respect to Lebesgue-measure on $\mathbb{R}^2$.

To check Hörmander's condition H1 compute

$$\begin{aligned}[A_1, A_2] &= (x_1\partial_1 + \partial_2)\partial_2 - \partial_2(x_1\partial_1 + \partial_2) \\ &= 0\end{aligned}$$

therefore the Lie Algebra generated by A_1 and A_2 simply equals the span $\{A_1, A_2\}$ and is *not* the entire of $\mathbb{R}^2$ when evaluated at the degenerated area $\{x_1 = 0\}$ - exactly as expected.

Example 2 (good): Same setting but

$$A_1 = x_2\partial_1 + \partial_2, \quad A_2 = \partial_2.$$

Again,

$$\begin{aligned}\mathcal{L} &= \frac{1}{2}(V_1^2 + V_2^2) \\ &= \frac{1}{2}a_{ij}\partial_{ij} + b_i\partial_i\end{aligned}$$

Similarly we find

$$dX_t = A_1(X_t)d\beta_t^1 + A_2(X_t)d\beta_t^2 + B(X)dt.$$

with drift $B = (1, 0)^T$.

The situation looks similar. On the x_1-axis where $\{x_2 = 0\}$ we have $A_1 = A_2$, therefore diffusion happens in x_2-direction only. However, when we start at $\{x_2 = 0\}$ we are pushed in x_2-direction and hence immediatly leave the degenerated area.

To check Hörmander's condition H1 compute

$$\begin{aligned}[V_1, V_2] &= (x_2\partial_1 + \partial_2)\partial_2 - \partial_2(x_1\partial_1 + \partial_2) \\ &= -\partial_1.\end{aligned}$$

See that

$$\text{span}(V_2, [V_1, V_2]) = \mathbb{R}^2$$

for all points and our Theorem 14 applies.

Example 3 (How many driving BM?):

Consider the $m = 2$-dimensioal process driven by one BM ($d = 1$),

$$\begin{aligned}dX^1 &= d\beta, \\ dX^2 &= X^1 dt.\end{aligned}$$

From this extract $A_1 = \partial_1$, drift $A_0 = x^1\partial_2$ and since $[A_1, A_0] = \partial_2$ Hörmander's condition holds for all points on $\mathbb{R}^2$. Actually, it is an easy exercise to see that (X^1, X^2) is a zero mean Gaussian process with covariance matrix

$$\begin{pmatrix} t & t^2/2 \\ t^2/2 & t^3/3 \end{pmatrix}.$$

Hence we can write down explicitly a density with repsect to 2-dimensional Lebesgue-measure. Generally, one BM together with the right drift is enough for having a density.

Example 3 (Is Hörmander's condition necessary?):

No! Take $f : \mathbb{R} \to \mathbb{R}$ smooth, bounded etc such that $f^{(n)}(0) = 0$ for all $n \geq 0$ (in particular $f(0) = 0$) and look at

$$2\mathcal{L} = \big(\partial_1\big)^2 + \big(f(x_1)\partial_2\big)^2.$$

as arising from $m = d = 2$, $A_1 = \partial_1$, $A_2 = f(x_1)\partial_2$. Check that $A_2, [A_1, A_2], \ldots$ are all 0 when evaluated at $x_1 = 0$ (simply because the Lie-brackets make all derivatives of f appear.) Hence Hörmander's condition is not satisfied when starting from the degenerated region $\{x_1 = 0\}$. On the other hand, due to A_1 we will immediatly leave the degenerate region and hence there is a density (some argument as in example 2).

Chapter 4

Hypelliptic PDEs

4.1

Let $V_0, \ldots, V_d$ be smooth vectorfields on some open $U \subset \mathbb{R}^n$, let c be a smooth function on U. Define the second order differential operator (where c operates by multiplication)

$$\mathcal{G} := \sum_{k=1}^{d} V_k^2 + V_0 + c.$$

Let $f, g \in \mathcal{D}'(U)$, assume

$$\mathcal{G} f = g$$

in the distributional sense, which means (by definition)

$$< f, \mathcal{G}^* \varphi >=< g, \varphi >$$

for all test-functions $\varphi \in D(U)$. We call the operator $\mathcal{G}$ *hypoelliptic* if, forall open $V \subset U$,

$$g\,|_V \in C^\infty(V) \quad \Rightarrow \quad f\,|_V \in C^\infty(V).$$

Hörmander's Theorem, as proved in [Kohn], states:

Theorem 15 *Assume*

$$Lie\ [V_0|_y, \ldots, V_d|_y] = \mathbb{R}^n$$

for all $y \in U$. Then the operator $\mathcal{G}$ as given above is hypoelliptic.

Remark: An example the Hörmander's Theorem is a sufficient condition for hypoellipticity but not a necessary one goes along Example 4 from the last chapter.

4.2

Take X as in section (3.8), take $U = (0, \infty) \times \mathbb{R}^m$ and let $\varphi \in \mathcal{D}(U)$. For T large enough

$$\mathbb{E}[\varphi(T, X_T)] = \mathbb{E}[\varphi(0, X_0)] = 0,$$

hence, by Ito's formula,

$$0 = \mathbb{E}\int_0^T (\partial t + \mathcal{L})\varphi(t, X_t)dt.$$

By Fubini and $T \to \infty$ this implies

$$\begin{aligned} 0 &= \int_0^\infty \int_{\mathbb{R}^m} (\partial t + \mathcal{L})\varphi(t, y)p_t(dy)dt \\ &= \int_0^\infty \int_{\mathbb{R}^m} \psi(t, y)p_t(dy)dt \end{aligned}$$

for $\psi \in \mathcal{D}(U)$ as defined through the last equation. This also reads

$$0 =< \Phi, (\partial t + \mathcal{L})\varphi >=< \Phi, \psi >$$

for some distribution $\Phi \in \mathcal{D}'(U)$. [1] In distributional sense this writes

$$(\partial_t + \mathcal{L})^* \Phi = (-\partial_t + \mathcal{L}^*)\Phi = 0, \tag{4.1}$$

saying that Φ satisfies the *forward Fokker-Planck equation.* If we can guarantee that $-\partial_t + \mathcal{L}^*$ is hypoelliptic then, by Hörmander's theorem there exists $p(t, y)$ smooth in both variables s.t.

$$\begin{aligned} < \Phi, \varphi > &= \int_{(0,\infty)\times\mathbb{R}^m} p(t, y)\varphi(t, y)dtdy \\ &= \int_{(0,\infty)\times\mathbb{R}^m} \varphi(t, y)p_t(dy)dt. \end{aligned}$$

This implies

$$p_t(dy) = p(t, y)dy$$

for $p(t, y)$ smooth on $(0, \infty) \times \mathbb{R}^m$. [2]

4.3

We need sufficient conditions to guarantee the hypoellipticity of $\mathcal{G} = -\partial_t + \mathcal{L}^*$ as operator on $U = (0, \infty) \times \mathbb{R}^m \subset \mathbb{R}^n$ with $n = m + 1$.

Lemma 16 *Given a first order differential operator $V = v^i\partial_i$ its adjoint is given by*

$$V^* = -(V + c_V)$$

where $c_V = \partial_i v^i$ is a scalar-field acting by multiplication.

Proof: Easy. □

As corollary,

$$(V^2)^* = (V^*)^2 = (c_V + V)^2 = V^2 + 2c_V V + c$$

[1] The distribution Φ is also represented by the (finite-on-compacts-) measure given by the *semi-direct product* of the kernel $p(s, dy)$ and Lebesgue-measure $ds = d\lambda(s)$.

[2] Note that the smoothness conclusion via Malliavin calculus doesn't say anything about smoothness in t, i.e. our conclusion is stronger.

for some scalar-field c. For $\mathcal{L}$ as given in (3.26) this implies

$$\mathcal{L}^* = \frac{1}{2}\sum_{k=1}^{d} A_k^2 - (A_0 - c_{A_k} A_k) + c$$

for some (different) scalar-field c. Defining

$$\tilde{A}_0 = A_0 - c_{A_k} A_k \tag{4.2}$$

this reads

$$\mathcal{L}^* = \frac{1}{2}\sum_{k=1}^{d} A_k^2 - \tilde{A}_0 + c.$$

We can trivially extend vectorfields on $\mathbb{R}^m$ to vectorfields on $U = (0,\infty) \times \mathbb{R}^m$ ("time-independent vectorfields"). From the differential-operator point of view it just means that we have that we are acting only on the space-variables and not in t. Then

$$\mathcal{G} = \frac{1}{2}\sum_{k=1}^{d} A_k^2 - (\tilde{A}_0 + \partial_t) + c$$

is an operator on U, in Hörmander form as needed. Define the vector $\hat{A} = \tilde{A}_0 + \partial_t \in \mathbb{R}^n$. [3] Hence Hörmander's (sufficient) condition for $\mathcal{G}$ being hypoellitpic reads

$$\text{Lie } \{A_1|_y, \ldots, A_d|_y, \hat{A}|_y\} = \mathbb{R}^n \tag{4.3}$$

for all $y \in U$. Note that for $k = 1, \ldots, d$

$$[A_k, \partial_t] = A_k^i \partial_i \partial_t - \partial_t A_k^i \partial_i = 0$$

since the A_k^i are functions in space only. It follows that [4]

$$[A_k, \hat{A}] = [A_k, \tilde{A}_0],$$

and similarly no higher bracket will yield any component in t-direction. From this it follows that (4.3) is equivalent to

$$\text{Lie } \{A_1|_y, \ldots, A_d|_y, [A_1, \tilde{A}_0]|_y, \ldots, [A_d, \tilde{A}_0]|_y\} = \mathbb{R}^m \tag{4.4}$$

for all $y \in \mathbb{R}^m$. Using (4.2) we can replace $\tilde{A}_0$ in condition (4.4) by A_0 without changing the spanned Lie-algebra. We summarize

Theorem 17 *Assume that Hörmander's condition (H2) holds:*

$$\textit{Lie } \{A_1|_y, \ldots, A_d|_y, [A_1, A_0]|_y, \ldots, [A_d, A_0]|_y\} = \mathbb{R}^m.) \tag{4.5}$$

for all $y \in \mathbb{R}^m$. Then the law of the process X_t has a density $p(t,y)$ which is smooth on $(0,\infty) \times \mathbb{R}^m$.

[3] As vector, think of having a 1 in the 0th position (time), then use the coordinates from $\tilde{A}_0$ to fill up positions 1 to m.

[4] We abuse notation: the Bracket on the l.h.s. is taken in $\mathbb{R}^n$ resulting in a vector with no component in t-direction which, therefore, is identified with the $\mathbb{R}^m$-vector an the r.h.s., result of the bracket-operation in $\mathbb{R}^m$.

Remarks: - Compare conditions H1 and H2, see (3.23) and (4.5). The only difference is that H2 is required for all points while H1 only needs to hold for $x = X(0)$. (Hence H2 is a stronger condition.)
- Using H1 (ie Malliavin's approach) we don't get (a priori) information about smoothness in t.
- Neither H1 nor H2 allow A_0 (alone!) to help out with the span. The intuitive meaning is clear: A_0 alone represents the drift hence doesn't cause any diffusion which is the origin for a density of the process X_t.
- We identified the distribution Φ as uniquely associated to the (smooth) function $p(t, y) = p(t, y; x)$. Hence from (4.1)

$$\partial_t p = \mathcal{L}^* p \quad (\mathcal{L}^* \text{ acts on } y)$$

and $p(0, dy)$ is the Dirac-measure at x. All that is usually summarized by saying that p is a *fundamental solution* of the above parabolic PDE and our theorem gives smoothness-results for it.
- Let $\sigma = (A_1|\dots|A_d)$ and assume that $E = \sigma\sigma^T$ is uniformly elliptic. We claim that in this case the vectors $\{A_1, \dots, A_d\}$ already span $\mathbb{R}^m$ (at all points), so that Hörmander's condition is always satisfied.
Proof: Assume $v \in \text{span } \{A_1, \dots, A_d\}^\perp$. Then, for all k,

$$0 =< v, A_k >^2 = |v^i A_k^i|^2 = v^i A_k^i A_k^j v^j = v^T E v.$$

Since E is symmetric, positive definite we see that $v = 0$. $\square$

Bibliography

[Bass] Bass *Elliptic operators and diffusion*, Springer, 1997.

[Bell] Bell *The Malliavin Calculus*, Pitman Monographs 34, 1987.

[BH] Bouleau, Hirsch *Dirichlet Forms ...*, Springer, 1997.

[DaPrato] DaPrato, Zabczyk *Stochastic Equations in Infinite Dimensions*, Cambridge University Press, 1992.

[Evans] L.C. Evans *Partial Differential Equations* AMS.

[IW] Ikeda Watanabe *Stochastic Differential Equations and Diffusion Processes*, North-Holland, 1989

[Ito] K. Ito *Multiple Wiener integral*, J.Math.Soc.Japan 3 (1951), 157-169

[KS] Karatzas, Shreve *Brownian Motion and Stochastic Calculus, 2nd Ed*, Springer.

[Kohn] J.J, Kohn *Pseudo-differential Operators and Hypoellipticity* Proc. Symp. Pure Math, 23, A.M.S. (1073) p61-69

[Malliavin1] P. Malliavin *Probability and Integration* Springer.

[Malliavin2] Malliavin *Stochastic Analysis* Springer.

[Norris] Norris *Simplified Malliavin Calculus*

[Nualart] D. Nualart *Malliavin Calculus and Related Topics* Springer.

[Ocone] Ocone *A guide to stochastic calculus of variations* LNM 1316, 1987.

[Oksendal1] B. Oksendal *Stochastic Differential Equations* Springer 1995.

[Oksendal2] B. Oksendal *An Introduction to Malliavin Calculus with Application to Economics* Lecture Notes, 1997, available on www.nhh.no/for/dp/1996/

[Pazy] A. Pazy *Semigroups of Linear Operators and Applications to Partial Differential Equations* Springer, 1983.

[RR] Renardy Rogers *Introduction to Partial Differential Equations*, Springer.

[RU] W. Rudin *Functional Analysis*, McGraw-Hill, 1991.

[RY] Revuz, Yor *Brownian Motion and Continuous Martingales*, Springer.

[Stein] Stein *Singular Integrals* Princeton UP, 1970.

[Sugita] H. Sugita *Sobolev spaces of Wiener functionals and Malliavin's calculus*, J. Math. Kyoto Univ. 25-1 (1985) 31-48

[Uestuenel] A.S. Uestuenel *An Introduction to Analysis on Wiener Space*, Springer LNM 1610.

[Warner] Warner *Foundations of Differentiable Manifolds and Lie Groups*, Springer.

[Williams] Williams *To begin at the beginning.*, Stochastic Inergrals, Lecture Notes in Math. 851 (1981), 1-55