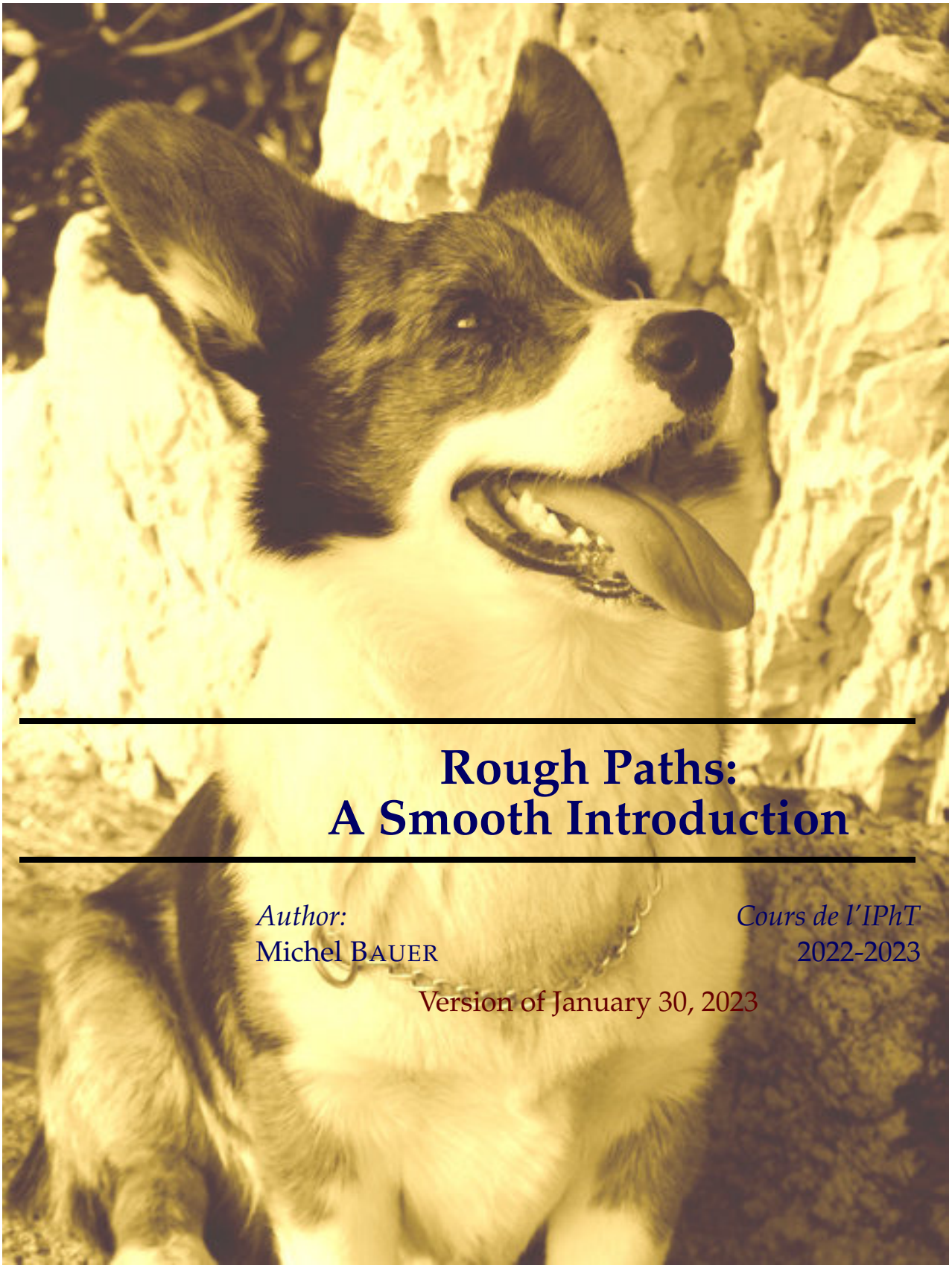
INSTITUT DE PHYSIQUE THÉORIQUE DE SACLAY

IPHT



Rough Paths: A Smooth Introduction

Author:
Michel BAUER

Cours de l'IPhT
2022-2023

Version of January 30, 2023

Rough paths:
A Smooth Introduction

Michel Bauer

January 30, 2023

Warning: The following notes are under construction.
They contain many inaccuracies, repetitions, and last but not least misprints. Every bit should be taken with a grain of salt.

Foreword

These notes are written as the basis of an eight hours lecture series given at the Institut de physique théorique de Saclay (IPhT).

The aims are quite modest, as the reader may check on the table of contents.

The first part contains few formally stated theorems and no proofs. There are many (maybe even too much) explicit computations on simple examples. We hope that this helps the reader to get a precise if non-rigorous perspective on the most basic rough paths concepts. Some mathematical elaborations are presented in the second part.

When I first heard introductory seminars or tried to read the few textbooks on the subject, I really felt at sea and it took me quite a few hours of intense efforts before something clicked. As usual, once it appended, I realized that the textbooks were in fact extremely clear and well-written and I could hardly remember why I did not grasp the basic definitions on the spot.

This is why I decided to offer a different starting point in these notes, with the hope that it might help some readers so they will waste less time than I did. I apologize to the others. Anyway, either to start or to get deeper, I can only recommend the textbooks [2] and (at a more slightly more advanced level) [3].

Another aim of these notes is to build on some physical intuition for certain of the phenomena and constructions encountered in rough paths theory. We shall try to reinterpret these features using the vocabulary of the renormalization group. The analogy is far from perfect, limited but nevertheless illuminating. One of the important simplification is that there are no anomalous dimensions in rough paths theory, which essentially deals with paths (!) that is with one-dimensional objects. The ideas of rough path theory can be generalized to fields, yielding to the theory of regularity structures, which has even closer links with renormalization theory, but which requires a much higher technical background (in distribution theory for instance) than rough paths theory. We shall not at all deal with regularity structures in these notes, but they have been used to tackle some long standing open (mathematical) questions, one example being the solution of the Kardar-Parisi-Zhang equation.

Rough paths theory has a deep interplay with continuous stochastic processes, and I this is apparent even at the modest level of these notes. Brownian motion, and its fractional cousins at a more advanced level, serve as a constant source of examples.

Contents

Part I

Basic Concepts

Introduction

It is a frequent situation in mathematics that certain notions are defined via a limiting procedure.

Think for instance as integrals as limits of sums: integrals are approximated by sums, sums involve only very basic algebra whereas integrals involve some analysis, so there is a price to pay. But once integrals are defined, they can be used in the opposite direction, to approximate sums. And integrals are more flexible tools than sums, mainly because the possibility of changing variables.

Another example is the deep relationship between random walks and Brownian motion. Among the multitude of definition of Brownian motion, quite a few are via the approximation by random walks. Again, random walks are rather concrete objects (in particular the simple symmetric random walk) whose definition requires minimal mathematical investment, whereas Brownian motion requires a more involved mathematical setting. But Brownian motion is nevertheless a sharp and invaluable tool to study (asymptotic properties of) random walks, a salient example being the law of the iterated logarithm. And again Brownian motion, the limiting object, is a more flexible tool than random walks. It is fully characterized by a few axioms in which it is not so easy to detect the relationship to random walks, though ignoring this relationship would really be a sad omission.

It is to be noted that some random walks are not good approximations to Brownian motion (or vice versa). Sometimes, they approximate other processes, like Levy processes. But there is always a flavor of what physicists call universality: a huge zoo of random walks and a more manageable menagerie of processes.

Rough paths with words

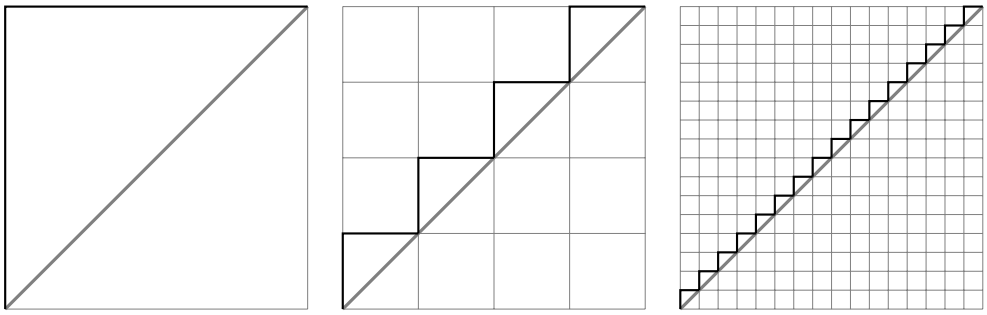
Rough path theory can also be viewed as such an interplay involving approximations, limits and the like. Just a Brownian motion or Levy processes, rough paths can be defined axiomatically without talking of any approximation scheme. And just as Levy processes theory, rough path theory tames a wildlife. This time it is about the possible limiting behaviors of paths and their iterated integrals.

The basic observation is that whereas the iterated integrals of a piecewise smooth[1] function are defined without ambiguity and can be recovered from the

[1]Contrary to standard practice, is these notes the term "smooth" refers to differentiable, not

function, this can be completely scrambled by taking limits even when they exist.

Let us illustrate this on a trivial example: the approximation of the diagonal of the unit square in the plane by a path on the square lattice, as pictured below with mesh $1, 1/4, 1/16$:



The lattice path gets closer and closer to the diagonal, but its length remains 2, and in particular does not converge to the length of the diagonal, $\sqrt{2}$. Though the analogy is limited, the rough paths philosophy would be to take as the limiting object of the lattice paths not only the diagonal of the unit square, but also to keep track of the anomalous scale. Rough path theory would also put this additional information to good use.

Returning to the general setting, the failure of the limit of functions/paths to describe faithfully what is going on has two main manifestations. First, good approximations to a functions can lead to different approximations of its iterated integrals, even if one is approximating smooth objects. Second, functions and their line integrals may converge, but there may be no direct integration theory to define the iterated integrals of the limiting function because it is too irregular.

The axiomatic definition of rough paths is a way to take those two elementary observations into account. Basically, one needs to break the strict bond between functions and their iterated integrals and give those some freedom. So a rough path is a collection of objects, with a standard function/path as its most basic object, and some substitutes for its iterated integrals. Those substitutes have a part of arbitrariness but they are constrained by some combinatorial and analytical conditions that reflect those of bona-fide iterated integrals. The combinatorial conditions are essentially Chasles relation in disguise. Another way to say the same thing, the combinatorial conditions ensure the closure of the flow in the context of solutions of differential equations. The analytical conditions endow rough paths with a topology which allows to compare them to various objects, discrete or continuous.

It turns out that the combinatorial plus analytical conditions ensure that in fact only a finite number of iterated integrals need to be specified, and then the others are fully determined. This leads to a first analogy with renormalization in quantum field theory. In a renormalizable theory, only finite number of conditions are needed to eliminate infinities and ambiguities, making all correlation functions finite. There are (at least) two important differences. First, in quantum field

necessarily infinitely differentiable.

theory a finite number of constants need to be specified, whereas in rough paths theory one needs a finite number of functions. Second, whereas renormalization is needed because naive computations lead to infinities, rough paths theory is needed because naive computations lead to undefined (but not systematically unbounded) results.

Before we give a more detailed and technical motivation for rough paths via differential equations, let us note that scale invariance plays a fundamental role in rough path theory, but there are no anomalous dimensions. Regularity structures, a subject we shall not elaborate on, are a generalization of rough paths that allow to tackle more realistic problems involving true renormalization. This approach culminates in a rigorous treatment of a number of singular stochastic partial differential equations, like the famous Kardar-Parisi-Zhang equation.

Rough paths theory was built in the 1990's mainly under the impulse of Terry Lyons. A friendly but serious introduction to rough path is [2], to which we refer the reader for a deeper treatment of the subject. A more difficult but more complete reference is [3].

CHAPTER 1

Motivations

The following discussion is a bit technical but stresses a few important points of the rough paths philosophy. We introduce the notion of controlled differential equations. If one tries to adapt the traditional Euler scheme for solving ordinary differential equations to this new situation, problems arise: it may happen that the Euler scheme does not apply naively, or that it simply fails. An attempt to improve it quickly leads to a new structure, that of a rough path.

As we mentioned briefly, among the origins of rough paths theory is the behavior of iterated integrals: if $Q := (Q^i)_{i\in[\![1,n]\!]}$ is a collection of smooth functions from $[a,b]\text{to}\mathbb{R}$, the iterated integrals of Q are the tensors $\int_{a<s_1<\cdots<s_k} dQ^{i_1}_{s_1}\cdots dQ^{i_k}_{s_k}$ for $k\in\mathbb{N}^*$. The point is that if $Q(\varepsilon)_{\varepsilon>0}$ is a family of such maps and there is a limit $Q := \lim_{\varepsilon\downarrow 0} Q(\varepsilon)$ exists in some appropriate sense, several (related) pathologies are possible. First, it may happen that the limit Q is smooth but the limit of iterated integrals does not exist, or does not coincide with the iterated integrals of the limit. Second, the iterated integrals might have a limit, but Q itself is not smooth enough for its iterated integrals to make any apriori sense. We shall see a number of examples in what follows. Controlled differential equations will quickly confront us with those questions.

1.1 Controlled differential equations

Even if the name itself is not well-known, most of us are probably familiar with the concept of controlled differential equations. They make their appearance when a system does not respond directly to the passage of time, but to some auxiliary time dependent quantities taken as input. A controlled differential equation has the following generic form

$$dY_t = V(Y_t, X_t)\, dX_t \text{ for } t\in[a,b] \text{ with initial condition } Y_a = y_a$$

where $X := (X_t)_{t\in[a,b]}$ is a given path in some vector space E, $Y := (Y_t)_{t\in[a,b]}$ is some unknown path in another vector space F and, for each $(y,x)\in F\times E$, $V(y,x)$ is a linear map from E to F. It is often the case that $V(y,x)$ is not defined globally but

only in a neighborhood of (y_a, x_a) and then we look for at least a local solution for $t \in [a, a+\delta]$.

For the time being, we have given no meaning to controlled differential equations. If the path X is smooth, i.e. if the derivative $\frac{dX_t}{dt}$ is well-defined for $t \in [a,b]$ we can *interpret* the above controlled differential equation to fall back to the framework of ordinary differential equations by setting $\tilde{V}(Y_t, t) := V(Y_t, X_t)\frac{dX_t}{dt}$ an turning the controlled differential equation into

$$dY_t = \tilde{V}(Y_t, t)\, dt \text{ for } t \in [a,b] \text{ with initial condition } Y_a = y_a.$$

Let us stress that this is an *interpretation*: we do not relate two meaningful things, but turn an apriori meaningless one to a meaningful one. Rough paths theory gives a precise meaning to controlled differential equations for sources X that may be far from smooth. It turns out that for smooth sources rough paths theory is consistent (it better be!) with the above interpretation.

We can and shall often, introduce local coordinates, say $E \cong \mathbb{R}^n$, $F \cong \mathbb{R}^d$ so that the controlled differential equation rewrites

$$\text{For } \mu = 1, \cdots, d: \ dY_t^\mu = \sum_{i=1}^{n} V_i^\mu(Y_t^1, \cdots, Y_t^d, X_t^1, \cdots, X_t^n)\, dX_t^i.$$

One example that is well-known to physicists is when $n = 2, d = 1$ and $X = \begin{pmatrix} B_t \\ t \end{pmatrix}_{t \geq 0}$ where B is a Brownian motion and $V(y,x) = (\sigma(y), \nu(y))$ so that,

$$dY_t = \nu(Y_t)\, dt + \sigma(Y_t) dB_t \text{ or } \dot{Y}_t = \nu(Y_t) + \sigma(Y_t)\xi_t \text{ in physicists notation,}$$

where ξ_t is a white noise (hence a distribution, not a function). In this context, the name "stochastic differential equation" is used in place of controlled differential equation. To make sense of this diffusion equation, one turns it into an integral equation:

$$Y_t = y_0 + \int_0^t \nu(Y_t)\, dt + \int_0^t \sigma(Y_s)\, dB_s.$$

Then a solution to the diffusion equation is a process Y such that first both integrals on the right-hand side are well-defined, and second such that the two sides turn out to be equal. Several mathematical remarks are in order. Assume that σ is not constant (i.e. independent on the position y). First, a new integration theory, stochastic integration, has to be available. Second, the stochastic integral theory is defined as a limit of discrete sums but the limit depends on conventions: mathematicians usually work with the Itō convention, when physicists tend to favor the Stratanovich convention. And last but not least, the stochastic integral is not defined pathwise, but in mean square or in probability: informally, the statement is that when the mesh is small, the discretized sum has a probability close to 1 to be close to the integral – this is far from saying that sample by sample the discretized sum goes to the integral when the mesh goes to 0. These subtleties cause

little or no trouble usually. Stochastic calculus (à la Itō or Stratanovich) justifies blind manipulation, and numerical computations are not really sensitive to the problem because one works usually only with regular time steps – adaptative methods are already harder to justify.

The intrinsic study of stochastic differential equations on manifolds is notably difficult, one of the reason being the absence of a pathwise definition of the stochastic integral. Let us clarify this point in the general context of control. Instead of vector spaces E and F, practical applications may force to consider manifolds M and N, a given curve X on M parameterized by time and a family $V(y,x)$, $x \in M, y \in N$ of linear maps: for given $(x,y) \in M \times N$, $V(y,x)$ is a linear map taking a tangent vector to M at x as input and yielding a vector tangent to N at y as output. The form of the equation is unchanged:

$$dY_t = V(Y_t, X_t)\, dX_t \text{ for } t \in [a,b] \text{ with initial condition } Y_a = y_a,$$

and we can always take local coordinates, in $\mathbb{R}^n$ for M and $\mathbb{R}^d$ for N. It this geometric context, it is crucial however that the meaning given to a controlled differential equation is intrinsic. In terms of local coordinates, the solutions over different coordinate patches should knit together nicely. Rough paths theory is successful in that aspect to.

The notion of pathwise versus non-pathwise solution is important in that rough paths theory allows to solve stochastic differential equations via a pathwise procedure. But the general context if that one is given a single X (not a sample space of Xs) so there is no choice but to work pathwise.

In the next section we specialize to the one-dimensional setting and illustrate a number of issues on a very simple example.

1.2 The generic one-dimensional setting

Think of making sense, or solving numerically, the equation

$$dY_t = V(Y_t)dX_t \text{ for } t \in [a,b] \text{ with initial condition } Y_a = y_a,$$

where V is some smooth function, X is a given real source defined on $[a,b]$ and Y is a real unknown function.

This is called a controlled differential equation because the variations of Y respond to those of a function X. When $X_t := t$ we recover a standard ordinary differential equation when the variations of Y respond to the passage of time. When X is a (piecewise) smooth function of t, it is natural to interpret $dY_t = V(Y_t)dX_t$ as $\frac{dY_t}{dt} = V(Y_t)\frac{dX_t}{dt}$ which is a special case of the familiar $\frac{dY_t}{dt} = U(Y_t, t)$.

We return to the controlled setting.

If f is a smooth function, we expect naively that $df(Y_t) = f'(Y_t)V(Y_t)dX_t$ which, assuming the integral to make sense, should be a tantamount for $f(Y_t) = f(Y_s) + \int_s^t f'(Y_u)V(Y_u)dX_u$. Thinking of t as close to s, the Euler scheme approxi-

mates $f'(Y_u)V(Y_u)$ on the interval $[s,t]$ by $f'(Y_s)V(Y_s)$ yielding

$$\begin{aligned} f(Y_t) &= f(Y_s) + \int_s^t f'(Y_s)V(Y_s)dX_u + \text{error} \\ &= f(Y_s) + f'(Y_s)V(Y_s)\int_s^t dX_u + \text{error} \\ &= f(Y_s) + f'(Y_s)V(Y_s)(X_t - X_s) + \text{error}. \end{aligned}$$

We've made some hair-splitting, the point being that some definitions of integrals do not allow to factor out constants,[1] and that as we have said nothing about the regularity of X, the "obvious" $\int_s^t dX_u = X_t - X_s$ is maybe not so obvious. This is what is produced if dX_u is interpreted as an exact differential, and also what discretization suggests. Hence this interpretation is not challenged in the following.

The idea is then to propagate the solution from the initial to the final time by small steps, with the hope that the errors do not accumulate to a sizable quantity.

Let us check this idea on one of the simplest examples.

Example 1.1. The case when $V(y) := y$. Specializing the Euler scheme to this case, we write $Y_t \simeq Y_s(1 + (X_t - X_s))$. Thus if $\Delta : a = t_0 < t_1 < \cdots < t_n = b$ is a subdivision of $[a,b]$ we define Y^Δ at subdivision points by

$$Y^\Delta_a = y_a \quad Y^\Delta_{t_{m+1}} = Y^\Delta_{t_m}(1 + X_{t_{m+1}} - X_{t_m}) \text{ for } 0 \le m \le n-1,$$

and extend the definition of Y^Δ_t for $t \in [a,b]$ by linear interpolation for instance. We assume that X is continuous, so that if $s,t \in [a,b]$ with $|t-s|$ small enough, say $|t-s| \le \delta$ then $|X_t - X_s| \le 1/2$. If the mesh of the subdivision Δ , $\text{mesh}(\Delta) := \max_{0\le m\le n-1} t_{m+1} - t_m$, is $\le \delta$ we can take logarithms: $\log \frac{Y^\Delta_{t_m}}{y_a} = \sum_{l=0}^{m_1} \log(1 + X_{t_{m+1}} - X_{t_m})$. Using the elementary bound $-x^2 \le \log(1+x) - x \le -x^2/3$ for $|x| \le 1/2$ we infer that $\log \frac{Y^\Delta_b}{y_a} - (X_b - X_a) \in [-Q_\Delta, -Q_\Delta/3]$ where $Q_\Delta := \sum_{m=0}^{n-1}(X_{t_{m+1}} - X_{t_m})^2$, the quadratic variation of X along Δ. The fate of Y^Δ as $\text{mesh}(\Delta) \downarrow 0$ is clear if X has vanishing 2-variation on $[a,b]$ which by definition means that Q_Δ goes to 0 at small mesh.[2] Then there is a limiting Y which is $Y_t = y_a e^{X_t - X_a}$.

It is a (not so well-known) theorem that, as X is assumed to be continuous, there is a sequence $(\Delta_k)_{k\in\mathbb{N}}$ of finer and finer partitions of $[a,b]$ such that Q^{Δ_k} goes to 0 at large k. This has two consequences. First, we could decide to restrict to such sequences, but this would have major drawbacks because they need a detailed knowledge of X to be constructed whereas we want an algorithm that works for arbitrary partitions in the small mesh limit. Second, if arbitrary partitions are to be considered, there is a dichotomy: either Q_Δ goes to 0 at small

[1] A prominent example is the Skorokhod stochastic integral.

[2] This happens in particular if $(X_t - X_s)^2$ is uniformly a $o(t-s)$, which occurs for instance if there is a $\sigma > 0$ and a constant K such that $|X_t - X_s| < K|t-s|^{1/2+\sigma}$ for $s,t \in [a,b]$.

mesh, or it diverges[3], in which case the discretization leads to a dead end. We may expect that the situation will not be better when V is generic!

We take this opportunity of discuss an important test case for rough paths ideas, Brownian motion. If $(\Delta_k)_{k\in\mathbb{N}}$ is a sequence of finer and finer partitions of $[a,b]$ and X is a Brownian motion then, with probability 1, Q^{Δ_k} converges to $b-a$ at large k. This is the usual quadratic variation of Brownian motion, and this result is at the heart of Itō's stochastic calculus.

Example 1.2. The case when $V(y) := y$ (continued). If $(\Delta_k)_{k\in\mathbb{N}}$ is a given sequence of finer and finer partitions of $[a,b]$ and X is a Brownian motion then, with probability 1, the approximation Y^{Δ_k} based on the Euler scheme above approaches a limiting Y which is $Y_t = y_a e^{X_t - X_a - (t-a)/2}$. Indeed, we can refine the above inequality for the log to: $-x^2/(2(1-\varepsilon)) \leq \log(1+x) - x \leq -x^2/(2(1-\varepsilon))$ for $|x| \leq \varepsilon$. By assumption $\delta_k := \mathrm{mesh}(\Delta_k)$ goes to 0 at large k. Then $\varepsilon_k := \sup_{s,t\in[a,b],\,|t-s|\leq\delta_k} |X_t - X_s|$ goes to 0 at large k as well(Brownian motion is continuous!) and $\log \frac{Y_b^{\Delta_k}}{y_a} - (X_b - X_a) \in [-Q^{\Delta_k}/(2(1-\varepsilon_k)), -Q^{\Delta_k}/(2(1+\varepsilon_k))]$. This settles the case when $t = b$. We leave it to the reader to make the obvious modifications needed to deal with a generic $t \in [a,b]$.

Brownian motion could seem to be a counter-example to the above statement that 0 is the only possible finite limit for 2-variation. It is not, and the subtlety is the following. It is another (again not so well-known) theorem that if X is a Brownian sample, there is, with probability 1, a sequence $(\Delta_k)_{k\in\mathbb{N}}$ of finer and finer partitions of $[a,b]$ such that Q^{Δ_k} goes to $+\infty$ at large k.[4]

The situation here is not too bad because when we do numerical computation we usually fix a single partition with small mesh (or a few partitions to test stability) and use them for a number of samples. The use of adaptative methods is already more questionable but can be dealt with. However, let us stress that the rough paths philosophy insists apriori that rough paths theory should work pathwise. That is, given the Brownian sample X we want some Y to exist such that for every partition of sufficiently small mesh Y^Δ is close to Y. This could be judged as too stringent a condition but, as we shall see below, it can be fulfilled with some modification of the Euler scheme. Moreover, we have not made any mention whatsoever of conventions for stochastic integrals in the above discussion. The result of our naive approach, that the solution of $dY_t = Y_t\, dX_t$ is $Y_t = y_a e^{X_t - X_a - (t-a)/2}$ for X a Brownian motion, should look strange: we have automatically (should we say automagically?) implemented the Itō convention

[3] Remember that divergence is just the negation of convergence, we do not mean "diverges to infinity" in general.

[4] Thus if the sequence of partitions is given in advance and used for each and every Brownian sample then the 2-variation behaves well. However, if the we give the Brownian sample in advance we can tailor partitions for which the 2-variation is as small or as large as we wish.

even if we dealt with classical manipulations of differentials. We shall see below that the rough paths philosophy revives the possibility of several different conventions.

Then again, how should we deal in general with sources X that do not have vanishing 2-variation ? The clue is a closed (but implicit) formula for the error: applying to $f'V$ the formula we had for f, we get $(f'V)(Y_u) = (f'V)(Y_s) + \int_s^u (f'V)'(Y_v)V(Y_v)dX_v$ leading to

$$f(Y_t) = f(Y_s) + \int_s^t \left((f'V)(Y_s) + \int_s^u (f'V)'(Y_v)V(Y_v)dX_v \right) dX_u.$$

This formula has (at least) two useful applications. Before turning to those, let us mention that we could iterate again, this time using a representation of $(f'V)'V)(Y_v) - (f'V)'V)(Y_s)$ as an integral and so on, a close analog of the Born expansion in quantum mechanics.[5]

The first application is that the error in our previous computation is

$$\text{error } = \int_s^t \left(\int_s^u (f'V)'(Y_v)V(Y_v)dX_v \right) dX_u.$$

Very naively, we expect that this error (an integral involving the data over a triangle) to be of the order of the square of the term retained in the approximation (an integral involving the data on a segment). In the case when X is smooth, it is clear that the line integral is $O(t-s)$ and the surface integral is $O((t-s)^2)$ and the accumulated error over a finite interval is of order the mesh of the set of points chosen to interpolate between the initial and the final point, leading to a convergent approximation at small mesh.

As a second application we may approximate $(f'V)'(Y_v)V(Y_v)$ on the interval $[s,u]$ by $(f'V)'(Y_v)V(Y_v)$, leading to

$$\begin{aligned} f(Y_t) &= f(Y_s) + (f'V)(Y_s)(X_t - X_s) + ((f'V)'V)(Y_s) \int_s^t \left(\int_s^u dX_v \right) dX_u + \text{error} \\ &= f(Y_s) + (f'V)(Y_s)(X_t - X_s) + ((f'V)'V)(Y_s) \int_s^t (X_u - X_s)dX_u + \text{error}. \end{aligned}$$

Getting to the next order in the Born expansion would show that when X is smooth the error is uniformly $O((t-s)^3)$, leading to an improved convergence, the accumulated error over a finite interval being of the order of the square of the mesh. All this is well-known, but our whole point is to deal with the case when X is not smooth...

It is again tempting to make the obvious guess that $\int_s^t (X_u - X_s)dX_u = (X_t - X_s)^2/2$, interpreting $(X_u - X_s)dX_u$ as an exact differential. Let us see where it leads us, i.e. explore the behavior of Y^Δ. We do this again for our simple example.

[5] And we use the name "Born expansion" for the procedure in the sequel.

Example 1.3. The case when $V(y) := y$ (continued). Using the second order Born approximation, we are led to

$$Y^\Delta_a = y_a \quad Y^\Delta_{t_{m+1}} = Y^\Delta_{t_m}(1 + (X_{t_{m+1}} - X_{t_m}) + (X_{t_{m+1}} - X_{t_m})^2/2) \text{ for } 0 \le m \le n-1.$$

If $s, t \in [a, b]$ with $|t - s|$ small enough, say $|t - s| \le \delta$ then $|X_t - X_s| \le 2$. We can take the logarithm again and bound with $|\log(1 + x + x^2/2) - x| \le e^{|x|}|x|^3/6$ for $x \ge -2$ to get that $|\log \frac{Y^{\Delta_k}_{t_m}}{y_a} - (X_{t_m} - X_a)| \le C^\Delta e^2/6$ if $\text{mesh}(\Delta) \le \delta$ where $C^\Delta := \sum_{m=0}^{n-1} |X_{t_{m+1}} - X_{t_m}|^3$ is the cubic variation of X along Δ. This time the fate of Y^Δ as $\text{mesh}(\Delta) \downarrow 0$ is clear if X has vanishing 3-variation on $[a, b]$ which by definition means that C^Δ goes to 0 at small mesh.[6] Then there is a limiting Y which is $Y_t = y_a e^{X_t - X_a}$. Thus, if X has vanishing 3-variation, using a second order Born expansion and a naive integration formula $\int_s^t (X_u - X_s)dX_u = (X_t - X_s)^2/2$ we recover the naive solution of $dY_t = Y_t\, dX_t$ namely $Y_t = y_a e^{X_t - X_a}$.

Let us see some consequences when X is a Brownian motion. Then X has vanishing 3-variation with probability 1 (in the strong, pathwise, sense: we can choose the sample and then choose any subdivision with small mesh to approach the 3-variation). It is reassuring that implementing the Stratanovich convention for the integral $\int_s^t (X_u - X_s)dX_u$ i.e. setting its value to $(X_t - X_s)^2/2$ (for which there is no pathwise justification via a discretization) the use of the second order Born approximation leads pathwise to the Stratanovich solution of $dY_t = Y_t\, dX_t$.

Notice that "integration" as the operation "inverse of differentiation" was almost a definition before Riemann (though Archimedes already used discretization to compute areas and volumes). However, 175 years later we recognize that this fact, the fundamental theorem of calculus, is a consequence of an independent definition of the integral via discrete approximations. Moreover, it is easy to generalize the Born expansion to the case when $X = (X^i)_{i=1,n}$ (and Y) have several components, see D.2. Instead of one double integral, the above formula would involve a linear combination of $\int_s^t (X^i_u - X^i_s)dX^j_u$ with possibly different components i, j of X and then no exact differential miracle could save us from the boredom of really dealing with another definition of the iterated integral. Itō integration gives a definition (though not a pathwise one) of $\int_s^t (X_u - X_s)dX_u$ via a discretization. Let us see where this definition, when applied to the second order Born approximation, leads us to. But before that, we propose the following exercise to the reader

Exercise 1.4. Check that the naive Born expansion (valid if X is differentiable) to k^{th} order for $dY_t = Y_t\, dX_t$ is

$$\begin{aligned} Y_t &= Y_s \left(1 + \int_{s \le u_1 \le t} dX_{u_1} + \cdots + \int_{s \le u_1 \le \cdots \le u_k \le t} dX_{u_1} \cdots dX_{u_k}\right) \\ &\quad + \int_{s \le u_1 \le \cdots \le u_{k+1} \le t} Y_{u_1} dX_{u_1} \cdots dX_{u_{k+1}}. \end{aligned}$$

[6] This happens in particular if $|X_t - X_s|^3$ is uniformly a $o(t - s)$, which occurs for instance if there is a $\sigma > 0$ and a constant K such that $|X_t - X_s| < K|t - s|^{1/3+\sigma}$ for $s, t \in [a, b]$.

Check that the k^{th} iterated integral is, if X is differentiable, $(X_t - X_s)^k/k!$.

Check that if X is the (pointwise) limit of a sequence of differentiable maps, the limit of the k^{th} is again $(X_t - X_s)^k/k!$.[7]

Check that under the naive assumption that this formula holds also for a less regular source, the corresponding k^{th} order Euler scheme leads to a convergent procedure if X has vanishing $(k+1)$-variation, and the corresponding solution is the naive solution of $dY_t = Y_t\, dX_t$ namely $Y_t = y_a e^{X_t - X_a}$.

Check (or accept) that if X has vanishing $(k+1)$-variation then all its higher variations vanish as well and infer that the procedure is stable under a change of the order of the Euler scheme.

Example 1.5. The case when $V(y) := y$ (continued). We suppose that X is a Brownian motion and we use the second order Born approximation, but this time with the Itō convention $2\int_s^t (X_u - X_s)dX_u = (X_t - X_s)^2 - (t-s)$. We are led to $Y_a^\Delta = y_a$ and

$$Y^\Delta_{t_{m+1}} = Y^\Delta_{t_m}(1 + (X_{t_{m+1}} - X_{t_m}) + (X_{t_{m+1}} - X_{t_m})^2/2 - (t_{m+1} - t_m)/2) \text{ for } 0 \le m \le n-1.$$

We observe that

$$\log(1 + x + x^2/2 + r/2) = x + r/2 - \frac{x^3}{6}(1 + c_1(x)) - \frac{xr}{3}(1 + c_2(x)) - \frac{r^2}{8}(1 + c_3(x, r)),$$

where c_1, c_2, c_3 vanish at the origin and are analytic close to the origin. We infer that for small enough x, y we have $|\log(1 + x + x^2/2 + r/2) - x - r/2| \le |x|^3/3 + |xr| + r^2/4$. Using that $\sum_{m=0}^{n-1} |X_{t_{m+1}} - X_{t_m}|^3$, $\sum_{m=0}^{n-1} |X_{t_{m+1}} - X_{t_m}|(t_{m+1} - t_m)$ and $\sum_{m=0}^{n-1}(t_{m+1} - t_m)^2$ are small if Δ has a small mesh (the first is because Brownian motion has vanishing 3-variation, the second because Brownian motion is continuous), we infer the there is a limiting Y, namely $Y_t = y_a e^{X_t - X_a - (t-a)/2}$.

It is reassuring again that implementing the Itō convention for the integral $\int_s^t (X_u - X_s)dX_u$ i.e. setting its value to $(X_t - X_s)^2/2 - (t-s)/2$, the use of the second order Born approximation leads pathwise to the Itō solution of $dY_t = Y_t\, dX_t$.

To summarize this slightly lengthy discussion,

- The first order Euler scheme for $dY_t = Y_t\, dX_t$
  - Yields a limiting Y if X has vanishing 2-variation.

[7] This holds even if X is irregular enough that no known procedure allows to make sense of the integral directly. Thus the naive assumption in the next question is natural somehow. Nevertheless, pointwise convergence is not the only way to approach X. Moreover, as already mentioned, when we turn to integrals involving several components of a path, it may happen that the limiting X is regular enough for a direct definition of the integral, but which is not the limit of the integrals of the approximations, even if the convergence is better than pointwise, see Section 3.1 and D.1for an illustration.

  - Breaks down if X does not have vanishing 2-variation, tough for Brownian motion, renouncing to a pathwise procedure, is leads to the Itō solution.

- The second order Euler scheme for $dY_t = Y_t\, dX_t$
  - Yields the Stratanovich solution if X has vanishing 3-variation (in particular if X is a Brownian motion) and if the integral $\int_s^t (X_u - X_s) dX_u$ is taken in the Stratanovich sense.
  - Yields the Itō solution if X has vanishing 3-variation (in particular if X is a Brownian motion) and if the integral $\int_s^t (X_u - X_s) dX_u$ is taken in the Itō sense.

Thus, even if we would restrict our attention to Brownian motion, the second order Euler scheme is better behaved than the first order Euler scheme: it leads to a pathwise solution and leaves room for the different conventions in a consistent way.

For Brownian motion, the Itō and Stratanovich convention are natural and the most used in practice, but they are certainly not the only ones. And if an X is given, of vanishing 3-variation for instance, but yet with important short distance "fluctuations", there are few clues to decide what $\int_s^t (X_u - X_s) dX_u$ should be. As we have observed at the beginning of this chapter, approximating X with smooth paths does not lead to an unambiguous definition (if any) of $\int_s^t (X_u - X_s) dX_u$. So why no give it a name, i.e. set "$\int_s^t (X_u - X_s) dX_u := \mathbb{X}_{s,t}$" and see what happens?

Example 1.6. The case when $V(y) := \lambda y$. We introduce a parameter, λ, for convenience, so that the second order Born approximation to go from time s to time u is $Y_u \simeq Y_s(1 + \lambda(X_u - X_s) + \lambda^2 \mathbb{X}_{s,u})$. On the other hand, adding another point in the game, t, we may go from s to u via t, leading to $Y_u \simeq Y_s(1 + \lambda(X_t - X_s) + \lambda^2 \mathbb{X}_{s,t})(1 + \lambda(X_u - X_t) + \lambda^2 \mathbb{X}_{t,u})$. How do these two approximation compare? The difference between the second and the first is seen to yield $1 - 1 = 0$ at order λ^0, $(X_u - X_s) - ((X_u - X_t) + (X_t - X_s)) = 0$ at order at order λ^1. Then come

$$\mathbb{X}_{s,u} - (\mathbb{X}_{s,t} + \mathbb{X}_{t,u} + (X_t - X_s)(X_u - X_t)) \text{ at order } \lambda^2,$$

$$-(\mathbb{X}_{s,t}(X_u - X_t) + (X_t - X_s)\mathbb{X}_{t,u}) \text{ at order } \lambda^3,$$

and $-\mathbb{X}_{s,t}\mathbb{X}_{t,u}$ at order λ^4. It is readily checked that if the Itō or Stratanovich interpretations of $\int_s^t (X_u - X_s) dX_u$ are substituted for $\mathbb{X}_{s,t}$ the term of order λ^2 vanishes identically. It is also easy to relate this vanishing to Chasles' relation, or, what amounts to the same in the case at hand, to the closure of the flow in the putative solution of the controlled differential equation. Doing the same substitutions in the higher order terms in λ does not yield 0 but the result at order λ^3 is negligible when the times steps are small and X is such that "$(X_t - X_s) = o((t-s)^{1/3})$ in which case $\int_s^t (X_u - X_s) dX_u$ (Itō or Stratanovich) is $o((t-s)^{2/3}$. Then the λ^4 contribution is harmless.

To summarize, if $(X, \mathbb{X})$ is chosen in such a way that

Combinatorial condition: $\mathbb{X}_{s,u} - (\mathbb{X}_{s,t} + \mathbb{X}_{t,u} + (X_t - X_s)(X_u - X_t)) = 0$,

Regularity conditions: $(X_t - X_s) = o((t-s)^{1/3})$ and $\mathbb{X}_{s,t} = o((t-s)^{2/3})$,

the second order Euler scheme for $dY_t = \lambda Y_t dX_t$ with the interpretation $\int_s^t (X_u - X_s) dX_u := \mathbb{X}_{s,t}$ will be convergent.

Exercise 1.7. Show that the cocycle relation $\mathbb{X}_{s,u} - (\mathbb{X}_{s,t} + \mathbb{X}_{t,u} + (X_t - X_s)(X_u - X_t)) = 0$ is enough to ensure the closure of the flow for the second order Euler scheme associated to the general equation $dY_t = V(Y_t)dX_t$, i.e. $Y_t \simeq Y_s + V(Y_s)(X_t - X_s) + (V'V)(Y_s)\mathbb{X}_{s,t}$.

Exercise 1.8. Show that if X is smooth, $\mathbb{X}_{s,t} := \int_s^t (X_u - X_s) dX_u$ (note what defines what here) satisfies automatically the cocycle relation $\mathbb{X}_{s,u} - (\mathbb{X}_{s,t} + \mathbb{X}_{t,u} + (X_t - X_s)(X_u - X_t)) = 0$. Infer that if $X(\varepsilon)_{\varepsilon>0}$ is a family of such maps and there is a limit when $\varepsilon \downarrow 0$, say X for the paths and $\mathbb{X}_{s,t}$ for the iterated integrals, then the limit satisfies the cocycle relation. This is another reason to consider this relation as fundamental.

This suggests that to make sense of a numerically convergent scheme for a controlled differential equation when the driving function X is irregular (typically, when the driving function does not have vanishing quadratic variation), one needs to supplement X with other data which involve some arbitrariness. The path X supplemented with additional components playing the role of integrals, subject to certain natural conditions (the object called $\mathbb{X}$ above) is what is called a rough path.

Let us note that we have not really defined what it means for Y to be a solution of $dY_t = V(Y_t)dX_t$. But we are close enough. First, we should acknowledge that X must be supplemented with a $\mathbb{X}$, set $\mathbf{X} := (X, \mathbb{X} =$ and rewrite the equation as $dY_t = V(Y_t)d\mathbf{X}_t$. The we say that Y solves the equation (on some interval) if for s, t in that interval $Y_t - Y_s - V(Y_s)(X_t - X_s) - (V'V)(Y_s)\mathbb{X}_{s,t} = o(t-s)$.

This is hopefully enough motivation for the usefulness of a notion of rough path and we turn to a formal definition.

Appendix

These notes take for granted some familiarity with Brownian motion. The following sections section hopefully may help a reader who lacks this familiarity. Of course, this presentation only covers a microscopic fraction of the subject. Though all definitions are (hopefully) correct, the way we formulate them is far from optimal.

1.A Basic definitions from probability theory

The reader is advised to skip this section at first reading and jump directly to the next one, coming back here only when faced with an unknown notion or notation. We recall a few basic definitions.

Notion of σ-algebra A σ-algebra on a set Ω is a subset $\mathcal{F}$ of $\mathcal{P}(\Omega)$, the set of subsets of Ω (also denoted by 2^{Ω}) such that:

1. The empty set $\emptyset \in \mathcal{F}$
2. If $A \in \mathcal{F}$ then its complement $\Omega \setminus A$ also belongs to $\mathcal{F}$.
3. If sets $A_n \in \mathcal{F}$ for $n \in \mathbb{N}$ are given, the $\cup_{n\in\mathbb{N}} A_n$ also belongs to $\mathcal{F}$.

A member of $\mathcal{F}$ is called $\mathcal{F}$-measurable or simply measurable when there is no risk of confusion. An element of $\mathcal{F}$ is also called an event. Suppose that P is a property of some elements of Ω, i.e. that $\{\omega \in \Omega,\ P(\omega)\}$ defines a subset of Ω. The property P is called measurable if $\{\omega \in \Omega,\ P(\omega)\}$ is an event. It is customary in probability theory to abbreviate $\{\omega \in \Omega,\ P(\omega)\}$ simply by P, that is talk of "the event P".

The pair $(\Omega, \mathcal{F})$ is called a measurable space.

Notion of random variable If $(\Omega, \mathcal{F})$ is a measurable space, a map $X : \Omega \text{to} \mathbb{R}$ is a (real-valued) random variable if for every interval $I \subset \mathbb{R}$ the inverse image $X^{-1}(I) := \{\omega \in \Omega,\ X(\omega) \in I\}$ belongs to $\mathcal{F}$. As an example, for each $A \in \mathcal{F}$ there is a random variable $\mathbf{1}_A$, called the indicator of A defined by $\mathbf{1}_A : \Omega \text{to} \mathbb{R},\ \omega \mapsto \mathbf{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if} \omega \notin A \end{cases}$.

Notion of probability measure If $(\Omega, \mathcal{F})$ is a measurable space, a probability measure on $(\Omega, \mathcal{F})$ is a map $p\colon \mathcal{F} \to [0,1]$ such that if $A_n \in \mathcal{F}$ for $n \in \mathbb{N}$ are disjoint then $p(\cup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} p(A_n)$. This last condition if rephrased as "p is σ-additive".

The triple $(\Omega, \mathcal{F}, p)$ is called a probability space.

An event A such that $p(A) = 0$ is called negligible. One defines $\mathcal{N} := \{B \subset \Omega, \exists A \in \mathcal{F}, p(A) = 0\}$. One shows that[8] $\overline{\mathcal{F}} := \{B \subset \Omega, \exists A \in \mathcal{F} \text{ such that } A\Delta B \in \mathcal{N}\}$ is a σ-algebra on Ω called the completion of $\mathcal{F}$ with respect to p, and $\mathcal{F}$ is said to be complete if $\mathcal{F} = \overline{\mathcal{F}}$, i.e. if $\mathcal{N} \subset \mathcal{F}$. One shows that there is a unique probability measure $\overline{p}$ on $(\Omega, \overline{\mathcal{F}})$ such that $\overline{p}_{|\mathcal{F}}$, the restriction of $\overline{p}$ to $\mathcal{F}$ coincides with p. Thus, it is usually harmless to assume that $\mathcal{F}$ is complete to start with.

Notion of expectation If $(\Omega, \mathcal{F}, p)$ is a probability space, one defines the expectation of an indicator by $\mathbb{E}(\mathbf{1}_A) := p(A)$ for $A \in \mathcal{F}$.

A simple random variable is a finite linear combination with real coefficients of indicators of measurable sets. If $X := \sum_{m\in[\![1,n]\!]} \lambda_m \mathbf{1}_{A_m}$, where $A_m \in \mathcal{F}$ and $\lambda_m \in \mathbb{R}$ are given for $m \in [\![1,n]\!]$, is a simple function, one sets $\mathbb{E}(X) := \sum_{m\in[\![1,n]\!]} \lambda_m p(A_m)$.

If X is an arbitrary positive (i.e. ≥ 0) random variable one sets $\mathbb{E}(X) := \sup_{Y \text{ simple}, Y\leq X} \mathbb{E}(Y)$, a member of $[0, +\infty]$; one says that X is integrable if $\mathbb{E}(X) < +\infty$. If X is an arbitrary random variable, one says that X is integrable $|X|$ is integrable. Then $X^+ := X\mathbf{1}_{X\geq 0}$ and $X^- := -X\mathbf{1}_{X\leq 0}$ are integrable (i.e. $\mathbb{E}(X^+) < +\infty$ and $\mathbb{E}(X^-) < +\infty$) and one sets $\mathbb{E}(X) = \mathbb{E}(X^+) - \mathbb{E}(X^-)$.

This construction is a special case of the construction of the Lebesgue integral, and a more standard notation for $\mathbb{E}(X)$ outside probability theory would be $\int_\Omega X(\omega)\, dp(\omega)$.

Spaces of integrable random variables The space of integrable random variables is denoted by $\mathbb{L}^1(\Omega, \mathcal{F}, p)$, or $\mathbb{L}^1$ when no confusion is possible. One shows that $\mathbb{L}^1(\Omega, \mathcal{F}, p)$ is a vector space, that $X \mapsto \mathbb{E}(X)$ is a linear map from $\mathbb{L}^1$ to $\mathbb{R}$. For $X, Y \in \mathbb{L}^1$, $\mathbb{E}(|X+Y|) \leq \mathbb{E}(|X|) + \mathbb{E}(|Y|)$. Moreover, for $X \in \mathbb{L}^1$ and $\lambda \in \mathbb{R}$, $\mathbb{E}(|\lambda X|) = |\lambda|\mathbb{E}(|X|)$ and $\mathbb{E}(|X|) = 0$ if and only if $p(X \neq 0) = 0$, a condition which defines a linear subspace of $\mathbb{L}^1$ called the subspace of negligible random variables. The map $\mathbb{E}(\cdot)$ descends to the quotient of $\mathbb{L}^1$ by this subspace. It is a common abuse of notation that the class modulo negligible random variables of a random variable X is still denoted by X.[9] In the quotient, the function $\mathbb{E}(|\cdot|)$ defines a norm, and the quotient has an important property: it is complete.

For $q \geq 1$, the space of q-integrable random variables ($|X|^q$ is integrable) is denoted by $\mathbb{L}^q(\Omega, \mathcal{F}, p)$. By Hölder's inequality, it is a vector space and

[8] Recall that $A\Delta B := \{x \in A, x \notin B\} \cup \{x \in B, x \notin A\}$, the symmetric difference of A and B.

[9] In fact $\mathbb{L}^1$ is usually the notation of the quotient space.

going to the quotient modulo negligible random variables, the function $\mathbb{E}(|\cdot|^q)^{1/q}$ defines a norm for which $\mathbb{L}^q$ is complete, i.e. every Cauchy sequences converges. A sequence $(X_n)_n \in \mathbb{N}$ of members of $\mathbb{L}^q$ satisfies the Cauchy criterion if $\forall \varepsilon > 0$, $\exists n \in \mathbb{N}$ such that, for $l, m \geq n$, $\mathbb{E}((X_m - X_l)^q) \leq \varepsilon$. Then, as $\mathbb{L}^q$ is complete, there is a random variable $X \in \mathbb{L}^q$ such that $\lim_{n\to\infty} \mathbb{E}((X_n - X)^q) = 0$.

A very important example is $q = 2$. Convergence in $\mathbb{L}^2$ is also called convergence in mean square.

Exercise 1.9. Check that the sum and product of two random variables are random variables.

Check that if X is an arbitrary random variable, $X^+ := X\mathbf{1}_{X\geq 0}$, $X^- := -X\mathbf{1}_{X\leq 0}$ and $|X|$ are positive random variables.

Check that a simple function is a random variable.

Check that the definition of $\mathbb{E}(X)$ when X is a simple function is consistent despite the fact that X can have several representations as $\sum_{m\in[\![1,n]\!]} \lambda_m \mathbf{1}_{A_m}$. Hint: show that if $\sum_{m\in[\![1,n]\!]} \lambda_m \mathbf{1}_{A_m} = 0$ (the function vanishing everywhere on Ω) then $\sum_{m\in[\![1,n]\!]} \lambda_m p(A_m) = 0$.

Check that if $A_n \in \mathcal{F}$ and $\lambda_n \in \mathbb{R}$ are given for $n \in \mathbb{N}$ with $\sum_{n\in\mathbb{N}} |\lambda_n| < +\infty$ then $X := \sum_{n\in\mathbb{N}} \lambda_n \mathbf{1}_{A_n}$ is a random variable.Check that $\mathbb{E}(X) = \sum_{n\in\mathbb{N}} \lambda_n p(A_n)$.

1.B A quick reminder on Brownian motion

Brownian motion A Brownian motion on a probability space $(\Omega, \mathcal{F}, p)$ is a map $B\colon [0, +\infty[\times\Omega \to \mathbb{R},\ (t,\omega) \mapsto B_t(\omega)$ such that

1. For each fixed $t \in [0, +\infty[$, the map $B_t\colon \Omega \to \mathbb{R},\ \omega \mapsto B_t(\omega)$ is a random variable.
2. For each fixed $\omega \in \Omega$ the map (trajectory) $B(\omega)\colon \mathrm{sTime} \to \mathbb{R},\ t \mapsto B_t(\omega)$ is continuous.
3. The probabilistic laws governing Brownian motion are:

   Brownian motion is a Gaussian process The finite linear combinations $\sum_{m\in[\![1,n]\!]} \lambda_m B_{t_m}$ where $\lambda_1, \cdots \lambda_n \in \mathbb{R}$ and $0 < t_1 < \cdots < t_n < +\infty$ are Gaussian random variables.

   Brownian motion starts at the origin With probability 1, $B_0 = 0$.

   Brownian motion has independent increments For $0 \leq s \leq t \leq u \leq v < +\infty$ the random variables $B_t - B_s$ and $B_v - B_u$ are independent. For a Gaussian process, this independence reduces to the fact that $\mathbb{E}((B_t - B_s)(B_v - B_u)) = 0$.

   Law of increments For $0 \leq s \leq t < +\infty$, $B_t - B_s$ is a Gaussian random variable with mean 0 and variance $t - s$, i.e. $\mathbb{E}(B_t - B_s) = 0$ and $\mathbb{E}\left((B_t - B_s)^2\right) = t - s$.

The probabilistic laws governing Brownian motion can be rewritten in explicit terms that suggest a path integral representation: for $0 = t_0 < t_1 < \cdots < t_n \in [0, +\infty[$ and $I_0, I_1, \cdots, I_n$ intervals of $\mathbb{R}$

$$p(B_{t_0} \in A_0, B_{t_1} \in A_1, \cdots, B_{t_n} \in A_n) = \mathbf{1}_{0 \in A_0} \int_{I_1 \times \cdots \times I_n} dx_1 \cdots dx_n \prod_{m=1}^{n} K(t_m - t_{m-1}, x_m - x_{m-1})$$

where $x_0 := 0$ and $K(t, x) := \frac{1}{\sqrt{2\pi t}} e^{-x^2/(2t)}$ is the Gaussian kernel. Note that $K(t, x) = \frac{1}{\sqrt{t}} K(1, x/\sqrt{t})$, the density of a standard centered Gaussian random variable of mean 0 and variance 1. This scaling property lies at the heart of the intricacies of Brownian trajectories.

The fact that all those properties can be satisfied is non-trivial.

1.C Two simple computations

In this section, $B\colon [0, +\infty[\times \Omega \to \mathbb{R}$, $(t, \omega) \mapsto B_t(\omega)$ is a Brownian motion defined on some probability space.

We start with the quadratic variation. Take an interval $[a, b] \subset [0, +\infty[$. If $\Delta : a = t_0 < t_1 < \cdots < t_n = b$ is a subdivision, we set $Q^\Delta := \sum_{m=0}^{n-1} (B_{t_{m+1}} - B_{t_m})^2$, which with more details would read $Q^\Delta(\omega) := \sum_{m=0}^{n-1} (B_{t_{m+1}}(\omega) - B_{t_m}(\omega))^2$, stressing that is is a random variable. .

By the basic rules of expectations and the defining properties of Brownian motion, $\mathbb{E}\left(Q^\Delta\right) = \sum_{m=0}^{n-1} (t_{m+1} - t_m) = b - a$. Then

$$\begin{aligned}
\mathbb{E}\left((Q^\Delta)^2\right) &= \sum_{l,m=0}^{n-1} \mathbb{E}\left((B_{t_{l+1}} - B_{t_l})^2 (B_{t_{m+1}} - B_{t_m})^2\right) \\
&= 2 \sum_{0 \le l < m < n} \mathbb{E}\left((B_{t_{l+1}} - B_{t_l})^2 (B_{t_{m+1}} - B_{t_m})^2\right) + \sum_{m=0}^{n-1} \mathbb{E}\left((B_{t_{m+1}} - B_{t_m})^4\right).
\end{aligned}$$

By independence of increments,

$$\begin{aligned}
2 \sum_{0 \le l < m < n} \mathbb{E}\left((B_{t_{l+1}} - B_{t_l})^2 (B_{t_{m+1}} - B_{t_m})^2\right) &= 2 \sum_{0 \le l < m < n} (t_{l+1} - t_l)(t_{m+1} - t_m) \\
&= \sum_{l,m=0}^{n-1} (t_{l+1} - t_l)(t_{m+1} - t_m) - \sum_{m=0}^{n-1} (t_{m+1} - t_m)^2 \\
&= (b-a)^2 - \sum_{m=0}^{n-1} (t_{m+1} - t_m)^2.
\end{aligned}$$

Using that the fourth moment of a standard centered Gaussian random variable is 3 we get by scaling that $\mathbb{E}\left((B_{t_{m+1}} - B_{t_m})^4\right) = 3(t_{m+1} - t_m)^2$. To summarize, $\mathbb{E}\left((Q^\Delta)^2\right) = (b-a)^2 + 2\sum_{m=0}^{n-1} (t_{m+1} - t_m)^2$ and

$$\mathbb{E}\left((Q^\Delta - (b-a))^2\right) = 2 \sum_{m=0}^{n-1} (t_{m+1} - t_m)^2 \le (b-a)\mathrm{mesh}(\Delta).$$

Using the definition of mean square convergence, we infer that the family of random variables Q^Δ converges in mean square towards $b-a$, a non-random random variable. This limit is called the quadratic variation of Brownian motion on $[a,b]$. One also says the quadratic variation of Brownian motion is $Q_t := t$ because the quadratic variation on $[a,b]$ is $Q_b - Q_a$.

We turn to the computation of $\int_a^b B_s\, dB_s$ à la Itō. The Itō algorithm define integrals as limits of Riemann-like sums is to use retarded sums: the time at which the integrand (here B_s) is evaluated is always before the times of the increment of the integrator (here dB_s). Concretely this means that $\int_a^b B_s\, dB_s$ is defined to be the limit at small mesh of $S^\Delta := \sum_{m=0}^{n-1} B_{t_m}(B_{t_{m+1}} - B_{t_m})$.[10] Of course, nothing guarantees in advance that the limit exists, or the sense in which it exists. For this simple case, things are easy. We just have to note that $S^\Delta + Q^\Delta/2$ is a telescopic sum, namely

$$S^\Delta + \frac{1}{2}Q^\Delta = \frac{1}{2}\sum_{m=0}^{n-1}(B_{t_{m+1}}^2 - B_{t_m}^2) = \frac{1}{2}(B_b^2 - B_a^2).$$

As Q^Δ converges towards $b-a$ in mean square at small mesh by our first computation, we infer that at small mesh S^Δ converges towards $((B_b^2 - b) - (B_a^2 - a))/2$ in mean square. This is in fact the mode of convergence used for the Itō integral, and we have shown that

$$\int_a^b B_s\, dB_s = \frac{1}{2}(B_b^2 - b) - \frac{1}{2}(B_a^2 - a).$$

We note the appearance of an anomalous term with respect to the naive integral.

As already explained in the main text, for (p-almost) every sample $B(\omega)$ there are subdivisions $\Delta(\Omega)$ of arbitrary small mesh such that $Q^{\Delta(\omega)}(\omega)$ is arbitrary small and others for which it is arbitrary large : one can fine-tune the subdivisions to the sample to get wildly different results for the quadratic variation Q^Δ, hence for the approximations S^Δ. Thus, for (p-almost) every sample $B(\omega)$, $S^\Delta(\omega)$ varies wildly when Δ ranges over all subdivisions of arbitrary small mesh. There is no pathwise definition of the integral $\int_a^b B_s\, dB_s$. And rough path theory will not attempt to define this particular integral pathwise. What it will do it take it as given and use it to provide a pathwise definition of integrals with integrator dB_s but more complicated integrands.

Note that for this simple case, one can show however that if $(\Delta)_{k\in\mathbb{N}}$ is a sequence of subdivisions of $[a,b]$ with $\lim_{k\to\infty} \mathrm{mesh}(\Delta_k) = 0$ then, for (p-almost) every sample $B(\omega)$, S^{Δ_k} goes to $((B_b^2 - b) - (B_a^2 - a))/2$ at large k. For all these subtleties, a nice reference (and the only one I know) is [1].

[10] Again, a more detailed notation stressing the status of random variable would be $S^\Delta(\omega) := \sum_{m=0}^{n-1} B_{t_m}(\omega)(B_{t_{m+1}}(\omega) - B_{t_m}(\omega))$.