NOTES IN MALLIAVIN-STEIN CALCULUS MINICOURSE

ARTURO JARAMILLO

1. Introduction

These notes are devoted to the study of limit theorems via the Malliavin-Stein method, and are largely based on the exposition found in [1]. Malliavin calculus is, in essence, more of a tool than a self-contained theory, although one could certainly choose to view it as such. What matters most for purposes of this minicourse, is to keep in mind a clear picture of on single, very modest concrete problem to address by using it. In general, our main goal is to study Gaussian distributional limit theorems through the lens of Malliavin calculus. The type of problem we will have in mind throughout has the following structure:

Ingredients:

- A Gaussian process $G = \{G_t\,;\, t \in T\}$, indexed by some set T lurking in the background.
- A random variable that is measurable with respect to G, and which we interpret as a functional of the trajectory of G.

Conditions:

- The random variable of interest depends on a parameter n, and we assume that its behavior becomes interesting as $n \to \infty$. Thus, we consider a sequence Z_n of such random variables.
- After appropriate normalization, we can assume that each Z_n has mean zero and variance one.

Meta-problems:

- Suppose we suspect that the sequence Z_n converges in distribution to a standard Gaussian random variable. What kinds of conditions can we establish to rigorously guarantee this convergence?
- Assuming the above convergence does hold, how can we quantify the rate at which Z_n converges to the Gaussian limit? (More precisely, we will seek to bound this rate rather than describe it exactly, though the latter is sometimes possible.)

The reader might think of the example that feels most familiar or interesting to them. But to make things more concrete, I'll share the example that means the most to me personally: the BreuerMajor theorem.

Date: July 8, 2025.

Before talking about that, let's quickly go over the classic central limit theorem to really understand the motivation. Suppose we have a sequence $X_1, X_2, \dots$ of independent and identically distributed random variables, each with mean zero and finite third moment. Then we know that the sum

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k$$

converges in distribution, as n tends to infinity, to a Gaussian random variable with some variance $\sigma^2 > 0$.

There's also a more precise version called the Berry-Esseen theorem. It gives a bound on how fast this convergence happens, using the Kolmogorov distance:

$$d_K(\mu, \nu) := \sup_{z \in \mathbb{R}} |\mu((-\infty, z]) - \nu((-\infty, z])|.$$

This theorem says theres a universal constant $C > 0$ such that

$$d_K(Z_n/\sigma, N) \leq \frac{C}{\sqrt{n}\sigma^3} \mathbb{E}[|\xi_1|^3],$$

where $Z_n = \sum_{k=1}^{n} \xi_k$, and N is a standard Gaussian. This result might be a bit less well-known than the central limit theorem, but its still pretty common. And it naturally raises a few important questions. For example:

(i) What if the variables aren't identically distributed?
(ii) What if the limit isn't Gaussian?
(iii) What if we don't assume independence?

Charles Stein came up with a method (now known as Stein's method) that gives very clean answers to the first two questions. We'll talk about this more in the second lecture. The nice thing is that the same kind of ideas can be used when the limit is something other than Gaussian, like Poisson, exponential, gamma, beta, Wigner, or even the Dickman distribution. Point number (iii) is by far more delicate, and it is precisely this point that we address in the following example:

The Breuer-Major problem
Lets look at a specific case where the source of randomness is a stationary, standardized Gaussian time series $\xi = \{\xi_k\}_{k \geq 1}$. This means that ξ is a Gaussian process where each ξ_k has mean zero and variance one, and the sequence is stationary in the sense that

$$\mathcal{L}(\xi_1, \xi_2, \dots) = \mathcal{L}(\xi_{1+r}, \xi_{2+r}, \dots)$$

for every integer $r \geq 1$, where $\mathcal{L}$ denotes the law (i.e., the joint distribution) of the process.

In the Gaussian case, the full distribution of ξ is determined by its covariance function. So we define

$$\rho(r) := \mathbb{E}[\xi_1 \xi_{1+r}]$$

as the correlation at lag r. We are particularly interested in situations where $\rho(r)$ stays nonzero for many values of r, meaning the process has long memoryunlike white noise, where $\rho(r) = \delta_{0,r}$. This kind of structured noise appears frequently in real-world applications.

Now, suppose we don't observe ξ_k directly, but instead see a transformed version through a function φ. That is, we define a new sequence $X_k := \varphi(\xi_k)$, where φ is a fixed function chosen to match the needs of the problem. We'll assume that φ satisfies some regularity or integrability conditions, enough to allow the use of limit theorems, while keeping things as flexible as possible.

Recall that for centered Gaussian random vectors, being independent is equivalent to being uncorrelated. So in our setting (where $\rho(r)$ is nonzero for a large set of values of r), we definitely can't treat the sum

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k \tag{1.1}$$

as we would in the classical central limit theorem. But intuitively, if the dependence between variables is weak or decays fast enough, it still seems reasonable to expect that some kind of asymptotic Gaussian behavior could emerge.

This brings us to what I'll call the Breuer-Major problem: under what conditions on the correlation structure ρ and on the function φ can we guarantee that the sum (1.1) converges in distribution to a standard Gaussian variable? Breuer and Major were able to provide surprisingly mild conditions to ensure this, and we'll discuss the precise statement of their theorem in the final lecture. For now, I just want to use this problem as motivation (it is a great example of why the tools we'll study in these notes are worth learning). Actually, this type of problem is a special case of a broader class: the study of functionals of Gaussian processes. If this structure feels familiar, you might be right to suspect your own favorite problem falls into the same general framework. I highly recommend checking out the following website, which presents a wide range of results and developments in this area:

https://sites.google.com/site/malliavinstein/home

The methodology we're about to follow is built around two key components. At first, they might seem a bit abstract, but I promise theyll become very familiar as we go along:

- Stein's method, which gives a way to measure how far the distribution of a random variable X is from the standard Gaussian distribution. It does so by analyzing expressions of the form
$$\mathbb{E}[Xf(X) - f'(X)], \tag{1.2}$$
where f runs over a suitable class of test functions for which f' makes sense.
- The challenge in estimating (1.2) is handled by introducing a new random variable $\Gamma[X]$, which depends on X, and is chosen so that
$$\mathbb{E}[Xf(X) - f'(X)\Gamma[X]] = 0 \tag{1.3}$$

holds for all functions f in our class.

These two ideas together give us a path forward: the quality of the approximation to Gaussianity can be understood by measuring how close $\Gamma[X]$ is to 1.

Of course, I haven't yet said how to construct this $\Gamma[X]$, and that's exactly where Malliavin calculus comes in. Although originally developed for very different purposes, the tools from Malliavin calculus provide a remarkably clean and powerful way to build such a $\Gamma[X]$. It's worth noting that this construction is far from unique, but Malliavin's framework offers a natural and elegant route that fits perfectly with our goals. I invite the reader to bare with the reading of the following sections with the promise that all the language will eventually be utilized for solving the concrete application of the resolution of the Breuer Major problem

2. Malliavin calculus in one dimension

In this section, we introduce the basic tools of Malliavin calculus in the most accessible way possible. Following the approach in [1], we focus on the simplest case: Gaussian noise generated by a single random variable, rather than a full process. This helps us get familiar with the ideas without getting overwhelmed by technicalities.

Throughout this section, all random variables will be defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. As a concrete and manageable starting point, we take $\Omega = \mathbb{R}$, $\mathcal{F}$ as the Borel σ-algebra, and $\mathbb{P}$ as the standard Gaussian measure, which we denote by γ, and which is defined by

$$\gamma(dx) := \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} dx.$$

Just a quick heads-up: later on, we'll use the same symbol γ to refer to the Gaussian distribution in higher dimensions. Hopefully, this won't be too confusing when we get there. The main theme of this section is to describe random variables on Ω using an orthogonal basis, and to explore different ways to interpret these decompositions. The point of working with orthogonal decompositions is that applying expectations becomes much easier.

Let's begin with the basic building block: the Malliavin derivative. In what follows, we denote by $\mathcal{S}$ the set of C^∞-functions $f : \mathbb{R} \to \mathbb{R}$ such that f and all of its derivatives have at most polynomial growth. We call any element of $\mathcal{S}$ a smooth function. For a function $f \in \mathcal{S}$, the Malliavin derivative is just the usual derivative: $Df = f'$. Things get more interesting when we want to differentiate functions outside of $\mathcal{S}$, like

$$f(x) = (x - a)_+,$$

for some $a \in \mathbb{R}$. There are many ways to define derivatives in such cases, but since well eventually work with spaces that have more structure, well follow an approach inspired by Sobolev spaces. The precise idea is as follows:

- We already know how to define D on $\mathcal{S}$, so we want to extend this definition by approximation. To do this, well need a way to measure distances between functions something that justifies writing $f_n \approx f$.
- This is done by introducing a norm. For functions $f \in \mathcal{S}$, we define

$$\|f\|_{\mathbb{D}^{1,2}} := \left(\|f\|^2_{L^2(\Omega)} + \|f'\|^2_{L^2(\Omega)} \right)^{1/2}.$$

  This turns $\mathcal{S}$ into a normed space (not yet complete).
- Since the space is not complete, let's just complete it. The resulting space will be denoted by $\mathbb{D}^{1,2}$
- By the construction in the previous step, an element F in $\mathbb{D}^{1,2}$ would have an approximating sequence f_n in the metric $\|\cdot\|_{\mathbb{D}^{1,2}}$. Since we have made the constuction to force the derivatives f_n to converge (in $L^2(\Omega)$), then we have a candidate $Df := \lim_n Df_n$.

The above program indeed works, and the first 3 steps are just definitions. The well-poseddness of the definition of Df, however, is not, and we require the following fundamental property: if f_n, g_n are sequences that are Cauchy in $\mathbb{D}^{1,2}$, and both converge in $L^2(\Omega)$ to the same random variable h, then it might become natural to expect that $Df_n - Dg_n$ converges to zero so that we can safely define $Dh = \lim_n Df_n$. Actually, by linearity, it suffices to think that the g_k are all equal to zero and the property of D being such that Df_n converges to zero in $L^2(\Omega)$ is called clossability. This property for D sounds reasonable but not exactly trivial to prove. The first couple of pages of the book [1] are devoted to prove this in detail. Here, we settle for describing the program for proving it and refer the reader to the aforementioned book for filling the details.

As in the theory of PDE's. the proof of this property is based on integration by parts, but now the integration by parts is carried not through the Lebesgue measure, but with respect to the Gaussian measure γ, so we have to formulate the appropriate integration by parts formula for the Gaussian distribution.

Lemma 2.1 (Integration by parts)**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be an absolutely continuous function such that $f' \in L^1(\gamma)$. Then the function $x \mapsto xf(x)$ belongs to $L^1(\gamma)$, and*

$$\int_{\mathbb{R}} xf(x)\, d\gamma(x) = \int_{\mathbb{R}} f'(x)\, d\gamma(x),$$

where γ denotes the standard Gaussian measure on $\mathbb{R}$.

Proof. We only make the sketch of the proof and in the exercise session you can do the details: we start with the identity

$$\int_{\mathbb{R}} xf(x)\, d\gamma(x) = \int_{\mathbb{R}} f(x) \left(x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right) dx = \int_{\mathbb{R}} f(x) \frac{d}{dx} \left(-\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right) dx.$$

An integration by parts formula yields the result. □

Now, let's go back to the closability property

Lemma 2.2 (Closability of D)**.** *The operator $D : \mathcal{S} \subset L^2(\gamma) \to L^2(\gamma)$ is closable.*

Proof. The key idea in this proof is to think of Df_n not just as a function, but as an operator acting on test functions in $L^2(\Omega)$. With that in mind, try imagining how you might approach the proof with just this perspective.

Now for the argument: let $(f_n) \subset \mathcal{S}$ be a sequence such that (i) $f_n \to 0$ in $L^2(\gamma)$, and (ii) $D^p f_n \to \eta$ in $L^2(\gamma)$, for some $\eta \in L^2(\gamma)$. We want to show that $\eta = 0$ almost everywhere. Take any test function $g \in \mathcal{S}$, and define the function

$$\delta g(x) := xg(x) - g'(x).$$

This object plays a key role in what's coming next. Using integration by parts, we compute:

$$\begin{aligned}\int_{\mathbb{R}} \eta(x)g(x)\,d\gamma(x) &= \lim_{n\to\infty} \int_{\mathbb{R}} Df_n(x)\,g(x)\,d\gamma(x)\\ &= \lim_{n\to\infty} \int_{\mathbb{R}} f_n(x)\,\delta g(x)\,d\gamma(x).\end{aligned}$$

Since $f_n \to 0$ in $L^2(\gamma)$ and $\delta g \in \mathcal{S} \subset L^2(\gamma)$, Hlders inequality tells us that the last integral tends to zero. So:

$$\int_{\mathbb{R}} \eta(x)g(x)\,d\gamma(x) = 0 \quad \text{for all } g \in \mathcal{S}.$$

Now, this means that the functional

$$g \mapsto \int_{\mathbb{R}} \eta(x)g(x)\,d\gamma(x)$$

vanishes on the whole space $\mathcal{S}$. From Proposition 1.1.5 in [1], we know that $\mathcal{S}$ is dense in $L^2(\gamma)$, so the functional must be zero on all of $L^2(\gamma)$. By the Riesz representation theorem, this implies that $\eta = 0$ γ-almost everywhere, as we wanted to show. □

One can play the same game for defining properly the derivative of higher order: Fix an integer $p \geq 1$. We define $\mathbb{D}^{p,2}$ as the closure of $\mathcal{S}$ with respect to the norm:

$$\|f\|_{\mathbb{D}^{p,2}} = \left(\int_{\mathbb{R}} |f(x)|^2\,d\gamma(x) + \int_{\mathbb{R}} |f'(x)|^2\,d\gamma(x) + \cdots + \int_{\mathbb{R}} |f^{(p)}(x)|^2\,d\gamma(x)\right)^{1/2}.$$

We can construct $\mathbb{D}^{p,2}$ similarly to $\mathbb{D}^{1,2}$ and define D^p by approximation as in the $p = 1$ case. The arguments for showing that the operator is closable are analogous. We will call D^p, defined over $\mathbb{D}^{p,2}$ Malliavin derivative of order p.

A closer examination of the proof of the closedness of the operator D allows us to see that the trick of "not looking at DX as a function but rather as looking at its action over other functions" implicitly required to consider the action of functions against δ. In other words, we used the following identity, which definetely deserves to be called integration by parts (thinking of integration by parts as a duality operation)

$$\int_{\mathbb{R}} Dg(x)f(x)\gamma(dx) = \int_{\mathbb{R}} g(x)\delta f(x)\gamma(dx)$$

for $f, x \in \mathbb{D}^{1,2}$.

Definition 2.3 (Definition of *Domδ*)**.** *We denote by* $\mathrm{Dom}(\delta^p)$ *the subset of* $L^2(\gamma)$ *composed of those functions* g *such that there exists a constant* $c > 0$ *satisfying the following property: for all* $f \in \mathcal{S}$,

$$\left| \int_{\mathbb{R}} f^{(p)}(x) g(x)\, d\gamma(x) \right| \le c \left(\int_{\mathbb{R}} f^2(x)\, d\gamma(x) \right)^{1/2}. \tag{2.1}$$

Now we define propertly the operator δ^r

Definition 2.4 (Divergence)**.** *The pth divergence operator* δ^p *is defined as follows: if* $g \in \mathrm{Dom}(\delta^p)$, *then* $\delta^p g$ *is the unique element of* $L^2(\Omega)$ *characterized by*

$$\int_{\mathbb{R}} f^{(p)}(x) g(x)\, d\gamma(x) = \int_{\mathbb{R}} f(x) \delta^p g(x)\, d\gamma(x). \tag{2.2}$$

which holds for $f \in \mathcal{S}$.

One could have also thought of formulating an iteration of the operation δ in the form $\delta^{r+1} := \delta[\delta^r]$. One can verify that both roads lead to the same result.

This note is a small reminder that we had a goal at the beginning, which was to construct an operation $\Gamma[X]$ that would satisfy (1.3). We are a couple of extra definitions apart from getting there. Be patient.

Definition 2.5 (Ornstein Uhlenbeck semigroup)**.** *The Ornstein–Uhlenbeck semigroup, written* $(P_t)_{t\ge 0}$, *is defined as follows. For* $f \in \mathcal{S}$ *and* $t \ge 0$,

$$P_t f(x) = \int_{\mathbb{R}} f\left(e^{-t}x + \sqrt{1-e^{-2t}}\, y \right) d\gamma(y), \quad x \in \mathbb{R},$$

where γ *denotes the standard Gaussian measure on* $\mathbb{R}$.

The semigroup P_θ has the fundamental property that (as any semigroup), its value at zero is the identity operator and in the other hand, we we evaluate P_θ at θ equal to infinity, we obtain the trivial operator that sends f to $\mathbb{E}[f(N)]$, where N is a standard Gaussian random variable. This gives the semigroup a nice interpretation of briding the operation of "not doing anything to test functions" with the operation of "acting over functions precisely like a standard Gaussian probability measure".

By elementary computations, one can check that P_t is a contraction that indeed the semigroup property $P_{s+t} = P_s \circ P_t$ holds. Being a semigroup, it has an associated generator, which we denote by L. More precisely, $\mathrm{Dom}\, L$ is defined as the collection of those $f \in L^2(\gamma)$ such that the expression

$$\frac{P_h f - f}{h}$$

converges in $L^2(\gamma)$ as $h \to 0$. By differentiating with respect to t in the expression of $P_t f(x)$,

$$\begin{aligned} \frac{d}{dt} P_t f(x) = &-x e^{-t} \int_{\mathbb{R}} f'\left(e^{-t}x + \sqrt{1-e^{-2t}}\, y \right) d\gamma(y) \\ &+ \frac{e^{-2t}}{\sqrt{1-e^{-2t}}} \int_{\mathbb{R}} f'\left(e^{-t}x + \sqrt{1-e^{-2t}}\, y \right) y\, d\gamma(y), \end{aligned}$$

so integrating by parts, we get

$$\frac{d}{dt}P_t f(x) = -xe^{-t}\int_{\mathbb{R}} f'\left(e^{-t}x + \sqrt{1-e^{-2t}}\,y\right) d\gamma(y) \\ + e^{-2t}\int_{\mathbb{R}} f''\left(e^{-t}x + \sqrt{1-e^{-2t}}\,y\right) d\gamma(y),$$

In particular, by evaluating the previous expression at $t = 0$, we get:

Proposition 2.1. *For any $X \in \mathcal{S}$, we have*

$$LX = -\delta DX.$$

Another important relationship between D and δ is the following relation, typically named as the Heisenberg commutativity relation

Proposition 2.2 (Heisenberg)**.** *For every $f \in \mathcal{S}$, it holds that*

$$(D\delta - \delta D)f = f.$$

Moreover,

$$(D\delta^p - \delta^p D)f = p\delta^{p-1}f, \quad \text{for all } f \in \mathcal{S}.$$

We can now have a sneak-peak of what is coming next: suppose for a moment that we could have X such that taking the inverse $L^{-1}[X]$ would make sense, so we could write $X = LL^{-1}X$. Then we have that

$$\mathbb{E}[Xf(X)] = -\mathbb{E}[\delta DL^{-1}f(X)] = \mathbb{E}[(-DL^{-1}X \cdot DX)f'(X)].$$

Then, the variable that we are looking for is

$$\Gamma[X] = (-DL^{-1}X) \cdot (DX).$$

Ok, I wrote a formula, but for the moment is not obvious what to do with it. In order to being able to use this effectively, we need a good basis to express the variable X in, and so that we could have nice compatibility with the measure γ. A common choice of basis of this sort are the orthogonal ones. With this in mind, we define the Hermite polynomials (note: there are some references that use a slightly modified version of the polynomials as we present them)

Definition 2.6 (Hermite polynomials)**.** *Let $p \geq 0$ be an integer. We define the pth Hermite polynomial as*

$$H_0(x) = 1 \quad \text{and} \quad H_p(x) := \delta^p 1, \quad \text{for } p \geq 1.$$

The reason why it is worth studying these polynomials is explained next

Proposition 2.3 (Properties of Hermite polynomials)**.**

(i) *For any $p \geq 0$, we have*

$$H_p' = pH_{p-1}, \quad LH_p = -pH_p, \quad \text{and} \quad P_tH_p = e^{-pt}H_p, \quad t \geq 0.$$

(ii) *For any $p \geq 0$,*

$$H_{p+1}(x) = xH_p(x) - pH_{p-1}(x).$$

(iii) *For any* $p, q \geq 0$,

$$\int_{\mathbb{R}} H_p(x) H_q(x)\, d\gamma(x) = \begin{cases} p! & \text{if } p = q, \\ 0 & \text{otherwise.} \end{cases}$$

(iv) *The family* $\left\{ \frac{1}{\sqrt{p!}} H_p : p \geq 0 \right\}$ *is an orthonormal basis of* $L^2(\gamma)$.

(v) *If* $f \in \mathbb{D}^{\infty,2}$, *then*

$$f = \sum_{p=0}^{\infty} \frac{1}{p!} \left(\int_{\mathbb{R}} f^{(p)}(x)\, d\gamma(x) \right) H_p \quad \text{in } L^2(\gamma).$$

Proof. (i) By the definition of H_p, we have

$$H_p' = D\delta^p 1.$$

Using Proposition 2.2, we get

$$H_p' = p\delta^{p-1} 1 + \delta^p D1 = pH_{p-1},$$

since $D1 = 0$. Now, using the fact that $L = -\delta D$, we find

$$LH_p = -\delta D H_p = -\delta H_p' = -\delta(pH_{p-1}) = -p\delta H_{p-1} = -pH_p.$$

To prove the third identity, fix $x \in \mathbb{R}$ and define $y_x(t) := P_t H_p(x)$. Then:

$$y_x(0) = P_0 H_p(x) = H_p(x),$$

and for $t > 0$,

$$y_x'(t) = \frac{d}{dt} P_t H_p(x) = P_t L H_p(x) = -pP_t H_p(x) = -p y_x(t).$$

Solving this differential equation gives $y_x(t) = e^{-pt} H_p(x)$, that is,

$$P_t H_p = e^{-pt} H_p.$$

(ii) Take $p \geq 1$. By definition, $H_{p+1} = \delta^{p+1} 1 = \delta\delta^p 1 = \delta H_p$. From the definition of the divergence operator δ, we have:

$$H_{p+1}(x) = xH_p(x) - H_p'(x).$$

Using the result from part (i), $H_p'(x) = pH_{p-1}(x)$, we deduce the recurrence:

$$H_{p+1}(x) = xH_p(x) - pH_{p-1}(x).$$

□

Now is the time to take some time to reflect on what we have done. We started with definitions which are very familiar to us: the notion of derivative that we appreciate since long ago was extended to a larger domain, then we defined the operation that is dual to D (hence the name divergence), and we denoted by δ. Some cryptic semigroup P_θ was introduced with the help of some explicit formula that relies in the standard Gaussian distribution somehow, and it was promised to be a handy tool that could interpolate the identity with a standard Gaussian behavior. This operator was utilized in a spirit very similar to MCMC to estimate variances by mean of the Poincaré inequality. The point here is that every single construction was motivated by objects that are familiar to us since quite some time ago. Then we are presented with a family of orthogonal polynomials, and if we look closely at the result, we don't only get the orthonormality (up to an explicit normalizing factor) with respect to the standard Gaussian distribution, but we also have very easy formulas for the actions of D, δ, P_θ, L over Hermite polynomials. This very simple, but extremely useful information can be used to stop thinking of D, δ, P_θ, L as familiar objects and simply think of them as abstract operations defined over the basis $H_0, H_1, H_2, ...$ Why is this helpful? suppose that you have an abstract expression of a given $L^2(\Omega)$ function f in the form

$$f(x) = \sum_{q\geq 0} a_q H_q.$$

(if you are very strict about convergence, just think of the sum as being finite). Clearly the operators D, δ, P_θ, L should act linearly over f, so that if $A \in \{D, \delta, P_\theta, L\}$, then

$$A[f](x) = \sum_{q\geq 0} a_q A[H_q].$$

Since we have a formula for $A[H_q]$ by the previous proposition we could instead of getting the explicit value of $A[H_q]$ as a consequence of our definitions, we could simply take it as a definition, and this perspective doesn't require at all the interpretation of Ω as being Euclidean, and actually, is the way to generalize these ideas to very abstract spaces (including Radamacher chaoses, Wigner chaos, etc).

3. The infinite dimensional case

In this section we present the implementation of the above ideas, but in the context of stochastic processes. In the sequel, $\mathcal{H}$ will denote the Hilbert space $\mathcal{H} := L^2([0,T])$, endowed with the Lebesgue measure, where T is a time horizon that is allowed to consider the case T equals to infinity.

Let B be a Brownian motion defined over $(\Omega, \mathcal{F}, \mathbb{P})$. In the sequel, we will assume that $\mathcal{F}$ is the σ-algebra generated by B. A very handy way of getting the randomness of B codified in a structured way is to consider the so called "isonormal Gaussian process associated to B", formally defined as the process $\{W_h \ ; \ h \in \mathfrak{H}\}$, with

$$W_h := \int_{[0,T]} h(x) W(dx). \tag{3.1}$$

Observe that it is equivalent to have access to B than to have access to W, since we can jump from B to W via (3.1) and reciprocally, we can recover B_t by noticing that $B_t = W_{\mathbb{1}_{[0,t]}}$. A fundamental property of the process W is that is centered Gaussian, with covariance given by

$$\mathbb{E}[W_{h_1} W_{h_2}] = \langle h_1, h_2 \rangle. \tag{3.2}$$

A very useful property of the Hermite polynomials is that they can be used to obtain a quite much more robust version of the isometry (3.2), as illustrated in the following proposition

Proposition 3.1. *Let $Z, Y \sim \mathcal{N}(0,1)$ be jointly Gaussian. Then, for all $n, m \geq 0$,*

$$\mathbb{E}[H_n(Z)H_m(Y)] = \begin{cases} \rho^n n!, & \text{if } n = m, \\ 0, & \text{otherwise}, \end{cases}$$

where $\rho = \mathbb{E}[ZY]$ and H_n denotes the nth Hermite polynomial.

Proof. Set $\rho = \mathbb{E}[ZY]$, and assume for the moment that $\rho > 0$. Let $N, N' \sim \mathcal{N}(0,1)$ be independent standard normal random variables, and observe that the pair (Z, Y) has the same law as $(N, \rho N + \sqrt{1-\rho^2}N')$. Hence, for $n, m \geq 0$, we compute $\mathbb{E}[H_n(Z)H_m(Y)]$ by replacing Z and Y accordingly. From here it follows that

$$\mathbb{E}[H_n(Z)H_m(Y)] = \mathbb{E}\left[H_n(N)\, P_{\log(1/\rho)} H_m(N)\right] = \rho^m\, \mathbb{E}[H_n(N)H_m(N)] = n!\mathbb{1}_{\{m=n\}}\rho^n.$$

The case $\rho < 0$ is an easy generalization that is left as an exercise to the reader. □

The above theorem gives us the building blocks for constructing an extension to the orthogonality decomposition of $\mathbb{R}$ (endowed with γ), but now extended to the whole Ω.

Definition 3.1. *For each $n \geq 0$, we write $\mathcal{H}_n$ to denote the closed linear subspace of $L^2(\Omega)$ generated by the random variables of the form $H_n(X(h))$, where $h \in \mathfrak{H}$ and $\|h\|_{\mathfrak{H}} = 1$. The space $\mathcal{H}_n$ is called the nth Wiener chaos of X.*

Clearly, $\mathcal{H}_0 = \mathbb{R}$ and $\mathcal{H}_1 = \{X(h) : h \in \mathfrak{H}\} = X$. By Proposition 2.2.1, if $n \neq m$, then $\mathcal{H}_n$ and $\mathcal{H}_m$ are orthogonal in the usual inner product of $L^2(\Omega)$. The following result is known as the WienerIt chaotic decomposition of $L^2(\Omega)$.

Theorem 3.2 (Chaos decomposition)**.** *The linear space generated by the class*

$$\{H_n(X(h)) : n \geq 0,\ h \in \mathfrak{H},\ \|h\|_{\mathfrak{H}} = 1\}$$

is dense in $L^q(\Omega)$ for every $q \in [1, \infty)$. Moreover, one has

$$L^2(\Omega) = \bigoplus_{n=0}^{\infty} \mathcal{H}_n.$$

That is, every random variable $F \in L^2(\Omega)$ admits a unique expansion as a (possibly infinite) sum of orthogonal components in each $\mathcal{H}_n$, and this sum converges in $L^2(\Omega)$.

The above theorem gives us a mild substitute of the orthogonal decomposition that we reviewed in dimension one, although the description will turn out to be quite more exact, as we will see a bit later. Before that, we copy paste absolutely everything that we have done in dimension one, but in infinite dimensions. The reader is referred to [] for rigurous proofs, and to have have a small leap of faith and believe me that the ideas presented in dimension one are exactly the same ones that are utilized in infinite dimensions, with just very small adjustments.

Our immediate goal is to generalize the definitions of D, δ, P_θ, L. We begin with the derivative. I emphasize that the material sounds like a reminescence of the one dimensional case because the constructions are basicaly the same. We begin by considering the set $\mathcal{S}$ of all random variables of the form

$$F = f\big(W_{\varphi_1}, \ldots, W_{\varphi_m}\big), \tag{3.3}$$

where $m \geq 1$, $f : \mathbb{R}^m \to \mathbb{R}$ is a C^∞-function such that f and all its partial derivatives have at most polynomial growth, and $\varphi_i \in \mathfrak{H}$, for $i = 1, \dots, m$. A random variable of the form (3.3) is said to be smooth.

Next we introduce the notion of Malliavin derivative. To this end, we introduce the notation

$$\mathfrak{H}^{\otimes q} := \{h : [0,T]^q \to ;\ h \text{ is square integrable with respect to Lebesgue}\}.$$

The notation alludes to the fact that the space in the right is isomorphic to the q-tensor of $\mathfrak{H}$, which can be constructed in an abstract way if we want, although for purposes of these notes, is completely fine to just work with this version. A special subset of $\mathfrak{H}^{\otimes}$ is the set $\mathfrak{H}^{\odot q}$ consisting of those elements of $\mathfrak{H}^{\otimes q}$, such that they are symmetric. This space is called the symmetrized tensor of order q.

Next, we introduce the notion of the Malliavin derivative in the space of smooth random variables. Let $F \in \mathcal{S}$ be given by (3.3), and let $p \geq 1$ be an integer. The pth Malliavin derivative of F (with respect to X) is defined as the element of $L^2(\Omega, \mathfrak{H}^{\odot p})$ (note the symmetric tensor product) given by

$$D^p F = \sum_{i_1,\dots,i_p=1}^{m} \frac{\partial^p f}{\partial x_{i_1} \cdots \partial x_{i_p}} (W_{\varphi_1}, \dots, W_{\varphi_m})\, \varphi_{i_1} \otimes \cdots \otimes \varphi_{i_p}.$$

At this point, a subtle issue arises: the representation of F in the form (3.3) is generally not unique. Therefore, one must verify that the definition of $D^p F$ does not depend on the specific representation chosen. This fact is true, although the full justification involves technical details and is left to the reader.

As in the one-dimensional case, the operator

$$D^p : \mathcal{S} \subset L^1(\Omega) \to L^1(\Omega, \mathfrak{H}^{\odot p})$$

is closable. This allows us to define, for $F \in \mathcal{S}$, the norm

$$\|F\|_{D^{p,2}} = \left(\mathbb{E}[|F|^2] + \mathbb{E}\left[\|DF\|_{\mathfrak{H}}^2\right] + \cdots + \mathbb{E}\left[\|D^p F\|_{\mathfrak{H}^{\odot p}}^2\right]\right)^{1/2}.$$

We can then complete $\mathcal{S}$ under this norm to define the Sobolev-type space $\mathbb{D}^{p,2}$, which will serve as the extended domain of the operator D^p. The Malliavin derivative satisfies the following chain rule

Proposition 3.3. *Let $\varphi : \mathbb{R}^m \to \mathbb{R}$ be a continuously differentiable function with bounded partial derivatives. Suppose that $F = (F_1, \dots, F_m)$ is a random vector whose components belong to $\mathbb{D}^{1,2}$. Then $\varphi(F) \in \mathbb{D}^{1,2}$, and*

$$D\varphi(F) = \sum_{i=1}^{m} \frac{\partial \varphi}{\partial x_i}(F) DF_i.$$

The notion of Malliavin derivative can be extended to the case of Hilbert valued random variables. Indeed, if we consider a separable Hilbert space $\mathfrak{U}$, we can define the space $\mathcal{S}_{\mathfrak{U}}$ of all smooth $\mathfrak{U}$-valued random elements of the type

$$F = \sum_{j=1}^{n} F_j v_j,$$

where $F_j \in \mathcal{S}$ and $v_j \in \mathfrak{U}$ (careful, here the v_j are not random!). For $k \geq 1$, the kth Malliavin derivative of any $F \in \mathcal{S}_{\mathfrak{U}}$ is given by the $\mathfrak{H}^{\otimes k} \otimes U$-valued random element

$$D^k F = \sum_{j=1}^{n} D^k F_j \otimes v_j.$$

A procedure as before can be performed, in order to deduce that this operator can be extended to a domain $\mathbb{D}^{k,2}(\mathfrak{U})$, which is defined completely analogusly to the cases we have considered. The reason why we insist of this level of generality is because $D^a F$ can be thought of as an element in $\mathfrak{U} = \mathfrak{H}^{\otimes a}$, and thus, now it makes sense to take D^b to this element. By looking at the above phenomena first over elements F that are smooth, we can show that

$$D^{a+b} F = D^a D^b F.$$

Something very important to notice is that the space of random elements in $\mathfrak{H}^{\otimes r}$ is a very familiar object. Indeed, the elements of this space are functions (random) over r parameters belonging to $[0, T]$. This way, we are thinking of the derivative as being inputed by a random variable and spitting a (multi-parameter) stochastic process.

We now continue with the extension of the divergence: Fix an integer $p \geq 1$. We will now define δ^p (the divergence operator of order p) as the adjoint of

$$D^p : \mathbb{D}^{p,2} \to L^2(\Omega, \mathfrak{H}^{\odot p}).$$

This is the exact analog of the operator δ^p we introduced before (line by line!).

Definition 3.2. *Let $p \geq 1$ be an integer. We denote by Dom(δ^p) the subset of $L^2(\Omega, \mathfrak{H}^{\otimes p})$ composed of those elements u such that there exists a constant $c > 0$ satisfying*

$$|\mathbb{E}[\langle D^p F, u\rangle_{\mathfrak{H}^{\otimes p}}]| \leq c\, \mathbb{E}[F^2] \quad \textit{for all } F \in \mathcal{S},$$

There exists a unique element in $L^2(\Omega)$, denoted by $\delta^p(u)$, such that

$$\mathbb{E}[\langle D^p F, u\rangle_{\mathfrak{H}^{\otimes p}}] = \mathbb{E}[F\, \delta^p(u)] \quad \textit{for all } F \in \mathcal{S}.$$

Definition 2.5.2. If $u \in \mathrm{Dom}(\delta^p)$, then $\delta^p(u)$ is the unique element of $L^2(\Omega)$ characterized by the following duality formula:

$$\mathbb{E}[F\, \delta^p(u)] = \mathbb{E}[\langle D^p F, u\rangle_{\mathfrak{H}^{\otimes p}}], \tag{2.5.2}$$

for all $F \in \mathcal{S}$. The operator

$$\delta^p : \mathrm{Dom}(\delta^p) \subset L^2(\Omega, \mathfrak{H}^{\otimes p}) \to L^2(\Omega)$$

is called the multiple divergence operator of order p. We define δ^0 to be equal to the identity. To make emphasis on the analogy with the one dimensional case, we will typically refer to the duality formula "integration by parts formula". Now, since δ is the dual of D, several properties of D should have translated consequences for the operator δ. The following proposition is the translation of the property of Leibnitz rule.

Proposition 3.4. *Let $F \in \mathbb{D}^{1,2}$ and $u \in \mathrm{Dom}(\delta)$ be such that the three expectations*

$$\mathbb{E}[F^2\|u\|_{\mathfrak{H}}^2], \quad \mathbb{E}[F^2\delta(u)^2], \quad \text{and} \quad \mathbb{E}[\langle DF, u\rangle_{\mathfrak{H}}^2]$$

are finite. Then $Fu \in \mathrm{Dom}(\delta)$, and

$$\delta(Fu) = F\delta(u) - \langle DF, u\rangle_{\mathfrak{H}}.$$

Proof. For any $G \in \mathcal{S}$, we have

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

Hence, using the assumptions, we conclude that $Fu \in \mathrm{Dom}(\delta)$ and

$$\delta(Fu) = F\delta(u) - \langle DF, u\rangle_{\mathfrak{H}}.$$

□

For a smooth random element of the form $u = \sum_{j=1}^n F_j h_j$, we then have

$$\delta(u) = \sum_{j=1}^n \delta(F_j h_j) = \sum_{j=1}^n \left(F_j W_{h_j} - \langle DF_j, h_j\rangle_{\mathfrak{H}}\right). \tag{2.5.3}$$

Consequently,

$$D\delta(u) = \sum_{j=1}^n \left(W_{h_j} DF_j + F_j h_j - \langle D^2 F_j, h_j\rangle_{\mathfrak{H}}\right).$$

On the other hand, it is immediate that

$$Du = \sum_{j=1}^n DF_j \otimes h_j,$$

Once more using the proposition, as well as the explicit expression of F_j as a finite sum of random numbers multiplied by non-random vectors in $\mathfrak{H}$, we obtain

$$\delta(Du) = \sum_{j=1}^n \delta(DF_j \otimes h_j) = \sum_{j=1}^n \left(X(h_j) DF_j - \langle D^2 F_j, h_j\rangle_{\mathfrak{H}}\right). \tag{2.5.4}$$

Combining these identities, we obtain the so-called Heisenberg commutativity relation

$$D\delta(u) - \delta(Du) = u. \tag{2.5.5}$$

Finally, in the same way we considered the derivative of Hilbert-valued random variables, we can consider the divergence of Hilbert valued random variables. The use of this, as in the case of the derivative, is that we can consider certain types of iterations of the divergence operator. Fix $k \geq 1$, let be a real separable Hilbert space, and let

$$u = \sum_{j=1}^n v_j \otimes h_j \in \mathfrak{U} \otimes \mathfrak{H}^{\otimes k},$$

where $v_j \in \mathfrak{U}$ and $h_j \in \mathfrak{H}^{\otimes k}$. We define $\delta_k(u)$ to be the $\mathfrak{U}$-valued random element

$$\delta_k(u) = \sum_{j=1}^{n} v_j \delta_k(h_j), \tag{3.4}$$

It can be shown that vectors of this type are dense in $\mathfrak{U} \otimes \mathfrak{H}^{\otimes k}$, so that δ_k can be extended to a bounded operator from $\mathfrak{U} \otimes \mathfrak{H}^{\otimes k}$ into $L^2(\Omega, \mathfrak{U})$.

Note that this construction allows a precise meaning to be given to the expression $\delta_k(f)$, where $f \in \mathfrak{H}^{\otimes p}$ and $p > k$. Indeed, since $\mathfrak{H}^{\otimes p} = \mathfrak{H}^{\otimes (p-k)} \otimes \mathfrak{H}^{\otimes k}$, we define $\delta_k(f)$ to be the element of $L^2(\Omega, \mathfrak{H}^{\otimes (p-k)})$ obtained by specializing the previous construction to the case $\mathfrak{U} = \mathfrak{H}^{\otimes (p-k)}$. Note that, for every $k = 1, \ldots, p-1$ and every $f \in \mathfrak{H}^{\otimes p}$, we also have

$$\delta_p(f) = \delta_{p-k}(\delta_k(f)). \tag{3.5}$$

One can get a bit confused when reading expressions that involve divergences and derivatives of Hilbert valued random objects, but the bright side of this is that the notation is quite robust and suggests analogies that were present in the one-dimensional case. The following proposition is one of them

Proposition 3.5. *Let $p \geq 1$ be an integer. For all $u \in \mathfrak{H}^{\otimes p}$, we have $\delta_p(u) \in \mathbb{D}^{1,2}$ and*

$$D\delta_p(u) = p\delta_{p-1}(u).$$

Proof. We proceed by induction. For $p = 1$, this is a direct consequence of Equation (2.5.5). Now, assume that $D\delta_p(u) = p\delta_{p-1}(u)$ holds for some $p \geq 1$ and all $u \in \mathfrak{H}^{\otimes p}$. Let $v \in \mathfrak{H}^{\otimes (p+1)}$. Then, again using (2.5.5),

$$D\delta_{p+1}(v) = D\delta(\delta_p v) = \delta(D(\delta_p v)) + \delta_p v = \delta(p\delta_{p-1}(v)) + \delta_p v = (p+1)\delta_p(v),$$

which completes the induction and proves the result for $p + 1$. □

As the reader might have noticed, we are making a repetition of everything presented in the one dimensional case. What about the generalization of the Hermite polynomials? well, by reasoning as before, we can consider the repeated action of δ^q acting over something. Since the domain of δ^q is $\mathfrak{H}^{\odot q}$. This motivates the following definition

Definition 3.6. *Multiple integrals Let $p \geq 1$ and $f \in \mathcal{H}^{\odot p}$. The pth multiple integral of f (with respect to W) is defined by*

$$I_p(f) = \delta^p(f).$$

What we are left to prove is that these are good generalizations of the Hermite polynomials. To this end, we first prove the following lemma

Proposition 3.7. *Let $p \geq 1$ and $f \in \mathcal{H}^{\odot p}$. Then, the multiple integral $I_p(f)$ belongs to the Sobolev space $\mathbb{D}^{\infty,q}$. Moreover, for all integers $r \geq 1$, the r-th Malliavin derivative of $I_p(f)$ satisfies:*

$$D^r I_p(f) = \begin{cases} \dfrac{p!}{(p-r)!} I_{p-r}(f) & \text{if } r \leq p, \\ 0 & \text{if } r > p. \end{cases}$$

Proof. For the proof of the fact that $I_p(f) \in \mathbb{D}^{\infty,2}$, please consult [1]. To show the explicit expression under consideration, notice that the first Malliavin derivative is given by:

$$DI_p(f) = D\delta^p(f) = p\delta^{p-1}(f) = pI_{p-1}(f),$$

which corresponds exactly to the formula for $r = 1$.

Applying this reasoning recursively, we deduce that $I_p(f) \in \mathbb{D}^{\infty,2}$, and we obtain the general expression for $D^r I_p(f)$ as above. □

Proposition 3.8 (Isometry property)**.** *Fix integers $1 \leq q \leq p$, and let $f \in \mathcal{H}^{\odot p}$, $g \in \mathcal{H}^{\odot q}$. Then we have In particular,*

$$\mathbb{E}[I_p(f)I_q(g)] = \begin{cases} p!\langle f, g\rangle_{\mathcal{H}^{\otimes p}} & \text{if } p = q, \\ 0 & \text{if } p > q. \end{cases}$$

Proof. It suffices to observe that

$$\mathbb{E}[I_p(f)I_q(g)] = \mathbb{E}[\delta^p(f)I_q(g)] = \mathbb{E}\left[\langle f, D^p I_q(g)\rangle_{\mathcal{H}^{\otimes p}}\right] = \begin{cases} p!\langle f, g\rangle_{\mathcal{H}^{\otimes p}} & \text{if } p = q, \\ 0 & \text{if } p > q. \end{cases}$$

The argumentation being exactly the same as in the one dimensional case. □

Finally, we connect all the dots. In particular, we introduce a relation between multiple integrals and Hermite polynomials, which ultimately yields a characterization of the chaoses $\mathcal{H}_q$, with $q \geq 1$.

Theorem 3.9. *Let $f \in \mathcal{H}$ be such that $\|f\|_{\mathcal{H}} = 1$. Then, for any integer $p \geq 1$, we have*

$$H_p(W_f) = I_p(f^{\otimes p}). \tag{2.7.7}$$

As a consequence, the linear operator I_p provides an isometry from $\mathcal{H}^{\odot p}$ to the p-th Wiener chaos $\mathcal{H}_p$ of X (equipped with the $L^2(\Omega)$-norm).

Proof. We prove identity (2.7.7) by induction on p. For $p = 1$, the result is immediate:

$$H_1(X(f)) = X(f) = \delta(f) = I_1(f).$$

Assume now that the property holds for all integers $1, 2, \ldots, p$. Then we compute:

$$\begin{aligned} I_{p+1}(f^{\otimes(p+1)}) &= \delta\left(\delta^p(f^{\otimes p})f\right) \\ &= \delta\left(I_p(f^{\otimes p})f\right) \\ &= I_p(f^{\otimes p})\delta(f) - \langle DI_p(f^{\otimes p}), f\rangle_{\mathcal{H}} \\ &= I_p(f^{\otimes p})W_f - pI_{p-1}(f^{\otimes(p-1)})\|f\|_{\mathcal{H}}^2 \\ &= H_p(W_f)W_f - pH_{p-1}(W_f) \\ &= H_{p+1}(W_f) \end{aligned}$$

This completes the induction and the proof. □

Now we generalize Stroock's formula for the stochastic processes version. The proof is analogous to the one dimensional case.

Corollary 3.10 (Stroock formula). *Every $F \in L^2(\Omega)$ can be expanded as*

$$F = \sum_{p=0}^{\infty} I_p(f_p), \tag{2.7.8}$$

for some unique collection of kernels $f_p \in \mathcal{H}^{\odot p}$, for all $p \geq 0$. Moreover, if $F \in \mathbb{D}^{n,2}$ for some $n \geq 1$, then for all $p \leq n$,

$$f_p = \frac{1}{p!}\mathbb{E}[D^p F].$$

Now we proceed with the formulation of the Ornstein Uhlenbeck semigroup

Definition 3.11. *The OrnsteinUhlenbeck semigroup $(P_t)_{t\geq 0}$ is defined, for all $t \geq 0$ and $F \in L^2(\Omega)$, by*

$$P_t(F) = \sum_{p=0}^{\infty} e^{-pt} J_p(F) \in L^2(\Omega),$$

where $J_p(F)$ denotes the projection of F onto the p-th Wiener chaos.

Next we connect with the formulation of Ornstein Uhlenbeck semigroup that we had before. Before stating exactly a type of result of this type, we observe that since F is measurable with respect to the Gaussian process W, we can view F as a measurable function of W, meaning $F = F(X)$ almost surely with respect to the law of X.

Now fix $t \geq 0$, and let X' be an independent copy of X, defined on a separate probability space. Then the expression

$$F\left(e^{-t}X + \sqrt{1-e^{-2t}}X'\right)$$

is a well-defined random variable, almost surely with respect to the product measure $\mathbb{P} \times \mathbb{P}'$. This is justified because the random vector $e^{-t}X + \sqrt{1-e^{-2t}}X'$ has the same distribution as X, so evaluating F at this input makes sense. As shown by an argument similar as before, the collection of operators

$$F \mapsto \mathbb{E}'\left[F\left(e^{-t}X + \sqrt{1-e^{-2t}}X'\right)\right], \quad t \geq 0,$$

where $\mathbb{E}'$ denotes expectation with respect to X', forms a semigroup that is well-defined for all $F \in L^1(\Omega)$. The next result shows that this semigroup coincides with the OrnsteinUhlenbeck semigroup $(P_t)_{t\geq 0}$ when restricted to $L^2(\Omega)$.

Theorem 3.12 (Mehlers formula). *For every $F \in L^2(\Omega)$ and every $t \geq 0$, we have*

$$P_t(F) = \mathbb{E}'\left[F\left(e^{-t}X + \sqrt{1-e^{-2t}}X'\right)\right], \tag{2.8.1}$$

where $\mathbb{E}'$ denotes expectation with respect to the independent copy X' of X.

Proof. To prove the identity, it is enough to verify it on a dense class of random variables in $L^2(\Omega)$. A natural choice is the linear span of exponentials of the form

$$F = \exp\left(X(h) - \tfrac{1}{2}\|h\|_{\mathcal{H}}^2\right), \quad h \in \mathcal{H}.$$

This class is dense in $L^2(\Omega)$, a fact that can be shown using arguments similar to those used in the one dimensional case. We therefore focus on the case where $F = \exp\left(X(h) - \frac{1}{2}\right)$, assuming without loss of generality that $\|h\|_{\mathcal{H}} = 1$. Fix $t \geq 0$, and let X' be an independent copy of X. Then, using the representation from the right-hand side of Mehlers formula, we compute:

$$\mathbb{E}'\left[F\left(e^{-t}X + \sqrt{1-e^{-2t}}X'\right)\right] = \mathbb{E}'\left[\exp\left(e^{-t}X(h) + \sqrt{1-e^{-2t}}X'(h) - \tfrac{1}{2}\right)\right].$$

Since $X'(h)$ is a standard Gaussian independent of $X(h)$, the expectation becomes:

$$\exp\left(e^{-t}X(h) - \tfrac{1}{2}e^{-2t}\right).$$

Using the Hermite expansion of the exponential function, we have:

$$\exp\left(e^{-t}X(h) - \tfrac{1}{2}e^{-2t}\right) = \sum_{p=0}^{\infty} \frac{e^{-pt}}{p!} H_p(X(h)).$$

On the other hand, by equation (2.7.7), we know that:

$$H_p(X(h)) = I_p(h^{\otimes p}),$$

so the above becomes:

$$\sum_{p=0}^{\infty} e^{-pt} I_p(h^{\otimes p}) = P_t(F).$$

This confirms the identity for all F in the span of exponentials, and since this class is dense in $L^2(\Omega)$, the result follows for all $F \in L^2(\Omega)$ by continuity of both sides. □

We now consider the associated generator

Definition 3.13. *We say that a random variable $F \in L^2(\Omega)$ belongs to the domain of the OrnsteinUhlenbeck generator, denoted* $\mathrm{Dom}(L)$, *if*

$$\sum_{p=1}^{\infty} p^2\, \mathbb{E}\left[J_p(F)^2\right] < \infty,$$

where $J_p(F)$ denotes the projection of F onto the p-th Wiener chaos.

The following result establishes a fundamental connection between the Malliavin derivative D, the divergence operator δ, and the Ornstein-Uhlenbeck generator L. It serves as the exact analogue of the one dimensional case.

Proposition 3.14. *Let $F \in L^2(\Omega)$. Then $F \in \mathrm{Dom}(L)$ if and only if $F \in \mathbb{D}^{1,2}$ and $DF \in \mathrm{Dom}(\delta)$. In this case, we have*

$$\delta(DF) = -LF.$$

It is worth noting that the quantity $-DL^{-1}F$ admits at least two different, yet equivalent, representations. Suppose $F \in \mathbb{D}^{1,2}$ and has mean zero, that is, $\mathbb{E}[F] = 0$. Then we have the following result.

Proposition 3.15. *Let $F \in \mathbb{D}^{1,2}$ with $\mathbb{E}[F] = 0$. Then*

$$-DL^{-1}F = \int_0^{\infty} e^{-t} P_t DF\, dt$$

Proof. Assume that $F = I_p(f)$ for some $p \geq 1$ and $f \in \mathcal{H}^{\odot p}$. Then we know that $L^{-1}F = -\frac{1}{p}F$, so by Proposition 2.7.4 we get:

$$-DL^{-1}F = \frac{1}{p}DF = I_{p-1}(f).$$

On the other hand, using the action of the Ornstein–Uhlenbeck semigroup on derivatives, we have:

$$P_t DF = pe^{-(p-1)t} I_{p-1}(f),$$

and therefore,

$$\int_0^\infty e^{-t} P_t DF \, dt = \int_0^\infty e^{-t} pe^{-(p-1)t} I_{p-1}(f) \, dt = I_{p-1}(f),$$

since the integral evaluates to $\frac{p}{p} = 1$. This proves the identity. □

References

[1] Ivan Nourdin and Giovanni Peccati. *Normal Approximations Using Malliavin Calculus: from Stein's Method to Universality*, volume 192 of *Cambridge Tracts in Mathematics*. Cambridge University Press, 2012.