Brief introduction to Malliavin calculus

January 13, 2021

Very big picture: Malliavin calculus is a calculus of variations for functionals of Brownian motion. For example: if X solves an SDE then X_t and $\sup_{0\le t\le 1}|X_t - \hat{X}_t|^2$ for some (adapted) discretization $\hat{X}$ are such functionals.
Notation: for clarity I write $L^2(A \to B, \mu)$ for the (Hilbert) space of square integrable B-valued functions on the measure space (A, μ). Unless specified the σ-algebra is the one generated by the Brownian motion.

1 The Malliavin derivative

1.1 Hands-on approach

Here we present an elementary approach to the construction of the Malliavin derivative, adapted from [Bal03]. Suppose we have an m-dimensional Brownian motion W and for concreteness let us work on the time horizon $[0,1)$. Denote $(\Omega, \mathcal{F}, \mathbb{P})$ the underlying probability space with $\mathcal{F} = \sigma((W(t))_{0\le t\le 1})$. For a scale $n \ge 1$ let use write $t_n^k = k/2^n$ for $0 \le k \le 2^n$ so that the intervals $I_n^k := [t_n^k, t_n^{k+1})$ form the dyadic partition of $[0,1)$. We now take the (independent) increments of the Brownian motion on this partition:

$$\Delta_n^k := W(t_n^{k+1}) - W(t_n^k).$$

Let $f_n : \mathbb{R}^{m\times 2^n} \to \mathbb{R}$ be smooth with all its partial derivatives at most polynomial, and set $X_n = f_n(\Delta_n^0, \dots, \Delta_n^{2^n-1})$ (sometimes called a simple functional of order n) noting that X_n has moments of all orders. Importantly we have uniqueness in the sense that $f_n(\Delta_n) = f_n'(\Delta_n)$ implies $f = f'$. Let $\mathcal{S}_n$ denote the space of all such X_n and $\mathcal{S} = \cup_n \mathcal{S}_n$. The following Lemma will be useful for the infinite dimensional extension of the Malliavin derivative.

Lemma 1.1 — *$\mathcal{S}$ is a dense linear subspace of $L^p(\Omega, \mathcal{F}, \mathbb{P})$ for every $p > 0$.*

For a random variable $X \in \mathcal{S}_n$ of the form $X = f(\Delta_n)$ we are ready to define its Malliavin derivative. For a given $t \in [0,1)$ let k_0 be such that $t \in I_n^{k_0}$. The Malliavin derivative is given by

$$\mathcal{D}_t^{(i)} X := \frac{\partial f}{\partial x_i^{k_0}}(\Delta_n) = \sum_{k=0}^{2^n-1} \frac{\partial f}{\partial x_i^k}(\Delta_n) \mathbb{1}_{I_n^k}(t).$$

Let us clarify the notation: $f = f(x_1, \dots, x_m)$ where $x_i \in \mathbb{R}^{2^n-1}$ are given as $x_i = (x_i^0, \dots, x_i^{2^n-1})$. Intuitively, $\mathcal{D}_t^{(i)}$ differentiates X with respect to the increment of the i'th dimension of W at time t, approximately. Formally we could write

$$\mathcal{D}_t^{(i)} = \frac{\partial X}{\partial \Delta_i^t}.$$

Note that $\mathcal{D}X = (\mathcal{D}_t^1 X, \dots, \mathcal{D}_t^m X)_{t\in[0,1)}$ is a random function from $[0,1)$ into $\mathbb{R}^m$. Thus, $\mathcal{D}$ maps $\mathcal{S}$ to random elements of $L^2([0,1) \to \mathbb{R}^m, \text{Leb})$. By the virtue of classical calculus, the operator $\mathcal{D}$

satisfies a chain rule like all good derivatives should:

$$\mathcal{D}F(X_1,\dots,X_l)=\sum_{j=1}^{l}\partial_j F(X_1,\dots,X_l)\mathcal{D}X_j$$

for all $X_j\in\mathcal{S}$ and nice $F:\mathbb{R}^l\to\mathbb{R}$. Note we are implicitly (correctly) assuming that the definition of $\mathcal{D}$ is consistent with the nested sequence

$$\dots\subseteq\mathcal{S}_n\subseteq\mathcal{S}_{n+1}\subseteq\dots$$

For $F\in\mathcal{S}$ we may define the norm

$$\|F\|_{1,2}^2:=\mathbb{E}|F|^2+\mathbb{E}\langle\mathcal{D}F,\mathcal{D}F\rangle=\mathbb{E}|F|^2+\mathbb{E}\int_0^1\sum_{i=1}^{m}|\mathcal{D}_t^{(i)}F|^2\,\mathrm{d}t.$$

Noting the resemblence to Sobolev norms the definition of $\|\cdot\|_{r,p}$ for $r,p\geq 1$ also becomes clear (modulo the precise definition of higher order Malliavin derivatives). The final step is to define $\mathbb{D}^{r,p}$ as the completion of $\mathcal{S}$ with respect to the norm $\|\cdot\|_{r,p}$ and for each $X\in\mathbb{D}^{r,p}$ one writes $\mathcal{D}X=\lim_{n\to\infty}\mathcal{D}X_n$ for any sequence $X_n\in\mathcal{S}$ converging to X w.r.t. $\|\cdot\|_{r,p}$ (this requires justification but we don't go into that).

1.2 Abstract definition

From here on we follow [Hai19, Nua98]. Before we give the abstract definition of the Malliavin derivative we must introduce Gaussian white noise. Let $H=L^2(\mathbb{R}_+\to\mathbb{R}^m,\mathrm{Leb})$. Gaussian white noise is a linear isometry $W:H\to L^2(\Omega\to\mathbb{R},\mathbb{P})$ for some probability space $(\Omega,\mathbb{P})$ such that $W(h)$ is a centered Gaussian random variable for each $h\in H$. Since W is an isometry by definition we have

$$\mathbb{E}W(h)W(g)=\langle h,g\rangle.$$

Such an isometry is easily constructed by taking a countable orthonormal basis $\{e_n\}_{n\geq 0}$ of H and setting $W(e_i)=\xi_i$ where the ξ_i are i.i.d. standard Gaussian. Importantly, the white noise W defines an m-dimensional Wiener process $W=(W_1,\dots,W_m)$ by setting

$$W_i(t)=W((0,\dots,0,\mathbb{1}_{[0,t)},0,\dots,0))=:W(\mathbb{1}_{[0,t)}^{(i)})$$

Checking then that $\mathbb{E}W_i(t)W_j(s)=\delta_{i,j}(s\wedge t)$ is straightforward. By approximation one might also check that for all $h=(h_1,\dots,h_m)\in H$

$$W(h)=\sum_{i=1}^{m}\int_0^\infty h_i(t)\,\mathrm{d}W_i(t)$$

where the RHS is an Ito integral. Let us formally write $\xi_i(t)=\frac{\mathrm{d}W_i}{\mathrm{d}t}$, to be interpreted as the increment of the i'th component of W at time t. Formally, for $h\in H$

$$W(h)=\sum_{i=1}^{m}\int_0^\infty h_i(t)\xi_i(t)\,\mathrm{d}t.$$

From the previous section, the Malliavin derivative $\mathcal{D}_t^{(i)}$ intuitively takes the derivative of functionals of W with respect to $\xi_i(t)$. Thus, it would be desirable to have

$$\mathcal{D}_t^{(i)}W(h)=h_i(t)\tag{1.1}$$

for all $h \in H$. Further, a chain rule such as

$$\mathcal{D}_t^{(i)} F(X_1, \ldots, X_n) = \sum_{k=1}^{n} \partial_k F(X_1, \ldots, X_n) \mathcal{D}_t^{(i)} X_k$$

ought to hold. Let $\mathcal{S} \subseteq L^2(\Omega, \mathbb{P})$ denote the analogue of the simple functionals $\mathcal{S}$ in this abstract setting: $X \in \mathcal{S}$ provided there is $n \geq 0$, $h_1, \ldots, h_n \in H$ and smooth $F : \mathbb{R}^n \to \mathbb{R}$ with partials at most polynomial such that $X = F(W(h_1), \ldots, W(h_n))$ (Clearly $\mathcal{S}$ is more general than the simple functionals considered in Section 1.1, just take the h_i to be indicator functions of the dyadic partition). We define the Malliavin derivative first on simple functionals as

$$\mathcal{D}X = \sum_{k=1}^{n} \partial_k F(W(h_1), \ldots, W(h_n)) h_k$$

noting that D maps $\mathcal{S}$ to random elements of H. Another interpretation is the following: regarding F as a function of the Brownian motion $W(\cdot)$, the directional derivative $\langle \mathcal{D}F, h \rangle$ is recovered via

$$\langle \mathcal{D}F, h \rangle \overset{P}{=} \lim_{\epsilon \downarrow 0} \frac{F(W + \epsilon \int_0^{\cdot} h(s)\, \mathrm{d}s) - F(W)}{\epsilon}$$

which brings out the calculus of variations interpretation. Once again $\mathcal{D}$ is independent of the choice of representation F and satisfies the Leibniz and chain rules. As one would hope [Nua98, Proposition 1.3.1], the derivative of a constant is 0 so that for any $c \in \mathbb{R}$ we have $\mathbb{1}_{F=c} \mathcal{D}F = 0$ and $\mathcal{D}F = 0$ implies $F = \mathbb{E}F$ ([Nua98, Remark p.134]). One of the most important tools, unexpectedly, is integration by parts:

Theorem 1.2 ([Hai19, Proposition 3.1]) — *For every $X \in \mathcal{S}$ and $h \in H$ one has the identity*

$$\mathbb{E}\langle \mathcal{D}X, h \rangle = \mathbb{E}XW(h).$$

Proof. Let $X = F(W(h_1), \ldots, W(h_n))$ and assume WLOG (Gram-Schmidt and linearity of white noise) that $\langle h_i, h_j \rangle = \delta_{ij}$ and extend them to an orthonormal basis $\{h_k\}_{k \geq 1}$. We have

$$\begin{aligned}
\mathbb{E}\langle \mathcal{D}X, h \rangle &= \sum_{k=1}^{n} \langle h_k, h \rangle \mathbb{E} \partial_k F(W(h_1), \ldots, W(h_n)) \\
&= \sum_{k=1}^{n} \frac{\langle h_k, h \rangle}{(2\pi)^{d/2}} \int_{\mathbb{R}^n} e^{-\frac{|x|^2}{2}} \partial_k F(x)\, \mathrm{d}x \\
&= \sum_{k=1}^{n} \frac{\langle h_k, h \rangle}{(2\pi)^{d/2}} \int_{\mathbb{R}^n} e^{-\frac{|x|^2}{2}} F(x) x_k\, \mathrm{d}x \\
&= \sum_{k=1}^{n} \mathbb{E}\left[XW(h_k) \langle h_k, h \rangle \right] \\
&= \mathbb{E}\left[XW\left(\sum_{k=1}^{n} \langle h_k, h \rangle h \right) \right] \\
&= \mathbb{E}\left[XW\left(\sum_{k=1}^{\infty} \langle h_k, h \rangle h \right) \right] \\
&= \mathbb{E}XW(h)
\end{aligned}$$

where in the third equality we used Gaussian integration by parts. □

By the Leibniz rule of classical calculus the Malliavin derivative also satisfies a Leibniz rule on $\mathcal{S}$ giving us the following.

Corollary 1.2.1 ([Hai19, Corollary 3.2]) — *For every* $X, Y \in \mathcal{S}$ *and* $h \in H$ *one has*

$$\mathbb{E}Y\langle \mathcal{D}X, h\rangle = \mathbb{E}\left[XYW(h) - X\langle \mathcal{D}Y, h\rangle\right].$$

Finally, we can extend $\mathcal{D}$ to the closure of $\mathcal{S}$ with respect to various Sobolev-type norms which we denote $\mathbb{D}^{r,p}$. The chain and Leibniz rules as well as integration by parts extend to the closure as well. We finally define the set of Malliavin smooth random variables to be $\mathbb{D}^\infty = \cap_{r,p\geq 1}\mathbb{D}^{r,p}$.

2 The Skorokhod integral

The Skorokhod integral δ is the adjoint of $\mathcal{D}$ and is defined by the integration by parts formula

$$\mathbb{E}\langle \mathcal{D}X, u\rangle = \mathbb{E}X\delta u$$

for all Malliavin smooth X and random $u \in H$ (i.e. $u \in \mathrm{dom}(\delta) \subseteq L^2(\Omega \to H, \mathbb{P})$). Clearly if u is deterministic, then Theorem 1.2 yields $\delta u = W(u)$. Two questions arise: what is the domain of δ and why is it called an integral? To answer the latter question: Ito integration deals with integrands that are adapted, while the random function $u : \mathbb{R}_+ \to \mathbb{R}^m$ can depend on $\mathcal{F} = \sigma((W(t))_{t\geq 0})$ arbitrarily and need not be adapted. We can however restrict our attention to u's that are adapted to the filtration $\mathcal{F}_t = \sigma((W(s))_{s\leq t})$. To this end let $L^2_a(\Omega \to H, \mathbb{P})$ denote the space of these adapted processes, defined as the closure of the space of simple processes of the form

$$u = \sum_{i=1}^{m}\sum_{k=1}^{n} Y_k^{(i)} \mathbb{1}^{(i)}_{[s_k,t_k)}$$

for $n \geq 0$, $0 \leq s_k < t_k < \infty$ and $Y_k^{(i)} \in L^2(\Omega, \mathcal{F}_{s_k}, \mathbb{P})$. On such processes the Ito integral is given as

$$\int_0^\infty Y(t)\,\mathrm{d}W(t) := \sum_{i=1}^{m}\sum_{k=1}^{n} Y_k^{(i)}(W_i(t_k) - W_i(s_k)) = \sum_{i=1}^{m}\sum_{k=1}^{n} Y_k^{(i)} W(\mathbb{1}^{(i)}_{[s_k,t_k)}).$$

The following clarifies the relation between δ and the Ito integral:

Theorem 2.1 ([Hai19, Theorem 3.5]) — $L^2_a(\Omega \to H, \mathbb{P}) \subseteq \mathrm{dom}\,\delta$ *and on it* δ *coincides with the Ito integral.*

Proof. Let $u = Y_s \mathbb{1}^{(i)}_{[s,t)}$ where Y_s is $\mathcal{F}_s$ measurable. For $X \in \mathcal{S}$ one has

$$\begin{aligned}\mathbb{E}\langle u, \mathcal{D}X\rangle &= \mathbb{E}Y_s\langle \mathbb{1}^{(i)}_{[s,t)}, \mathcal{D}X\rangle \\ &= \mathbb{E}Y_s X W(\mathbb{1}^{(i)}_{[s,t)}) - \mathbb{E}X\langle \mathcal{D}Y_s, \mathbb{1}^{(i)}_{[s,t)}\rangle\end{aligned}$$

by Corollary 1.2.1. Now, since Y_s is $\mathcal{F}_s$-measurable, so is $\mathcal{D}Y_s$. But this is equivalent to saying that $\mathrm{supp}\,\mathcal{D}Y_s \subseteq [0, s]$. In particular, $\langle \mathcal{D}Y_s, \mathbb{1}^{(i)}_{[s,t)}\rangle = 0$ almost surely. Noting that $W(\mathbb{1}^{(i)}_{[s,t)}) = W_i(t) - W_i(s)$, we have shown that

$$\mathbb{E}\langle u, \mathcal{D}X\rangle = \mathbb{E}\left[X \int_0^\infty u(t)\,\mathrm{d}W(t)\right]$$

for all $X \in \mathcal{S}$. By linearity and approximation, the above display holds for arbitrary $u \in L^2_a(\Omega \to H, \mathbb{P})$ and the result follows. □

A generalization of Ito isometry holds also.

Theorem 2.2 ([Hai19, Theorem 3.6]) — *For* $u \in \mathrm{dom}\,\delta$ *one has*

$$\mathbb{E}|\delta u|^2 = \mathbb{E}\int_0^\infty |u(t)|^2\,\mathrm{d}t + \mathbb{E}\sum_{i,j}\int_0^\infty\int_0^\infty \mathcal{D}_s^{(i)}u_j(t)\mathcal{D}_t^{(j)}u_i(s)\,\mathrm{d}s\,\mathrm{d}t.$$

For adapted u Ito isometry is recovered by noting that $\mathcal{D}_s^{(i)} u_j(t) = 0$ if $s > t$.

3 The density of random variables

Originally, Malliavin calculus was developed to study the regularity of solutions to certain PDEs. More specifically, given a diffusion, under what conditions does it have a density with respect to the Lebesgue measure, when is it smooth, what kind of regularity can one obtain. Via the Fokker-Planck equation this question can be framed in purely analytic terms. In his seminal work [Hör67] Fields medalist Lars Hörmander proved that under Hörmander's conditions (which corresponds to integrability conditions on the inverse Malliavin matrix) the solution is smooth, using only analytic arguments. Due to the clear probabilistic interpretation a probabilistic proof was sought for a long time, eventually achieved by Paul Malliavin in [Mal78]. In this section we give a very brief idea of how Malliavin calculus can be used to study regularity of densities.

Lemma 3.1 ([Hai19, Lemma 4.1]) — *Let X be an $\mathbb{R}^n$ valued random variable and suppose for each multiindex α there exists a constant C_α such that $|\mathbb{E}\partial^\alpha G(X)| \le C_\alpha \|G\|_\infty$ for every $G \in \mathcal{C}_0^\infty$. Then X has a smooth density with respect to the Lebesgue measure.*

Proof. Consequence of Sobolev embedding theorems. □

Suppose now that each component of $X \in \mathbb{R}^d$ is Malliavin smooth. To prove that X has a smooth density, the idea is to find for each i a random variable $Y_i \in H$ that may depend on X but not on the choice of $G \in \mathcal{C}_0^\infty$ such that

$$\partial_i G(X) = \langle \mathcal{D}G(X), Y_i \rangle \tag{3.1}$$

almost surely. If this is possible, then the definition of δ yields

$$|\mathbb{E}\partial_i G(X)| = |\mathbb{E}\langle \mathcal{D}G(X), Y_i\rangle| = |\mathbb{E}G(X)\delta Y_i| \le \|G\|_\infty \mathbb{E}|\delta Y_i|.$$

If δY_i is integrable then the condition of Lemma 3.1 holds with $|k| = 1$. Omitting the details, the result for $|k| > 1$ follows by the repeated application of integration by parts i.e. induction. So when does Y_i exist? By the chain rule

$$\langle \mathcal{D}G(X), Y_i\rangle = \sum_{j=1}^{d} \partial_j G(X) \langle \mathcal{D}X_j, Y_i\rangle$$

and we see that if $\langle \mathcal{D}X_j, Y_i\rangle = \delta_{ij}$ is solvable we are done. This is possible only if $\langle \mathcal{D}X, \cdot\rangle : H \to \mathbb{R}^n$ is surjective almost surely, i.e. if the Malliavin matrix $\mathcal{M}$ with entries

$$\mathcal{M}_{ij} = \langle \mathcal{D}X_i, \mathcal{D}X_j\rangle \tag{3.2}$$

is invertible almost surely in which case the minimum H-norm solution is given by

$$Y = \underbrace{(\mathcal{D}X)^*}_{\in H^d} \underbrace{\mathcal{M}^{-1}}_{\in \mathbb{R}^{d\times d}}$$

where the multiplication is defined in the obvious way. By induction, verifying the condition of Lemma 3.1 boils down to checking that $\mathcal{M}^{-1} \in L^p$ for all p.

Theorem 3.2 ([Hai19, Theorem 4.3]) — *Let X be a Malliavin smooth $\mathbb{R}^n$ valued random variable and suppose its Malliavin matrix $\mathcal{M}$ has inverse moments of all orders. Then the law of X has smooth density with respect to the Lebesgue measure.*

For more details and a full proof of Hörmander's Theorem see [Hai19].

References

[Bal03] Vlad Bally. An elementary introduction to malliavin calculus. 2003.

[Hai19] M Hairer. Advanced stochastic analysis, 2019.

[Hör67] Lars Hörmander. Hypoelliptic second order differential equations. *Acta Mathematica*, 119(1):147–171, 1967.

[Mal78] Paul Malliavin. Stochastic calculus of variations and hypoelliptic operators. In *Proc. Internat. Symposium on Stochastic Differential Equations, Kyoto Univ., Kyoto, 1976*. Wiley, 1978.

[Nua98] David Nualart. Analysis on wiener space and anticipating stochastic calculus. In *Lectures on probability theory and statistics*, pages 123–220. Springer, 1998.