Introduction

One of the remarkable properties of Brownian motion is that we can use it to construct (stochastic) integrals of the type

$$\int \ldots dB.$$

The reason this is remarkable is that almost every Brownian sample path $(B_t(\omega) : t \in [0,T])$ has infinite variation and there is no help from the classical Stieltjes integration theory. Instead, Itô's theory of stochastic integration relies crucially on the fact that B is a martingale and stochastic integrals themselves are constructed as martingales. If one recalls the elementary interpretation of martingales as fair games one sees that Itô integration is some sort of martingale transform in which the integrand has the meaning of a gambling strategy. Clearly then, the integrand must not anticipate the random movements of the driving Brownian motion and one is led to the class of so-called previsible processes which can be integrated against Brownian motion. When such integration is possible, it allows for a theory of stochastic differential equations (SDEs) of the form[1]

$$dY = \sum_{i=1}^{d} V_i(Y)\, dB^i + V_0(Y)\, dt\,, \quad Y(0) = y_0. \tag{$*$}$$

Without going into too much detail, it is hard to overstate the importance of Itô's theory: it has a profound impact on modern mathematics, both pure and applied, not to speak of applications in fields such as physics, engineering, biology and finance.

It is natural to ask whether the meaning of $(*)$ can be extended to processes other than Brownian motion. For instance, there is motivation from mathematical finance to generalize the driving process to general (semi-)martingales and luckily Itô's approach can be carried out naturally in this context.

We can also ask for a Gaussian generalization, for instance by considering a differential equation of the form $(*)$ in which the driving signal may be taken from a reasonably general class of Gaussian processes. Such equations have been proposed, often in the setting of fractional Brownian motion of Hurst parameter $H > 1/2$,[2] as toy models to study the ergodic behaviour

[1] Here $B = \left(B^1, \ldots, B^d\right)$ is a d-dimensional Brownian motion.

[2] Hurst parameter $H = 1/2$ corresponds to Brownian motion. For $H > 1/2$, one has enough sample path regularity to use Young integration.

of non-Markovian systems or to provide new examples of arbitrage-free markets under transactions costs.

Or we can ask for a Markovian generalization. Indeed, it is not hard to think of motivating physical examples (such as heat flow in rough media) in which Brownian motion B may be replaced by a Markov process X^a with uniformly elliptic generator in divergence form, say $\frac{1}{2}\sum_{i,j}\partial_i\left(a^{ij}\partial_j\cdot\right)$, without any regularity assumptions on the symmetric matrix $\left(a^{ij}\right)$.

The Gaussian and Markovian examples have in common that the sample path behaviour can be arbitrarily close to Brownian motion (e.g. by taking $H = 1/2 \pm \varepsilon$ resp. a uniformly ε-close to the identity matrix I). And yet, Itô's theory has a complete breakdown!

It has emerged over recent years, starting with the pioneering work of T. Lyons [116], that differential equations driven by such non-semimartingales can be solved in the rough path sense. Moreover, the so-obtained solutions are *not* abstract nonsense but have firm probabilistic justification. For instance, if the driving signal converges to Brownian motion (in some reasonable sense which covers $\varepsilon \to 0$ in the aforementioned examples) the corresponding rough path solutions converge to the classical Stratonovich solution of $(*)$, as one would hope.

While this alone seems to allow for flexible and robust stochastic modelling, it is not all about dealing with new types of driving signals. Even in the classical case of Brownian motion, we get some remarkable insights. Namely, the (Stratonovich) solution to $(*)$ can be represented as a deterministic and continuous image of Brownian motion and Lévy's stochastic area

$$A_t^{j\,k}(\omega) = \frac{1}{2}\left(\int_0^t B^j\,dB^k - \int_0^t B^k\,dB^j\right)$$

alone. In fact, there is a "nice" deterministic map, the *Itô–Lyons map*,

$$(y_0;\mathbf{x}) \mapsto \pi\,(0, y_0;\mathbf{x})$$

which yields, upon setting $\mathbf{x} = \left(B^i, A^{j,k} : i,j,k \in \{1,\ldots,d\}\right)$ a very pleasing version of the solution of $(*)$. Indeed, subject to sufficient regularity of the coefficients, we see that $(*)$ can be solved simultaneously for all starting points y_0, and even all coefficients! Clearly then, one can allow the starting point and coefficients to be random (even dependent on the entire future of the Brownian driving signals) without problems; in stark contrast to Itô's theory which struggles with the integration of non-previsible integrands. Also, construction of stochastic flows becomes a trivial corollary of purely deterministic regularity properties of the Itô–Lyons map.

This brings us to the (deterministic) main result of the theory: continuity of the Itô–Lyons map

$$\mathbf{x} \mapsto \pi\,(0, y_0;\mathbf{x})$$

in "rough path" topology. When applied in a standard SDE context, it quickly gives an entire catalogue of limit theorems. It also allows us to reduce (highly non-trivial) results, such as the Stroock–Varadhan support theorem or the Freidlin–Wentzell estimates, to relatively simple statements about Brownian motion and Lévy's area. Moreover, and at no extra price, all these results come at the level of stochastic flows. The Itô–Lyons map is also seen to be regular in certain perturbations of $\mathbf{x}$ which include (but are not restricted to) the usual Cameron–Martin space, and so there is a natural interplay with Malliavin calculus. At last, there is increasing evidence that rough path techniques will play an important role in the theory of stochastic *partial* differential equations and we have included some first results in this direction.

All that said, let us emphasize that the rough path approach to (stochastic) differential equations is *not* set out to replace Itô's point of view. Rather, it complements Itô's theory in precisely those areas where the former runs into difficulties.

We hope that the topics discussed in this book will prove useful to anyone who seeks new tools for robust and flexible stochastic modelling.

The story in a nutshell

1 From ordinary to rough differential equations

Rough path analysis can be viewed as a collection of smart estimates for differential equations of type

$$dy = V(y)\,dx \Longleftrightarrow \dot{y} = \sum_{i=1}^{d} V_i(y)\,\dot{x}^i.$$

Although a Banach formulation of the theory is possible, we shall remain in finite dimensions here. For the sake of simplicity, let us assume that the *driving signal* $x \in C^\infty\left([0,T], \mathbb{R}^d\right)$ and that the coefficients $V_1, \ldots, V_d \in C^{\infty,b}(\mathbb{R}^e, \mathbb{R}^e)$, that is bounded with bounded derivatives of all orders. We are dealing with a simple time-inhomogenous ordinary differential equation (ODE) and there is no question about existence and uniqueness of an $\mathbb{R}^e$-valued solution from every starting point $y_0 \in \mathbb{R}^e$. The usual first-order Euler approximation, from a fixed time-s starting point y_s, is obviously

$$y_t - y_s \approx V_i(y_s) \int_s^t dx^i.$$

(We now adopt the summation convention over repeated up–down indices.) A simple Taylor expansion leads to the following step-2 Euler approximation,

$$y_t - y_s \approx \underbrace{V_i(y_s) \int_s^t dx^i + V_i^k \partial_k V_j(y_s) \int_s^t \int_s^r dx^i dx^j}_{=\mathcal{E}(y_s, \mathbf{x}_{s,t})}$$

with

$$\mathbf{x}_{s,t} = \left(\int_s^t dx, \int_s^t \int_s^r dx \otimes dx \right) \in \mathbb{R}^d \oplus \mathbb{R}^{d \times d}. \tag{1}$$

Let us now make the following Hölder-type assumption: there exists c_1 and $\alpha \in (0,1]$ such that, for all $s < t$ in $[0,T]$ and all $i,j \in \{1,\ldots,d\}$,

$$(H_\alpha) : \left| \int_s^t dx^i \right| \vee \left| \int_s^t \int_s^r dx^i dx^j \right|^{1/2} \leq c_1 |t-s|^\alpha. \tag{2}$$

Note that $\int_s^t \int_s^r dx^i dx^j$ is readily estimated by $\ell^2 |t-s|^2$, where $\ell = |\dot{x}|_{\infty;[0,T]}$ is the Lipschitz norm of the driving signal, and so (H_α) holds, somewhat trivially for now, with $c_1 = \ell$ and $\alpha = 1$. [We shall see later that (H_α) also holds for d-dimensional Brownian motion for any $\alpha < 1/2$ and a random variable $c_1(\omega) < \infty$ a.s. provided the double integral is understood in the sense of stochastic integration. Nonetheless, let us keep x deterministic and smooth for now.]

It is natural to ask exactly how good these approximations are. The answer is given by *Davie's lemma* which says that, assuming (H_α) for some $\alpha \in (1/3, 1/2]$, one has the "step-2 Euler estimate"

$$|y_t - y_s - \mathcal{E}(y_s, \mathbf{x}_{s,t})| \le c_2 |t-s|^\theta$$

where $\theta = 3\alpha > 1$. The catch here is *uniformity*: $c_2 = c_2(c_1)$ depends on x only through the Hölder bound c_1 but not on its Lipschitz norm. Since it is easy to see that (H_α) implies

$$\mathcal{E}(y_s, \mathbf{x}_{s,t}) \le c_3 |t-s|^\alpha, \quad c_3 = c_3(c_1),$$

the triangle inequality leads to

$$|y_t - y_s| \le c_4 |t-s|^\alpha, \quad c_4 = c_4(c_1). \tag{3}$$

As often in analysis, uniform bounds allow for passage to the limit. We therefore take $x_n \in C^\infty([0,T], \mathbb{R}^d)$ with *uniform bounds*

$$\sup_n \left| \int_s^t dx_n^i \right| \vee \left| \int_s^t \int_s^r dx_n^i \, dx_n^j \right|^{1/2} \le c_1 |t-s|^\alpha$$

such that, uniformly in $t \in [0,T]$,

$$\left(\int_0^t dx_n^i, \int_0^t \int_0^r dx_n^i \, dx_n^j \right) \to \mathbf{x}_t \equiv \left(\mathbf{x}_t^{(1)}, \mathbf{x}_t^{(2)} \right) \in \mathbb{R}^d \oplus \mathbb{R}^{d \times d}.$$

The limiting object $\mathbf{x}$ is a path with values in $\mathbb{R}^d \oplus \mathbb{R}^{d\times d}$ and the class of $(\mathbb{R}^d \oplus \mathbb{R}^{d\times d})$-valued paths obtained in this way is precisely what we call the α-Hölder rough paths.[1]

Two important remarks are in order.

(i) The condition $\alpha \in (1/3, 1/2]$ in Davie's estimate is intimately tied to the fact that the condition (H_α) involves the first *two* iterated integrals.

(ii) The space $\mathbb{R}^d \oplus \mathbb{R}^{d\times d}$ is not quite the correct state space for $\mathbf{x}$. Indeed, the calculus product rule $d(x^i x^j) = x^i dx^j + x^j dx^i$ implies that[2]

$$\mathrm{Sym}\left(\int_0^t \int_0^r dx \otimes dx \right) = \frac{1}{2} \left(\int_0^t dx \right) \otimes \left(\int_0^t dx \right).$$

[1] To be completely honest, we call this a *weak geometric α-Hölder rough path.*
[2] $\mathrm{Sym}(A) := \frac{1}{2}(A + A^T)$, $\mathrm{Anti}(A) := \frac{1}{2}(A - A^T)$ for $A \in \mathbb{R}^{d\times d}$.

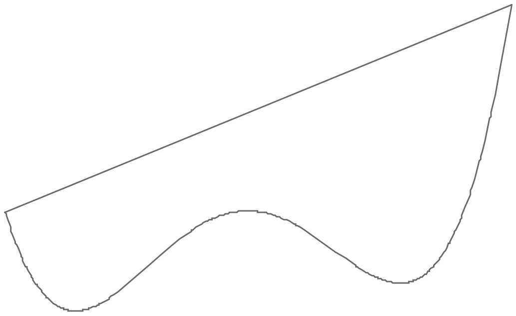

Figure 1. We plot $s \mapsto (x_s^i, x_s^j)$ and the chord which connects (x_0^i, x_0^j), on the lower left side, say, with (x_t^i, x_t^j) on the right side. The (signed) enclosed area (here positive) is precisely $\mathrm{Anti}(\mathbf{x}_t^{(2)})^{i,j}$.

This remains valid in the limit so that $\mathbf{x}(t)$ must take values in

$$\left\{\mathbf{x} = \left(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}\right) \in \mathbb{R}^d \oplus \mathbb{R}^{d\times d} : \mathrm{Sym}\left(\mathbf{x}^{(2)}\right) = \frac{1}{2}\mathbf{x}^{(1)} \otimes \mathbf{x}^{(1)}\right\}.$$

We can get rid of this algebraic redundancy by switching from $\mathbf{x}$ to[3]

$$\left(\mathbf{x}^{(1)}, \ \mathrm{Anti}(\mathbf{x}^{(2)})\right) \in \mathbb{R}^d \oplus so\,(d)\,.$$

At least for a smooth path $x(\cdot)$, this has an appealing geometric interpretation. Let $(x_\cdot^i, x_\cdot^j)$ denote the projection to two distinct coordinates (i,j); basic multivariable calculus then tells us that

$$\mathrm{Anti}(\mathbf{x}_t^{(2)})^{i,j} = \frac{1}{2}\left(\int_0^t \left(x_s^i - x_0^i\right) dx_s^j - \int_0^t \left(x_s^j - x_0^j\right) dx_s^i\right)$$

is the *area* (with multiplicity and orientation taken into account) between the curve $\{(x_s^i, x_s^j) : s \in [0,t]\}$ and the chord from (x_t^i, x_t^j) to (x_0^i, x_0^j). See Figure 1.

Example 1 Consider $d = 2$ and $x_n(t) = \left(\frac{1}{n}\cos\left(2n^2 t\right), \frac{1}{n}\sin\left(2n^2 t\right)\right) \in \mathbb{R}^2$. Then (H_α) holds with $\alpha = 1/2$, as may be seen by considering separately the cases where $1/n$ is less resp. greater than $(t-s)^{1/2}$. Moreover, the limiting rough path is

$$\mathbf{x}_t \equiv \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 & t \\ -t & 0 \end{pmatrix}\right), \tag{4}$$

since we run around the origin essentially $n^2 t/\pi$ times, sweeping out area π/n^2 at each round. □

---

[3] As will be discussed in Chapter 7, this is precisely switching from the step-2 free nilpotent Lie group (with d generators) to its Lie algebra.

We are now ready for the passage to the limit on the level of ODEs. To this end, consider $(y^n) \subset C([0,T], \mathbb{R}^e)$, obtained by solving, for each n, the ODE

$$dy^n = V(y^n)\, dx^n, \;\; y^n(0) = y_0.$$

By Davie's lemma the sequence (y_n) has a uniform α-Hölder bound c_4 and by Arzela–Ascoli we see that (y_n) has at least one limit point in $C([0,T], \mathbb{R}^e)$. Each such limit point is called a solution to the *rough differential equation* (RDE) which we write as

$$dy = V(y) d\mathbf{x}, \; y(0) = y_0. \tag{5}$$

The present arguments apply immediately for $V \in C^{2,b}$, that is bounded with two bounded derivatives, and more precisely for $V \in \mathrm{Lip}^{\gamma-1}, \gamma > 1/\alpha$, in the sense of Stein.[4] As in classical ODE theory, one additional degree of regularity (e.g. $V \in \mathrm{Lip}^{\gamma}, \gamma > 1/\alpha$) then gives uniqueness[5] and we will write

$$y = \pi_{(V)}(0, y_0; \mathbf{x})$$

for such a unique RDE solution. At last, it should not be surprising from our construction that the RDE solution map (a.k.a. *Itô–Lyons map*)

$$\mathbf{x} \mapsto \pi_{(V)}(0, y_0; \mathbf{x})$$

is continuous in $\mathbf{x}$ (e.g. under uniform convergence with uniform Hölder bounds).

Example 2 Assume $\mathbf{x}_t = \left(\int_0^t dx^i, \int_0^t \int_0^r dx^j dx^k \right)_{i,j,k \in \{1,\dots,d\}}$ with smooth x. Then

$$y = \pi_{(V)}(0, y_0; \mathbf{x})$$

is the classical ODE solution to $dy = V(y)\, dx, \; y(0) = y_0$. □

Example 3 Assume $\mathbf{x}$ is given by (4) and $V = (V_1, V_2)$. Then

$$y = \pi_{(V)}(0, y_0; \mathbf{x})$$

can be identified as the classical ODE solution to

$$dy = [V_1, V_2](y)\, dt$$

where $[V_1, V_2] = V_1^i \partial_i V_2 - V_2^i \partial_i V_1$ is the Lie bracket of V_1 and V_2. □

[4] Writing $\gamma = \lfloor \gamma \rfloor + \{\gamma\}$ with integer $\lfloor \gamma \rfloor$ and $\{\gamma\} \in (0,1]$ this means that V is bounded and has up to $\lfloor \gamma \rfloor$ bounded derivatives, the last of which is Hölder with exponent $\{\gamma\}$.

[5] With more effort, uniqueness can be shown under $\mathrm{Lip}^{1/\alpha}$-regularity.

Example 4 Assume $B = \left(B^1, \dots, B^d\right)$ is a d-dimensional Brownian motion. Define enhanced Brownian motion by

$$\mathbf{B}_t = \left(\int_0^t dB^i, \int_0^t B^j \circ dB^k \right)_{i,j,k \in \{1,\dots,d\}}$$

(where $\circ$ indicates stochastic integration in the Stratonovich sense). We shall see that $\mathbf{B}$ is an α-Hölder rough path for $\alpha \in (1/3, 1/2)$ and identify

$$Y_t(\omega) := \pi_{(V)}(0, y_0; \mathbf{B})$$

as a solution to the Stratonovich stochastic differential equation[6]

$$dY = \sum_{i=1}^d V_i(Y) \circ dB^i.$$

□

2 Carnot–Caratheodory geometry

We now try to gain a better understanding of the results discussed in the last section. To this end, it helps to understand the more general case of Hölder-type regularity with exponent $\alpha = 1/p \in (0,1]$. As indicated in remark (i), this will require consideration of more iterated integrals and we need suitable notation: given $x \in C^\infty\left([0,T], \mathbb{R}^d\right)$ we generalize (1) to[7]

$$\mathbf{x}_t := S_N(x)_{0,t} := \left(1, \int_0^t dx, \int_{\Delta^2_{[0,t]}} dx \otimes dx, \dots, \int_{\Delta^N_{[0,t]}} dx \otimes \dots \otimes dx \right), \tag{6}$$

called the *step-N signature* of x over the interval $[0,t]$, with values in

$$T^N\left(\mathbb{R}^d\right) := \mathbb{R} \oplus \mathbb{R}^d \oplus \left(\mathbb{R}^d\right)^{\otimes 2} \oplus \dots \oplus \left(\mathbb{R}^d\right)^{\otimes N}.$$

Observe that we added a zeroth scalar component in our definition of $\mathbf{x}_t$ which is always set to 1. This is pure convention but has some algebraic advantages. To go further, we note that $T^N\left(\mathbb{R}^d\right)$ has the structure of a (truncated) tensor-algebra with tensor-multiplication $\otimes$. (Elements with scalar component equal to 1 are always invertible with respect to $\otimes$.) Computations are simply carried out by considering the standard basis (e_i) of $\mathbb{R}^d$ as non-commutative indeterminants; for instance,

$$\left(a^i e_i\right) \otimes \left(b^j e_j\right) = a^i b^j (e_i \otimes e_j) \neq a^i b^j (e_j \otimes e_i).$$

---

[6] A drift term $V_0(y)\,dt$ can be trivially included by considering the time-space process (t, B).

[7] $\Delta^k_{[0,t]}$ denotes the k-dimensional simplex over $[0,t]$.

The reason we are interested in this sort of algebra is that the trivial

$$x_{s,t} \equiv (-x_s) + x_t = \int_s^t dx =: x_{s,t}$$

generalizes to

$$\mathbf{x}_{s,t} \equiv \mathbf{x}_s^{-1} \otimes \mathbf{x}_t = \left(1, \int_s^t dx, \int_{\Delta^2_{[s,t]}} dx \otimes dx, \ldots, \int_{\Delta^N_{[s,t]}} dx \otimes \cdots \otimes dx\right).$$

As a consequence, we have *Chen's relation* $\mathbf{x}_{s,u} = \mathbf{x}_{s,t} \otimes \mathbf{x}_{t,u}$, which tells us precisely how to "patch together" iterated integrals over adjacent intervals $[s,t]$ and $[t,u]$.

Let us now take on remark (ii) of the previous section. One can see that the step-N lift of a smooth path x, as given in (6), takes values in the free step-N nilpotent (Lie) group with d generators, realized as restriction of $T^N\left(\mathbb{R}^d\right)$ to

$$G^N\left(\mathbb{R}^d\right) = \exp\left(\mathbb{R}^d \oplus \left[\mathbb{R}^d, \mathbb{R}^d\right] \oplus \left[\mathbb{R}^d, \left[\mathbb{R}^d, \mathbb{R}^d\right] \oplus \ldots\right]\right) \equiv \exp\left(\mathfrak{g}^N\left(\mathbb{R}^d\right)\right)$$

where $\mathfrak{g}^N\left(\mathbb{R}^d\right)$ is the free step-N nilpotent Lie algebra and exp is defined by the usual power-series based on $\otimes$.

Example 5 $[N = 2]$ Note that $\left[\mathbb{R}^d, \mathbb{R}^d\right] = so(d)$. Then

$$\begin{aligned} &\exp\left(\mathbb{R}^d \oplus \left[\mathbb{R}^d, \mathbb{R}^d\right]\right) \\ = &\left\{\left(1, v, \frac{1}{2} v \otimes v + A\right) : v \in \mathbb{R}^d, \ A \in so(d)\right\} \end{aligned}$$

which is precisely the algebraic relation we pointed out in remark (ii) of the previous section. □

If the discussion above tells us that $T^N\left(\mathbb{R}^d\right)$ is too big a state space for lifted smooth paths, *Chow's theorem* tells us that $G^N\left(\mathbb{R}^d\right)$ is the correct state space. It asserts that for all $g \in G^N\left(\mathbb{R}^d\right)$ there exists $\gamma : [0,1] \to \mathbb{R}^d$, which may be taken to be piecewise linear such that $S_N(\gamma)_{0,1} = g$. One can then define the *Carnot–Caratheodory norm*

$$\|g\| = \inf\left\{\text{length }\left(\gamma|_{[0,1]}\right) : S_N(\gamma)_{0,1} = g\right\},$$

where the infimum is achieved for some Lipschitz continuous path $\gamma^* : [0,1] \to \mathbb{R}^d$, some sort of *geodesic path* associated with g. The *Carnot–Caratheodory distance* is then simply defined by $d(g,h) := \left\|g^{-1} \otimes h\right\|$. A Carnot–Caratheodory unit ball is plotted in Figure 2.

Example 6 Take $g = \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 & a \\ -a & 0 \end{pmatrix}\right) \in G^2\left(\mathbb{R}^2\right)$. Then γ^* is the shortest path which returns to its starting point and sweeps out area a. From basic isoperimetry, γ^* must be a circle and $\|g\| = 2\sqrt{\pi} a^{1/2}$. See Figure 3. □

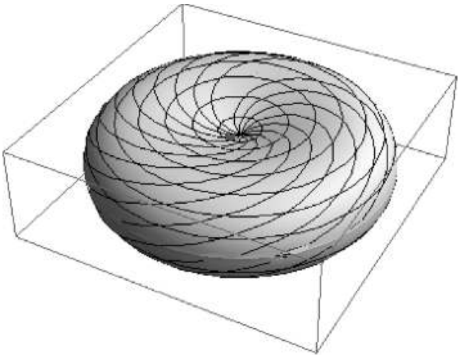

Figure 2. After identifying $G^2\left(\mathbb{R}^d\right)$ with the 3-dimensional Heisenberg group, i.e. $\left(\left(\begin{array}{c} x \\ y \end{array}\right), \left(\begin{array}{cc} 0 & a \\ -a & 0 \end{array}\right)\right) \equiv (x, y, a)$, we plot the (apple-shaped) unit-ball with respect to the Carnot–Caratheodory distance. It contains (and is contained in) a Euclidean ball.

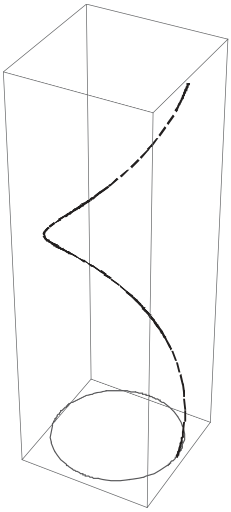

Figure 3. We plot the circle γ^*. The z-axis represents the wiped-out area and runs from 0 to a.