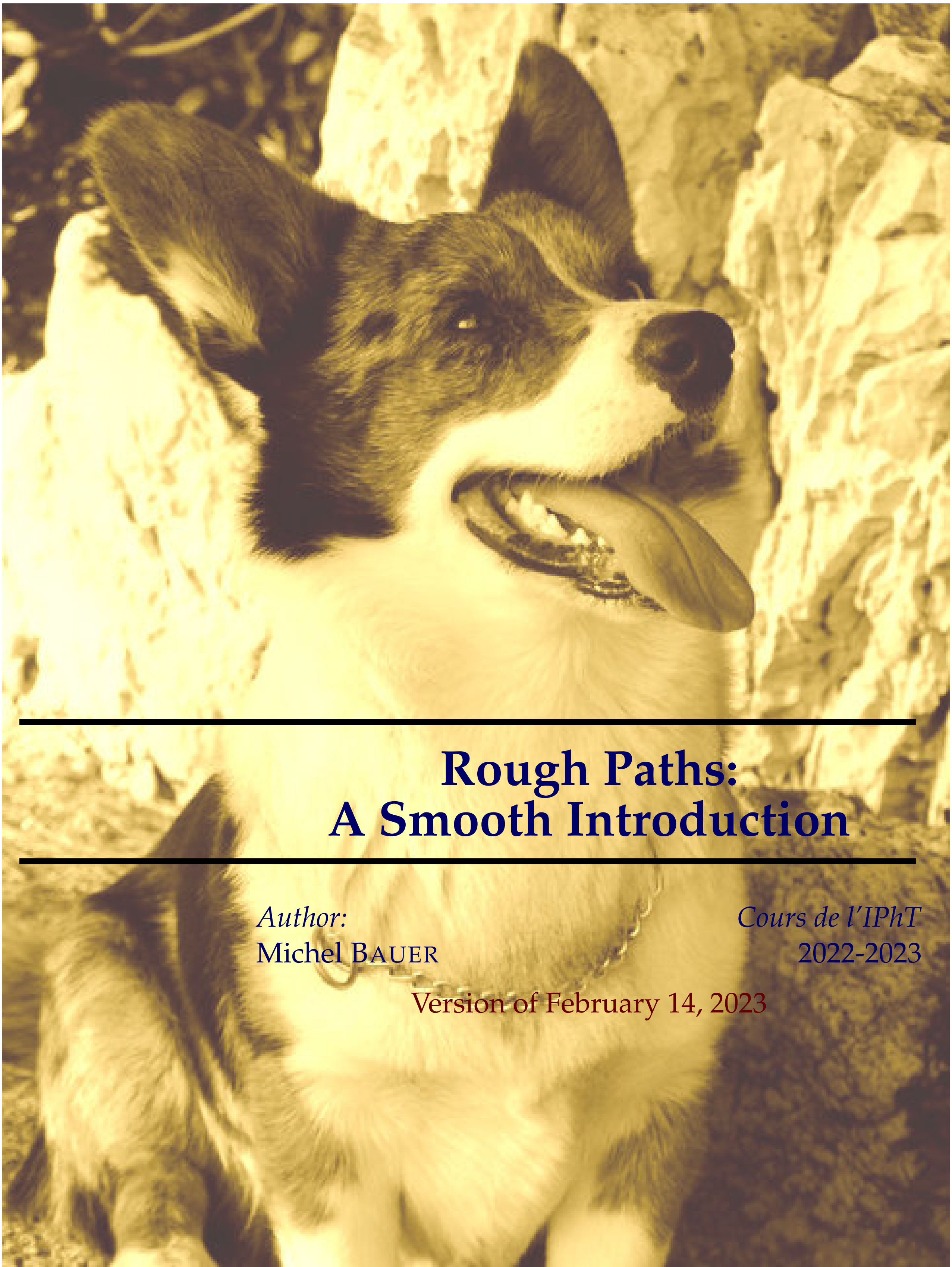
INSTITUT DE PHYSIQUE THÉORIQUE DE SACLAY

IPHT



Rough Paths: A Smooth Introduction

Author:
Michel BAUER

Cours de l'IPhT
2022-2023

Version of February 14, 2023

Rough paths:
A Smooth Introduction

Michel Bauer

February 14, 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

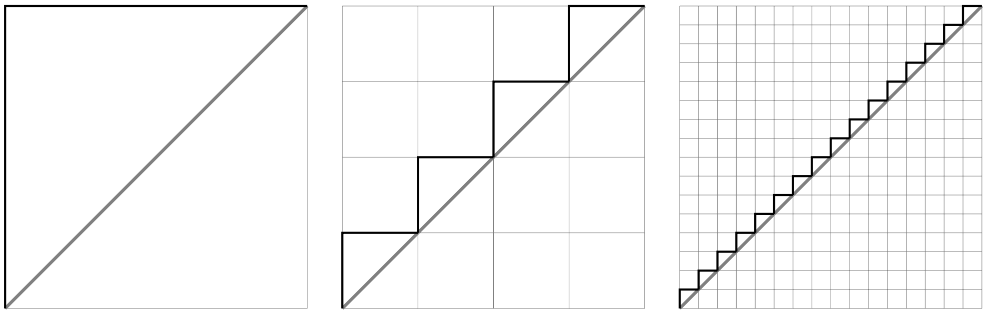
Rough path theory can also be viewed as such an interplay involving approximations, limits and the like. Just a Brownian motion or Levy processes, rough paths can be defined axiomatically without talking of any approximation scheme. And just as Levy processes theory, rough path theory tames a wildlife. 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 $[a, b]$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 some $\delta > 0$.

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

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

We shall often suppress the explicit summation sign and apply the Einstein summation convention. For instance i is repeated twice, once as a subscript and once as a superscript so we may dispense with the summation sign $\sum_i$.

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 approximates $f'(Y_u)V(Y_u)$ 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

$$Y_a^\Delta = y_a \quad Y_{t_{m+1}}^\Delta = Y_{t_m}^\Delta(1 + X_{t_{m+1}} - X_{t_m}) \text{ for } 0 \le m \le n-1,$$

and extend the definition of Y_t^Δ 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: $\log \frac{Y_{t_m}^\Delta}{y_a} = \sum_{l=0}^{m_1} \log(1 + X_{t_{m+1}} - X_{t_m})$. Using the elementary bound $-x^2 \le \log(1+x) - x \le -x^2/3$ for $|x| \le 1/2$ we infer that $\log \frac{Y_b^\Delta}{y_a} - (X_b - X_a) \in [-Q_\Delta, -Q_\Delta/3]$ 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 $\delta_k := \text{mesh}(\Delta_k)$ 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

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

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 $|\log \frac{Y_{t_m}^{\Delta_k}}{y_a} - (X_{t_m} - X_a)| \le C^\Delta e^2/6$ 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 Section 12.2. Instead of one double integral, the above formula would involve a linear combination of $\int_s^t (X_u^i - X_s^i) dX_u^j$ 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 $dY_t = V(Y_t)d\mathbb{X}_t$. 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 $\Omega \backslash A$ 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 $X : \Omega \mathrm{to} \mathbb{R}$ 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 $\mathbf{1}_A : \Omega \mathrm{to} \mathbb{R},\ \omega \mapsto \mathbf{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if} \omega \notin A \end{cases}$.

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] $\overline{\mathcal{F}} := \{B \subset \Omega, \exists A \in \mathcal{F}$ 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) $B(\omega)\colon sTime \to \mathbb{R}$, $t \mapsto B_t(\omega)$ 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 $\mathbb{R}$

$$p(B_{t_0} \in I_0, B_{t_1} \in I_1, \cdots, B_{t_n} \in I_n) = \mathbf{1}_{0\in I_0} \int_{I_1 \times \cdots \times I_n} dx_1 \cdots dx_n \prod_{m=1}^{n} K(t_m - t_{m-1}, x_m - x_{m-1})$$

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

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

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 $S^\Delta(\omega) := \sum_{m=0}^{n-1} B_{t_m}(\omega)(B_{t_{m+1}}(\omega) - B_{t_m}(\omega))$.

CHAPTER 2

Rough path in a nutshell, combinatorics and regularity

Informally, a rough path is a collection of objects whose first component is a standard path and the other components contain enough data to compute integrals involving the first component. Moreover, these additional data involve some arbitrariness, but they are not totally arbitrary. This is very vague but our survey of simple controlled differential equations has uncovered some salient features of the tools we need. Our next goal is to extract a precise definition. We start with combinatorics.

2.1 Rough paths, combinatorics

In this section, we focus on the simplest situation. From the viewpoint of controlled differential equations, it covers the case when the second order of the Born expansion is needed, but not more. Later we shall generalize to cases where higher orders are needed (see Exercise 1.4 for motivation).

Combinatorial rough path A combinatorial two-component rough path on the interval $[a, b]$ with values in a vector space E is a pair $\mathbf{X} = (X, \mathbb{X})$ where $X\colon [a, b] \to E$ is a function of one real variable with values in E (i.e. a path) and $\mathbb{X}\colon [a, b]^2 \to E \otimes E$ is a function of two real variables with values in $E \otimes E$ with the consistency relation (known as *Chen's relation*):

$$\mathbb{X}_{s,u} - \mathbb{X}_{s,t} - \mathbb{X}_{t,u} = (X_t - X_s) \otimes (X_u - X_t) \text{ for } a \leq s, t, u \leq b.$$

The Chen relation is precisely the multidimensional generalization of the relation uncovered in Example 1.6. It guarantees the closure of the flow if one attempts to solve the controlled differential equation $dY_t = V(Y_t)\, dX_t$ for $t \in [a, b]$ with initial condition $Y_a = y_a$ (where Y lives in some vector space) by a second order approximation $Y_t \simeq Y_s + V(X_t - X_s) + V'(Y_s)V(Y_s)\mathbb{X}_{s,t}$, or in coordinates

(with Einstein's summation convention)

$$Y_t^\mu \simeq Y_s^\mu + V_i^\mu(X_t^i - X_s^i) + \frac{\partial V_i^\mu}{\partial y^\nu}(Y_s)V_j^\nu(Y_s)\mathbb{X}_{s,t}^{ji}.$$

For a smooth X we could take $\mathbb{X}_{s,t}^{ji} := \int_s^t(\int_s^u dX_v^j)dX_u^i = \int_s^t(X_u^j - X_s^j)dX_u^i$. In that case the above formula is the second order Born approximation and Chen's relation is Chasles' relation for standard integrals.[1] Later, we shall study rough integrals against X (more precisely against the rough path $(X, \mathbb{X})$ because without surprise the definition involves $\mathbb{X}$) via limits of Riemann type sums. The sums approximating the rough integral $\int_s^t (X_u - X_t) \otimes dX_u$ will turn out to be telescopic by Chen's relation so that the rough integral $\int_s^t (X_u - X_t) \otimes dX_u$ will be found to be precisely $\mathbb{X}_{s,t}$.

The Chen relation, which is of combinatorial nature, leads to a consistent approximate flow, but does not guarantee that iteration with a small time increment leads to a convergent approximation. For that, some analytical constraints, to be introduced later, have to be put on X and $\mathbb{X}$. But first let us make a few remarks.

Gauge freedom If $(X, \mathbb{X})$ is a rough path, and $F\colon [a, b] \to E \otimes E$ a path with values in $E \otimes E$ and $\tilde{\mathbb{X}}_{s,t} := \mathbb{X}_{s,t} + F_t - F_s$ then $(X, \tilde{\mathbb{X}})$ is also a rough path, i.e. Chen's relation is satisfied.
Conversely, if $(X, \mathbb{X})$ and $(X, \tilde{\mathbb{X}})$ are two rough paths with the same first component, setting $\mathbb{F}_{s,t} := \tilde{\mathbb{X}}_{s,t} - \mathbb{X}_{s,t}$ one gets $\mathbb{F}_{s,t} + \mathbb{F}_{s,t} = 0$ and $\mathbb{F}_{s,t} = \mathbb{F}_{a,t} - \mathbb{F}_{a,s}$ for $a \le s, t \le b$, and defining $F_t := \mathbb{F}_{a,t}$ yields $\tilde{\mathbb{X}}_{s,t} = \mathbb{X}_{s,t} + F_t - F_s$.
Thus the ambiguity in the definition of the second component of a rough path is exactly the gauge freedom.

Scalar case Consequently, the general $\mathbb{X}$ if $E = \mathbb{R}$ is $\mathbb{X}_{s,t} = \frac{1}{2}(X_t - X_s)^2 + F_t - F_s$.

A rough path is a path More precisely, one recovers a rough path from a path:
Taking $u = t$ in Chen's relation yields $\mathbb{X}_{t,t} = 0$ for $a \le t \le b$. Then taking

[1] By Chasles relation

$$\mathbb{X}_{s,u} - \mathbb{X}_{s,t} = \int_s^u (X_v - X_s) \otimes dX_v - \int_s^t (X_v - X_s) \otimes dX_v = \int_t^u (X_v - X_s) \otimes dX_v.$$

Then using the linearity of the integral

$$\mathbb{X}_{s,u} - \mathbb{X}_{s,t} - \mathbb{X}_{t,u} = \int_t^u (X_v - X_s) \otimes dX_v - \int_t^u (X_v - X_t) \otimes dX_v = \int_t^u (X_t - X_s) \otimes dX_v.$$

As $X_t - X_s$ is a constant (the variable of integration is v), it factors out of the integral so

$$\mathbb{X}_{s,u} - \mathbb{X}_{s,t} - \mathbb{X}_{t,u} = (X_t - X_s) \otimes \int_t^u dX_v.$$

The remaining integral is $\int_t^u dX_v = X_u - X_t$ leading to Chen's relation.

$u = s$ yields $\mathbb{X}_{s,t} + \mathbb{X}_{t,s} = (X_t - X_s) \otimes (X_t - X_s)$ so that the symmetric part (in the arguments s, t, not in $V \otimes V$) of $\mathbb{X}$ is determined by X. Moreover, taking $s = a$ yields $\mathbb{X}_{t,u} = \mathbb{X}_{a,u} - \mathbb{X}_{a,t} - (X_t - X_a) \otimes (X_u - X_t)$ so that the knowledge of X_t and $\mathbb{X}_{a,t}$ for $a \le t \le b$ is enough to determine the rough path.

Any path can be lifted to a combinatorial rough path Conversely, consider an arbitrary path $(X^{(1)}, X^{(2)})$ on $[a, b]$ with values in $V \oplus V \otimes V$ with first component $X^{(1)}\colon [a, b] \to V$, and second component $X^{(2)}\colon [a, b] \to V \otimes V$, and an arbitrary point $c \in [a, b]$. Then it is immediate to check that $(X, \mathbb{X})$ defined by

$$X_s := X^{(1)}_s \text{ for } s \in [a, b] \text{ and } \mathbb{X}_{s,t} := X^{(2)}_t - X^{(2)}_s - (X^{(1)}_s - X^{(1)}_c) \otimes (X^{(1)}_t - X^{(1)}_s) \text{ for } s, t \in [a, b]$$

is a combinatorial rough path, i.e. satisfies the Chen relation.
Changing the reference point c to, say, $d \in [a, b]$ is easily seen to amount, as it should, to a (special) gauge transformation, with $F_t = (X^{(1)}_d - X^{(1)}_c) \otimes X^{(1)}_t$.
Changing $X^{(2)}$ obviously amounts to a gauge transformation, as it should.

In that sense (combinatorial) rough paths on $[a, b]$ with values in V (more precisely in $V \oplus V \otimes V$) are in correspondence with paths with values in $V \oplus V \otimes V$ in the usual sense, and the correspondence is one to one if one normalizes for instance the component of the path in $V \otimes V$ so that it vanishes at a.

2.2 Rough paths, regularity

It is time to impose some regularity conditions on combinatorial rough paths. This imposes to put some topology, on the target spaces E and $E \otimes E$ and we do so by introducing norms, which we denote by $\|\cdot\|_E$ and $\|\cdot\|_{E \otimes E}$.[2]

A rough path of Hölder type of order $\alpha \in]1/3, 1/2]$ on $[a, b]$ with values in E is a two-component combinatorial rough path $\mathbf{X} = (X, \mathbb{X})$ such that

First component regularity $X\colon [a, b] \to E$ is α-Hölder. This means that for some constant K and $\forall s, t \in [a, b]$, $\|X_t - X_s\|_E \le K|t - s|^\alpha$. Equivalently, X is α-Hölder if

$$\|X\|_\alpha := \sup_{s,t \in [a,b], s \ne t} \frac{\|X_t - X_s\|_E}{|t - s|^\alpha} < +\infty.$$

.

Second component regularity $\mathbb{X}\colon [a, b] \to E \otimes E$ is such that for some constant K and $\forall s, t \in [a, b]$, $\|\mathbb{X}_{s,t}\|_{E \otimes E} \le K|t - s|^{2\alpha}$. Equivalently,

$$\|\mathbb{X}\|_{2\alpha} := \sup_{a \le s \ne t \le b} \frac{\|\mathbb{X}_{s,t}\|_{E \otimes E}}{|t - s|^{2\alpha}} < +\infty.$$

These conditions define the space of rough paths, but also allow to define a metric on it. In particular, we may talk of convergence in the rough paths topology. With words, two rough paths $\mathbf{X}$ and $= \mathbf{Y}$ are close to each other if $\|X_a - Y_a\|_E$

[2] Or simply by $\|\cdot\|$ for both, it is easy from the context to decide in which space the norm is taken.

is small, as are $\|X - Y\|_\alpha$ and $\|\mathbb{X} - \mathbb{Y}\|_{2\alpha}$. The first condition is needed because the Hölder condition is insensitive to translation by a constant. See Subappendix 2.A for a precise definition.

Somme words of caution. First, the target spaces E and $E \otimes E$ are vector spaces but Chen's relation is quadratic, not linear, so the space of rough paths is not a vector space. Second, the "norms" for X and $\mathbb{X}$ do not have the same scaling: $\sqrt{\|\mathbb{X}\|_\alpha}$ is a natural object to consider, but it does not satisfy the triangular inequality needed in the definition of a metric. Third, trading $\|\cdot\|_E$ and $\|\cdot\|_{E\otimes E}$ for two equivalent norms gives the same space of rough paths endowed with an equivalent metric. We shall mostly concentrate on the case when E is finite-dimensional. Then so is $E \otimes E$, so all norms on E are equivalent, and so are those on $E \otimes E$. But it is to be noted that rough paths theory works and has important applications when E is an infinite-dimensional Banach space, and then one has 14 norms with nice properties[3] at disposal to define $E \otimes E$ as a Banach space, but only one, the projective norm, which allows to "linearize bilinear maps" makes life easy for rough paths.

The conditions we introduced are not exactly the ones that we uncovered heuristically at the end of Example 1.6.[4] Those conditions hold with the above definition of rough paths for every $\alpha > 1/3$, but not for $\alpha \leq 1/3$. Thus we may hope for success if we try to solve controlled differential equations with the second order Euler scheme (with $\mathbb{X}$ as a substitute for the second iterated integral) for $\alpha > 1/3$ but we expect problems when a combinatorial rough path $\mathbf{X} = (X, \mathbb{X})$ fullfils the regularity conditions for no $\alpha > 1/3$. As suggested by Exercise 1.4, dealing with less and less regular paths would imply the definition of substitutes for more and more iterated integrals. This generalization is touched upon in Chapter 12.

The reasons why we restrict to $\alpha \leq 1/2$ are also easy to understand. First, if one takes $\alpha > 1/2$ in the above definition, one can show that a naive integration theory can be defined via (avatars of) Riemann sums. This is the so-called Young integral that we shall introduce in Chapter 5. Second, the gauge freedom $\mathbb{X}_{s,t} \to \mathbb{X}_{s,t} + F_t - F_s$ disappears because the function F, which must be 2α-Hölder by the second component regularity, can only be a constant if $\alpha > 1/2$ (why?). Thus, even if the above definition makes sense for $\alpha > 1/2$ it does not bring anything new: given X, there is only one $\mathbb{X}$, and it is given by naive integration methods.

In one dimension, we know that any combinatorial $\mathbf{X}$ has a second component of the form $\mathbb{X}_{s,t} = (X_t - X_s)^2/2 + F_t - F_s$. If X is α-Hölder $(X_t - X_s)^2/2$ automatically satisfies the second regularity condition, and the additional F makes it is and only if it as α-Hölder $(X_t - X_s)^2/2$ so the situation is clear.

In the multidimensional setting, we have shown in the previous section how any X could be lifted to a combinatorial rough path. But it is easy to realize that the resulting $\mathbb{X}$ in general fails to fulfill the regularity conditions: if $\|X\|_\alpha$ is finite then so is $\|\mathbb{X}\|_{\alpha/2}$ but in general $\|\mathbb{X}\|_\alpha$ is not: in words, the short distance

[3] A counting due to Grothendieck!

[4] Recall that they where $(X_t - X_s) = o((t-s)^{1/3})$ and $\mathbb{X}_{s,t} = o((t-s)^{2/3})$.

fluctuations of a $\mathbb{X}_{s,t}$ constructed this way are generically of order $|t-s|^{\alpha}$ whereas the definition imposes that they should be of order $|t-s|^{2\alpha}$. However, there is a general theorem that says essentially[5] that any α-Hölder path, $\alpha \in]1/3, 1/2]$ has a lift to a rough path $(X, \mathbb{X})$. To conclude this overview, let us stress that even if the the "lifting to a rough path" theorem is conceptually crucial, it does not lead to a canonical way to lift. So it is no surprise the often a more direct approach, unrelated to rough path theory, is used. For instance, if X is a sample of some process for which a natural stochastic integrals exists, one defines $\mathbb{X}_{s,t} :=$ "the stochastic integral $\int_s^t (X_u - X_s) \otimes dX_u$". Usually, the Chen relation is automatic in such an approach. As we shall explain later, equipped with this $\mathbb{X}_{s,t}$ rough path theory allows to define integrals against dX or solve differential equations controlled by X without any further use of stochastic integration.

[5]There are a few exceptional borderline cases in infinite dimensions. The theorem is not an easy one. The proof has some similarities with the proof of the Kolmogorov-Centsov theorem.

Appendix

.

2.A The rough paths metric

We define more precisely the rough paths metric. Recall that $\alpha \in]1/3, 1/2]$ and we fixed norms on E and $E \otimes E$. For $X\colon [a,b] \to E$, $\|X\|_\alpha := \sup_{s,t\in[a,b], s\neq t} \frac{\|X_t - X_s\|_E}{|t-s|^\alpha}$ and we define $\mathcal{C}_1^\alpha([a,b], E, \|\ \|_E) := \{X\colon [a,b] \to E,\ \|X\|_\alpha < +\infty\}$. In the same vein, for $\mathbb{X}\colon [a,b] \to E \otimes E$, $\|\mathbb{X}\|_{2\alpha} := \sup_{a\leq s\neq t\leq b} \frac{\|\mathbb{X}_{s,t}\|_{E\otimes E}}{|t-s|^{2\alpha}}$ and we define $\mathcal{C}_2^\alpha([a,b]^2, E \otimes E, \|\ \|_{E\otimes E}) := \{\mathbb{X}\colon [a,b]^2 \to E,\ \|\mathbb{X}\|_\alpha < +\infty\}$. Then

$$\mathcal{C}_1^\alpha([a,b], E, \|\ \|_E) \oplus \mathcal{C}_1^{2\alpha}([a,b], E, \|\ \|_E)$$

is a vector space, which we can endow with a norm, for instance $\|\mathbf{X}\|_\alpha := \|X_a\|_E + \|X\|_\alpha + \|\mathbb{X}\|_{2\alpha}$ for $\mathbf{X} := (X, \mathbb{X})$.[6] This norm induces a distance on that vector space, hence on any subspace, in particular the (quadratic) subspace of rough paths $\mathcal{RP}^\alpha([a,b], E) \subset \mathcal{C}_1^\alpha([a,b], E, \|\ \|_E) \oplus \mathcal{C}_1^{2\alpha}([a,b], E, \|\ \|_E)$ made of pairs $(X, \mathbb{X})$ satisfying Chen's relation. In particular, we may talk of convergence in the rough paths topology.

[6]Of course $\max\{\|X_a\|_E, \|X\|_\alpha, \|\mathbb{X}\|_{2\alpha}\}$ for instance would define an equivalent norm. The $\|X_a\|_E$ contribution is needed because the Hölder condition (as well as the Chen relation for that matter) is insensitive to translation by a constant.

CHAPTER 3

An Illustrative Example

We shall illustrate the ideas mentioned in the Introduction, especially in the section "Rough paths with words" (see p. 3) on a very simple example whose avatars reappear several times in these lectures under different disguises.

3.1 A minimal example of a rough path

We shall study a family of smooth paths, depending on a parameter, and formalize what happens when the parameter goes to 0. Given $\varepsilon > 0$ we set

$$Z^{(\varepsilon)} \colon [0, +\infty[\to \mathbb{C},\ t \mapsto Z_t^{(\varepsilon)} := \sqrt{\varepsilon} e^{it/\varepsilon}$$

which we view as a path in $\mathbb{C} = \mathbb{R} + i\mathbb{R} \simeq \mathbb{R}^2$. This path winds on a circle of radius $\sqrt{\varepsilon}$ and shrinks to a point, i.e. a constant path we call $Z \equiv 0$, when $\varepsilon \downarrow 0$.

However, the winding gets faster as ε gets smaller, and this leaves a footprint on integrals. Writing $Z^{(\varepsilon)} = X^{(\varepsilon)} + iY^{(\varepsilon)}$ we compute that both

$$\int_s^t X_u^{(\varepsilon)}\, dX_u^{(\varepsilon)} = \frac{\varepsilon}{2}(\cos^2 t/\varepsilon - \cos^2 s/\varepsilon) \text{ and } \int_s^t Y_u^{(\varepsilon)}\, dY_u^{(\varepsilon)} = \frac{\varepsilon}{2}(\sin^2 t/\varepsilon - \sin^2 s/\varepsilon)$$

go to 0 (uniformly in s and t) when $\varepsilon \downarrow 0$ but both

$$\int_s^t X_u^{(\varepsilon)}\, dY_u^{(\varepsilon)} = \frac{1}{2}(t - s) + \frac{\varepsilon}{4}(\sin 2t/\varepsilon - \sin 2s/\varepsilon)$$

and

$$\int_s^t Y_u^{(\varepsilon)}\, dX_u^{(\varepsilon)} = -\frac{1}{2}(t - s) + \frac{\varepsilon}{4}(\sin 2t/\varepsilon - \sin 2s/\varepsilon)$$

have a non-trivial limit,

$$\lim_{\varepsilon \downarrow 0} \int_s^t X_u^{(\varepsilon)}\, dY_u^{(\varepsilon)} = \frac{1}{2}(t - s) \text{ and } \lim_{\varepsilon \downarrow 0} \int_s^t Y_u^{(\varepsilon)}\, dX_u^{(\varepsilon)} = -\frac{1}{2}(t - s)$$

(uniformly in s and t).

In particular

$$\frac{1}{2}\int_s^t (X_u^{(\varepsilon)}\,dY_u^{(\varepsilon)} - Y_u^{(\varepsilon)}\,dX_u^{(\varepsilon)}) = t - s,$$

is the area of the sector of the disk of radius $\sqrt{\varepsilon}$ based on the arc beginning at parameter s/ε and ending at parameter t/ε. It is mainly the desire to have a non trivial limit for the area swept by $Z^{(\varepsilon)}$ that justifies the choice of scaling in $Z^{(\varepsilon)}$.

A crucial point is that the "Euler approximation" to this integral,

$$\int_s^t (X_u^{(\varepsilon)}\,dY_u^{(\varepsilon)} - Y_u^{(\varepsilon)}\,dX_u^{(\varepsilon)}) \approx X_x^{(\varepsilon)}(Y_t^{(\varepsilon)} - Y_s^{(\varepsilon)}) - Y_x^{(\varepsilon)}(X_t^{(\varepsilon)} - X_s^{(\varepsilon)}) = \varepsilon \sin(t-s)/\varepsilon$$

exhibits a crossover: the approximation is accurate only for $|t-s| \ll \varepsilon$.

To summarize our results, we introduce a new object, in complex notation for convenience:

$$\mathbb{Z}_{s,t}^{(\varepsilon)} := \int_s^t \begin{pmatrix} (Z_u^{(\varepsilon)} - Z_s^{(\varepsilon)})dZ_u^{(\varepsilon)} & (Z_u^{(\varepsilon)} - Z_s^{(\varepsilon)})d\overline{Z}_u^{(\varepsilon)} \\ (\overline{Z}_u^{(\varepsilon)} - \overline{Z}_s^{(\varepsilon)})dZ_u^{(\varepsilon)} & (\overline{Z}_u^{(\varepsilon)} - \overline{Z}_s^{(\varepsilon)})d\overline{Z}_u^{(\varepsilon)} \end{pmatrix}.$$

We have established that

$$\lim_{\varepsilon\downarrow 0}(Z^{(\varepsilon)}, \mathbb{Z}^{(\varepsilon)}) = (Z, \mathbb{Z}),$$

where $Z_t = 0$ and $\mathbb{Z}_{s,t} = (t-s)\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$ for $s, t \in [0, +\infty[$. The way we obtained this limit is pointwise, which as usual is too weak a convergence for many purposes. It is easy to do better, the interested reader may look at Subappendix 3.A to see an explicit computation of convergence in the rough paths metric.

Integrals along the 0 path Z vanish, so we interpret the existence of a nontrivial $\mathbb{Z}$ as an anomaly: taking the integral of $Z^{(\varepsilon)}$ and then the limit $\varepsilon \downarrow 0$ is not the same as taking the limiting Z and then the integral. This phenomenon is of course very familiar. What is new with rough path theory is the way one exploits it. More on this on an example in the next section.

As a concluding remark, it may seem strange to talk of rough path to describe such smooth objects as the 0 path Z and its companion $\mathbb{Z}$. The point is first that the $Z^{(\varepsilon)}$s are indeed wild objects, and their limit Z must be a "rough" kind of 0 path Z to have integrals involving it (more precisely limits of integrals involving the $Z^{(\varepsilon)}$s) yield a nonzero result. But true, rough path theory is tailored in general to deal with rather singular objects, wilder than Brownian motion for instance.

3.2 A controlled differential equation

Consider the controlled differential equation $dW^{(\varepsilon)} = \overline{W}^{(\varepsilon)} dZ^{(\varepsilon)}$ (and its complex conjugate $d\overline{W}^{(\varepsilon)} = W^{(\varepsilon)} d\overline{Z}^{(\varepsilon)}$) on $[0, 2\pi]$, interpreted as the ordinary differential equation for $W^{(\varepsilon)}$

$$\frac{dW_s^{(\varepsilon)}}{ds} = \overline{W}_s^{(\varepsilon)} \frac{dZ_s^{(\varepsilon)}}{ds}$$

on $[0, 2\pi]$ with initial condition $W_0^{(\varepsilon)} = 1$. This is simple enough to be solved explicitly: taking derivations and using $Z^{(\varepsilon)}\overline{Z}^{(\varepsilon)} = 1/\varepsilon$ one infers that $W^{(\varepsilon)}$ satisfies an ordinary linear differential equation with constant coefficients, namely

$$\varepsilon \frac{d^2 W_s^{(\varepsilon)}}{ds^2} = i\frac{dW_s^{(\varepsilon)}}{ds} + W_s^{(\varepsilon)}.$$

This differential equation is singular when $\varepsilon \downarrow 0$ as confirmed by the expression of the solution: for $\varepsilon < 1/4$ (which we assume from now on) the characteristic exponents are imaginary and setting $\omega_\pm := \frac{1\pm\sqrt{1-4\varepsilon}}{2\varepsilon}$ one checks that

$$W_s^{(\varepsilon)} = \frac{1}{1+\sqrt{\varepsilon}\omega_+}e^{i\omega_+ s} + \frac{1}{1+\sqrt{\varepsilon}\omega_-}e^{i\omega_- s} =: W_s^{(\varepsilon,+)} + W^{(\varepsilon,-)}.$$

It is easily seen that $W^{(\varepsilon,-)}$ behaves nicely as $\varepsilon \downarrow 0$ because $\omega_- \downarrow 1$, leading to the fact that $\lim_{\varepsilon\downarrow 0} W^{(\varepsilon,-)} = Z^{(1)}$ the limit being valid in essentially any topology. On the other hand $W^{(\varepsilon,+)}$ can be rewritten as

$$W^{(\varepsilon,+)} = \frac{1}{\sqrt{\varepsilon_+} + \sqrt{1-\varepsilon_+}} Z^{(\varepsilon_+)}$$

with $\varepsilon_+ := 1/\omega_+$ going down to 0 when $\varepsilon \downarrow 0$. So $W^{(\varepsilon,+)}$ can be studied modulo minor modifications of the study of $Z^{(\varepsilon)}$, an exercise left to the reader after he has read Section D.1. For the time being, we content with the pointwise convergence of $W^{(\varepsilon)}$ to $Z^{(1)}$.

Due to the presence of the singular piece $W^{(\varepsilon,+)}$ in $W^{(\varepsilon)}$ it is perhaps not surprising that integrals have non-trivial limits when $\varepsilon \downarrow 0$. Explicit computations are elementary if tedious. We leave to the reader to check that, defining

$$\mathbb{W}_{s,t}^{(\varepsilon)} := \int_s^t \begin{pmatrix} (W_u^{(\varepsilon)} - W_s^{(\varepsilon)})dW_u^{(\varepsilon)} & (W_u^{(\varepsilon)} - W_s^{(\varepsilon)})d\overline{W}_u^{(\varepsilon)} \\ (\overline{W}_u^{(\varepsilon)} - \overline{W}_s^{(\varepsilon)})dW_u^{(\varepsilon)} & (\overline{W}_u^{(\varepsilon)} - \overline{W}_s^{(\varepsilon)})d\overline{W}_u^{(\varepsilon)} \end{pmatrix},$$

one is led to

$$\lim_{\varepsilon\downarrow 0}(W^{(\varepsilon)}, \mathbb{W}^{(\varepsilon)}) = (W, \mathbb{W}),$$

where $W_t = e^{it}$ and

$$\mathbb{W}_{s,t} = \begin{pmatrix} (e^{it} - e^{is})^2/2 & -i(t-s) + 1 - e^{-i(t-s)} \\ i(t-s) + 1 - e^{i(t-s)} & (e^{-it} - e^{-is})^2/2 \end{pmatrix} + (t-s)\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$$

for $s, t \in [0, +\infty[$. The first contribution is $\int_s^t \begin{pmatrix} (W_u - W_s)dW_u & (W_u - W_s)d\overline{W}_u \\ (\overline{W}_u - \overline{W}_s)dW_u & (\overline{W}_u - \overline{W}_s)d\overline{W}_u \end{pmatrix}$ which is somehow expected, but again there is an anomaly. Thus W ought to be complemented to a rough path.

It is natural to ask whether W, the limit of the solutions to a sequence of nice differential equations, can itself be seen as the solution of a differential equation.

The good news is that (at least naively) there is a limit for the differential equation $dW^{(\varepsilon)} = \overline{W}^{(\varepsilon)} dZ^{(\varepsilon)}$. Closing the eyes, the left-hand side should be dW_t and the right-hand side $\overline{W}_t dZ_t$.

But now comes the bad news: because Z is the 0 path, it is clear that for the W we obtained in the $\varepsilon \downarrow 0$ limit one has $dW_t \neq \overline{W}_t dZ_t$!

However, forgetting for a moment W and Z, think of solving numerically (and naively) the equation

$$dP_t = \overline{P}_t dQ_t \text{ with initial condition } P_0 = 1,$$

where Q is a known complex value function of time and P is to be found. Rewrite the differential equation as $P_t = P_s + \int_s^t \overline{P}_u dQ_u$, with its complex conjugate companion $\overline{P}_u = \overline{P}_s + \int_s^u P_v d\overline{Q}_v$. Performing another step of the Born expansion, we inject the second formula in the first and obtain

$$P_t = P_s + \int_s^t \left(\overline{P}_s + \int_s^u P_v d\overline{Q}_v \right) dQ_u = P_s + \overline{P}_s (Q_t - Q_s) + \int_s^t \left(\int_s^u P_v d\overline{Q}_v \right) dQ_u.$$

In the spirit of the (second order) Euler scheme, we set $P_v \simeq P_s$ in the inner integral to find

$$P_t \simeq P_s + \overline{P}_s (Q_t - Q_s) + P_s \int_s^t \left(\int_s^u d\overline{Q}_v \right) dQ_u.$$

We may "specialize" this formula for $P = W^{(\varepsilon)}$ and $Q = Z^{(\varepsilon)}$:

$$W_t^{(\varepsilon)} \simeq W_s^{(\varepsilon)} + \overline{W}_s^{(\varepsilon)} (Z_t^{(\varepsilon)} - Z_s^{(\varepsilon)}) + W_s^{(\varepsilon)} \int_s^t \left(\int_s^u d\overline{Z}_v^{(\varepsilon)} \right) dZ_u^{(\varepsilon)}.$$

The double integral is $\int_s^t \left(\overline{Z}_u^{(\varepsilon)} - \overline{Z}_s^{(\varepsilon)} \right) dZ_u^{(\varepsilon)}$, i.e the lower left corner of $\mathbb{Z}_{s,t}^{(\varepsilon)}$. Even if $Z^{(\varepsilon)}$ goes to the 0 path Z when $\varepsilon \downarrow 0$ it is not true for its iterated integrals and $\lim_{\varepsilon \downarrow 0} \mathbb{Z}^{(\varepsilon)} = \mathbb{Z}$ with $\mathbb{Z}_{s,t} = (t-s) \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$.Taking the limit when $\varepsilon \downarrow 0$ of the second order Born formula we obtain

$$W_t \simeq W_s + \overline{W}_s^{(\varepsilon)} .0 + W_s . i(t-s) = W_s (1 + i(t-s)).$$

It is reassuring if not surprising that the function satisfying the identity at small $t-s$ and taking value 1 at time 0 is indeed $W_t = e^{it}$.

The moral is that W is not a solution of the ordinary differential equation $dW_t = \overline{W}_t \, dZ_t$ with Z the 0 path, but that it solves the rough differential equation $dW_t = \overline{W}_t dZ_t$ (same formula, different meaning) controlled by the rough path $\mathbf{Z} := (0, \mathbb{Z})$. Note finally that the ordinary differential equation can be seen as the rough differential equation controlled by the trivial rough path $\mathbf{0} := (0, \mathbb{0})$. This may be a bit confusing, but one gets used to it. A way to clarify things would be to replace Z by $\mathbf{Z}$ in the controlled differential equation, i.e. introduce a new notation to rewrite it as $dW = \overline{W} \, d\mathbf{Z}$. The meaning being exactly that the solution is the limit at small mesh of the second order Euler scheme using $\mathbb{Z}$ instead of the naive iterated integral of Z.

Appendix

3.A An example of convergence in the rough paths metric

Recall that for $\varepsilon > 0$ we have set

$$Z^{(\varepsilon)} : [0, +\infty[\to \mathbb{C},\ t \mapsto Z_t^{(\varepsilon)} := \sqrt{\varepsilon} e^{it/\varepsilon}$$

The shrinking of $Z^{(\varepsilon)}$ towards $Z \equiv 0$ when $\varepsilon \downarrow 0$ happens uniformly along the path, in that $\sup_t |Z_t^{(\varepsilon)} - Z_t| = \sqrt{\varepsilon} \downarrow 0$.[1] In the same vein, $\mathbb{Z}^{(\varepsilon)}$ converges uniformly to $\mathbb{Z}$. However, for rough path theory, it is more relevant to know how variations (as physicists we might prefer the term "fluctuations") of $Z^{(\varepsilon)}$ and $\mathbb{Z}^{(\varepsilon)}$ on small scales evolve. These fluctuations are controlled by the rough path metric.

In the following, we use freely two obvious facts: for $x \in \mathbb{R}$, $|\sin x| \min\{|x|, 1\}$ and, for $x, y \in [0, +\infty[$ and $\alpha \in [0, 1]$, $\min\{x, y\} \leq x^\alpha y^{1-\alpha}$ (the min-inequality).

Fluctations of $Z^{(\varepsilon)} - Z$ To get a hold on $Z^{(\varepsilon)} - Z = Z^{(\varepsilon)}$ in $\mathcal{C}_1^\alpha([0, +\infty[, \mathbb{C})$ spaces we compute[2]

$$|Z_t^{(\varepsilon)} - Z_s^{(\varepsilon)}| = 2\sqrt{\varepsilon} |\sin((t-s)/(2\varepsilon))| \leq \min\{2\sqrt{\varepsilon}, |t-s|/\sqrt{\varepsilon}\}.$$

This inequality is rather crude, but captures well the crossover between the regime when $|t - s| \gg \varepsilon$ and when $|t - s| \ll \varepsilon$. Even if for fixed ε the paths $Z^{(\varepsilon)}$ are smooth, each derivative pulls out a factor $1/\varepsilon$ so to get a hold on what happens for varying ε at every scale of $|t - s|$ one has to accept less regularity. Using the min-inequality for $\alpha = 1/2$ we find that

$$|Z_t^{(\varepsilon)} - Z_s^{(\varepsilon)}| \leq \sqrt{2} |t-s|^{1/2}$$

for every epsilon, and for every $s, t \in [0, +\infty[$. So the family $Z^{(\varepsilon)}$ is bounded in $\mathcal{C}_1^{1/2}([0, +\infty[, \mathbb{C})$. For general α the min-inequality yields

$$|Z_t^{(\varepsilon)} - Z_s^{(\varepsilon)}| \leq 2^{1-\alpha} |t-s|^\alpha \varepsilon^{1/2-\alpha},$$

[1] We take this opportunity to recall that for the Riemann integral uniform convergence for a sequence of integrands on a bounded interval is enough to guarantee the convergence of the integral towards its naive limit. But here, the integrand *and* the integrator depend on a parameter, and as should be expected uniform convergence is not strong enough a criterion for the integrator.

[2] As $Z_0^{(\varepsilon)} \to Z_0$ when $\varepsilon \downarrow 0$, there is no issue with translation by a constant.

This impliess that for $\alpha \in]0,1/2[$ $\|Z^{(\varepsilon)}-Z\|_\alpha$ is a $O(\varepsilon^{1/2-\alpha})$ leading to convergence in $\mathcal{C}_1^\alpha([0,+\infty[,\mathbb{C})$ for every $\alpha \in]0,1/2[$. This can be rephrased as the fact that the fluctuations of $Z^{(\varepsilon)}$ are $1/2$-Hölder uniformly in $\varepsilon > 0$ and get close to those of the 0 path in the α-Hölder sense when $\varepsilon \downarrow 0$ for $\alpha < 1/2$. A simple scaling property shows that Hölder convergence does not hold for $\alpha < 1/2$.

See Subappendix 3.B for a generalization.

Fluctations of $\mathbb{Z}^{(\varepsilon)} - \mathbb{Z}$ Our computations have shown that

$$\mathbb{Z}^{(\varepsilon)}_{s,t} := (t-s)\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} + \begin{pmatrix} \frac{1}{2}(Z^{(\varepsilon)}_t - Z^{(\varepsilon)}_s)^2 & -Z^{(\varepsilon)}_s(\overline{Z}^{(\varepsilon)}_t - \overline{Z^{(\varepsilon)}}_s) \\ -\overline{Z}^{(\varepsilon)}_s(Z^{(\varepsilon)}_t - Z^{(\varepsilon)}_s) & \frac{1}{2}(\overline{Z}^{(\varepsilon)}_t - \overline{Z}^{(\varepsilon)}_s)^2 \end{pmatrix}.$$

We look at each matrix element of $\mathbb{Z}^{(\varepsilon)} - \mathbb{Z}$ which is given by the matrix on the right in the previous line. The computations parallel closely the ones for $Z^{(\varepsilon)}$. For instance the $(1,1)$ component is simply $\frac{1}{2}(Z^{(\varepsilon)}_t - Z^{(\varepsilon)}_s)^2$ so one simply has to take squares in the previous discussion. As the $(2,2)$ components is the complex conjugate of the $(1,1)$ component, the same remark applies. Finally, for the $(1,2)$ or the $(2,1)$ components (again complex conjugates of each other) we have, by the min-inequality

$$\left|-\overline{Z}^{(\varepsilon)}_s(Z^{(\varepsilon)}_t - Z^{(\varepsilon)}_s)\right| = \varepsilon|2\sin((t-s)/(2\varepsilon))| \leq \min\{|t-s|, 2\varepsilon\} \leq |t-s|^{2\alpha}(2\varepsilon)^{1-2\alpha}$$

for $\alpha \in [0,1/2]$. For $\alpha = 0$ we get uniform convergence and for $\alpha = 1/2$ we get boundedness in $\mathcal{C}_2^{1/2}([0,+\infty[^2, \mathbb{C}\otimes_{\mathbb{R}}\mathbb{C})$.

Taking all the components into account, we have found that $\|\mathbb{Z}^{(\varepsilon)} - \mathbb{Z}\|_1$ is bounded, and that for $\alpha \in]0,1/2[$ $\|\mathbb{Z}^{(\varepsilon)} - \mathbb{Z}\|_{2\alpha}$ is a $O(\varepsilon^{1-2\alpha})$ leading to convergence in $\mathcal{C}_2^{2\alpha}([0,+\infty[^2, \mathbb{C}\otimes_{\mathbb{R}}\mathbb{C})$ for every $\alpha \in]0,1/2[$.

3.B The interpolation property

In fact, the above computations illustrate a general result and which has an extension to rough paths (where Chen's relation proves again crucial).

Interpolation property Suppose $X^{(n)}$, $n \in \mathbb{N}$ is a sequence of paths on $[a,b]$ such that first, for some $\beta \in]0,1]$, $\sup_n \|X^{(n)}\|_\beta < +\infty$ and second $X^{(n)}_t \to X_t$ pointwise on $[a,b]$. Then X_t is β-Hölder, $X^{(n)}$ converges towards X uniformly on $[a,b]$ and $\|X^{(n)} - X\|_\alpha \to 0$ at large n for every $\alpha \in]0,\beta[$.

Proof. By hypothesis, there is a constant $C > 0$ such that $\left|X^{(n)}_t - X^{(n)}_s\right| \leq C|t-s|^\beta$ for every $n \in \mathbb{N}$ and every $s,t \in [a,b]$. Taking the limit $n \to \infty$ we infer that $|X_t - X_s| \leq C|t-s|^\beta$ for every $s,t \in [a,b]$, i.e. $X = \lim_{n\to+\infty} X^{(n)}$ is β-Hölder. This being established, $Y^{(n)} := X^{(n)} - X$ satisfies $\left|Y^{(n)}_t - Y^{(n)}_s\right| \leq 2C|t-s|^\beta$ for every $n \in \mathbb{N}$ and every $s,t \in [a,b]$, and $Y^{(n)}_t \to 0$ pointwise on $[a,b]$.

We show that this implies that $Y^{(n)} \to 0$ uniformly on $[a, b]$. For that, let $\varepsilon > 0$. Take a subdivision Δ of $[a, b]$ with $\mathrm{mesh}(\Delta)/2 \leq \left(\frac{\varepsilon}{4C}\right)^{1/\beta}$. Take m large enough that $\left|Y_s^{(n)}\right| \leq \varepsilon/2$ for every $n \geq m$ and every subdivision point s of Δ (there is only a finite number of subdivision points). For each $t \in [a, b]$ there is a subdivision point s_t of Δ such that $|t - s_t| \leq \mathrm{mesh}(\Delta)/2$ so that for $n \geq m$

$$\left|Y_t^{(n)}\right| \leq \left|Y_t^{(n)} - Y_{s_t}^{(n)}\right| + \left|Y_{s_t}^{(n)}\right| \leq 2C|t - s_t|^\beta + \varepsilon/2 \leq \varepsilon.$$

Thus convergence is indeed uniform, and there is a sequence ε_n, $n \in \mathbb{N}$ such that $\varepsilon_n \to 0$ at large n and $\left|Y_t^{(n)}\right| \leq \varepsilon_n$ for every n and every $t \in [a, b]$. We can use again the min-inequality: $\left|Y_t^{(n)} - Y_s^{(n)}\right| \leq \min\{2C|t-s|^\beta, 2\varepsilon_n\} \leq 2\left(C|t - s|^\beta\right)^{\alpha/\beta} \varepsilon_n^{1-\alpha/\beta}$ for $\alpha \in [0, \beta]$, for every $n \in \mathbb{N}$ and every $s, t \in [a, b]$ so that $\|Y^{(n)}\|_\alpha \to 0$ at large n for $\alpha \in]0, \eta[$. $\square$

Our computations with Z illustrate (in the case $\beta = 1/2$) that this result is sharp: in general $\|Y^{(n)}\|_\beta$ cannot be expected to go to 0 at large n.

CHAPTER 4

The sewing lemma

In this chapter, we change subject almost completely... or so it seems. What remains in the background is ta systematic construction of an integration theory well-suited to rough path.

4.1 Motivation

We want to abstract a construction that we are all familiar with in the context of the Riemann integral. In any situation when one needs to define an integral "à la Riemann" $\int_s^t (\delta\Gamma)_u$, depending on the upper and lower bounds s, t via a limiting procedure, one starts from an explicit expression $\Gamma_{s,t}$, wishing that the integral $\int_s^t (\delta\Gamma)_u$ we look forward is close enough to $\Gamma_{s,t}$ when $|t-s|$ is small so that the integral $\int_a^b (\delta\Gamma)_u$ may be approximated by a finite sum over subdivisions. This informal discussion emphasizes the role played by Chasles' relation, which is extremely fundamental.

For instance, to define the standard Riemann integral denoted by $\int_s^t Y_u du$ one takes $\Gamma^Y_{s,t} := Y_s(t-s)$ and it is immediate that a general Riemann sum $S_\Delta(Y, a, b) := \sum_{m=0}^{n-1} Y_{t_{2m+1}}(t_{2m+2} - t_{2m})$ along a tagged subdivision[1] $\Delta : a = t_0 \le t_1 \cdots \le t_{2n} = b$ of $[a, b]$ can be rewritten as $S_\Delta(Y, a, b) = \sum_{m=0}^{n-1} \Gamma^{(Y)}_{t_{2m+1},t_{2m+2}} - \Gamma^{(Y)}_{t_{2m+1},t_{2m}}$. By definition Y is Riemann-integrable if the Riemann sums associated to Y go to a finite limit when $\mathrm{mesh}(\Delta) \downarrow 0$.

In the same spirit, when we discussed the Euler scheme for controlled differential equations of the type $dY = V(Y)dX$ where Y is an unknown real function and X is a real function that appears as the first component of a two-components rough source $\mathbf{X} := (X, \mathbb{X})$, we were led to approximate a formal expression $\int_s^t V(Y_u)\, dX_u$ (more correctly $\int_s^t V(Y_u)\, d\mathbf{X}_u$) by $V(Y_s)(X_t - X_s) + V'(Y_s)V(Y_s)\mathbb{X}_{s,t}$ for small $|t-s|$.

[1] For the reader who feels the need for it, the formal definition of subdivisions and tagged subdivisions is recalled in Section A.1

In the same rough path context, we would approximate a formal expression $\int_s^t f(X_u)\, dX_u$ by $\Gamma^{(\mathbf{X},f)}_{s,t} := f(X_s)(X_t - X_s) + F'(X_s)\mathbb{X}_{s,t}$ for small $|t-s|$, and then declare that the rough integral $\int_s^t f(X_u)\, d\mathbf{X}_u$ exists if the sums along subdivisions $\sum_{m=0}^{n-1} \Gamma^{(\mathbf{X},f)}_{t_{2m+1},t_{2m+2}} - \Gamma^{(\mathbf{X},f)}_{t_{2m+1},t_{2m}}$ go to a finite limit when $\mathrm{mesh}(\Delta) \downarrow 0$.

It turns out to be fruitful to discuss more generally the existence of a limit for such sums when $\Gamma^{(Y)}$ or $\Gamma^{(\mathbf{X},f)}$ are replaced by an arbitrary function of two variables Γ on $[a,b]$.

4.2 Riemann sums and their limits

Let $\Gamma\colon [a,b]^2 \to \mathbb{R}$ be a map. Let$\Delta : a = t_0 \leq t_1 \cdots \leq t_{2n} = b$ be a tagged subdivision of $[a,b]$.

Generalized Riemann sums The Riemann sum of Γ along Δ is defined to be

$$S_\Delta(\Gamma, a, b) := \sum_{m=0}^{n-1} \Gamma_{t_{2m+1},t_{2m+2}} - \Gamma_{t_{2m+1},t_{2m}}.$$

A map from $[a,b]^2$ to $\mathbb{R}$ is also called a summand (with value in $\mathbb{R}$) in what follows. Instead of the term "Riemann sum of Γ" we often write "Γ-Riemann sum", or simply "Riemann sum" when there is no risk of confusion.

Retarded and advanced Riemann sums of Γ play a useful role in the discussion.

Advanced and retarded Γ-Riemann sums If $\Delta : a = t_0 < t_1 \cdots < t_n = b$ is a subdivision (not a tagged subdivision) of $[a,b]$ we set

$$S^{\mathrm{ret}}_\Delta(\Gamma, a, b) := \sum_{m=0}^{n-1} \Gamma_{t_m,t_{m+1}} - \Gamma_{t_m,t_m} \qquad S^{\mathrm{adv}}_\Delta(\Gamma, a, b) = \sum_{m=0}^{n-1} \Gamma_{t_{m+1},t_{m+1}} - \Gamma_{t_{m+1},t_m}.$$

Those are special cases of general Riemann sums. Recall from Section A.1 that if $\Delta : a = t_0 \leq t_1 \cdots \leq t_{2n} = b$ is a tagged subdivision of $[a,b]$ then $\Delta_* : a = t_0 < t_2 \cdots < t_{2n} = b$ is a subdivision of $[a,b]$. If Δ is such that $t_{2m+1} = t_{2m}$ (resp. $t_{2m+1} = t_{2m+2}$) for $m = 0, \cdots, n-1$ the Riemann sum along Δ is just a retarded (resp. advanced) Riemann sum along Δ_*.

We want to explore the small mesh behavior of Riemann sums. If $\lim_{\mathrm{mesh}(\Delta)\downarrow 0} S_\Delta(\Gamma, a, b)$ exists, we denote it by $\int_a^b (\delta\Gamma)$ and call it the integral of Γ on the interval $[a,b]$.

Here is a list of basic facts, the details and proofs can be found in Appendix C.

Locality If Γ vanishes in a neighborhood of the diagonal, then $\int_a^b (\delta\Gamma)$ exists and is 0. By linearity, the existence and then the value, of $\int_a^b (\delta\Gamma)$ depend only on the germ of Γ along the diagonal, or in physical term on the short distance behavior of $\Gamma_{s,t}$, $|t-s|$ small.

Subintervals Let $a \leq s \leq t \leq b$. We can restrict Γ to $[s,t]^2$. If $\int_a^b (\delta\Gamma)$ exists then so does $\int_s^t (\delta\Gamma)$.

Chasles' relation If $a \leq t \leq b$ and $\int_a^b(\delta\Gamma)$ exists then $\int_a^b(\delta\Gamma) = \int_a^t(\delta\Gamma) + \int_t^b(\delta\Gamma)$. Thus there is a function $\check{\Gamma}\colon [a,b] \to \mathbb{R}$, well-defined up to an additive constant such that $\int_s^t(\delta\Gamma) = \check{\Gamma}_t - \check{\Gamma}_s$. As usual, this allows to define $\int_s^t(\delta\Gamma)$ for $t \leq s$ as well in a way consistent with Chasles relation.[2]Joined with the previous fact, this allows to define $\int(\delta\Gamma)\colon [a,b]^2 \to \mathbb{R}$ by $(\int(\delta\Gamma))_{s,t} := \int_s^t(\delta\Gamma)$.

Integrability conditions, locality (II) If $\int_a^b(\delta\Gamma)$ exists then $\sum_{m=0}^{n-1} |\int_{t_m}^{t_{m+1}} -(\delta\Gamma) - (\Gamma_{t_m,t_{m+1}} - \Gamma_{t_m,t_m})|$ and $\sum_{m=0}^{n-1} |\int_{t_m}^{t_{m+1}} -(\delta\Gamma) - (\Gamma_{t_{m+1},t_{m+1}} - \Gamma_{t_{m+1},t_m})|$ are small when the mesh of the subdivision $\Delta : a = t_0 < t_1 \cdots < t_n = b$ is small. Conversely, if there is a function $\check{\Gamma}\colon [a,b] \to \mathbb{R}$ such that $\sum_{m=0}^{n-1} |(\check{\Gamma}_{t_{m+1}} - \check{\Gamma}_{t_m}) - (\Gamma_{t_m,t_{m+1}} - \Gamma_{t_m,t_m})|$ and $\sum_{m=0}^{n-1} |(\check{\Gamma}_{t_{m+1}} - \check{\Gamma}_{t_m}) - (\Gamma_{t_{m+1},t_{m+1}} - \Gamma_{t_{m+1},t_m})|$ are small when the mesh of the subdivision $\Delta : a = t_0 < t_1 \cdots < t_n = b$ is small then $\int_a^b(\delta\Gamma)$ exists and equals $\check{\Gamma}_b - \check{\Gamma}_a$.

Reparameterization invariance If $\varphi\colon [c,d] \to [a,b]$ is an arbitrary function we can define the pullback $\varphi^*\Gamma\colon [c,d]^2 \to \mathbb{R}$ of Γ by φ by $(\varphi^*\Gamma)_{u,v} := \Gamma_{\varphi(u),\varphi(v)}$. If φ is continuous and increasing, and if $\int_a^b(\delta\Gamma)$ exists then $\int_c^d(\delta(\varphi^*\Gamma))$ exists as well and the two are equal.

Exercise 4.1. Find $\Gamma\colon [-1,1]^2 \to \mathbb{R}$ such that $\int_{-1}^0(\delta\Gamma)$ and $\int_0^1(\delta\Gamma)$ both exist, but $\int_{-1}^1(\delta\Gamma)$ does not.

4.3 An attempt at an analogy

In this section, we concentrate on retarded Riemann sums for convenience. Thus the only values of $\Gamma_{s,t}$ that count are those for which $s \leq t$ and we define $[a,b]^2_{\leq} := \{s,t \in [a,b],\ s \leq t\}$. Accordingly, when we write that $\int_a^b(\delta\Gamma)$ exists, we only mean that retarded Riemann sums have a limit at small mesh.[3]

We denote by $\mathfrak{G}_{[a,b]}$ the (vector) space of all functions $\Gamma\colon [a,b]^2_{\leq} \to \mathbb{R}$ which vanish at coinciding points. We turn $\mathfrak{G}_{[a,b]}$ into an algebra with unit by pointwise multiplications: $(\Gamma' \cdot \Gamma'')_{s,t} := \Gamma'_{s,t}\Gamma''_{s,t}$. We want to view $\mathfrak{G}_{[a,b]}$ as a space analogous to a space of field theories (with a cutoff).

If $u \in [a,b]$ and $\Gamma \in \mathfrak{G}_{[a,b]}$ we define $R_u\Gamma \in \mathfrak{G}_{[a,b]}$ i.e. $(R_u\Gamma)_{s,t}$ for $(s,t) \in [a,b]^2_{\leq}$ by $(R_u\Gamma)_{s,t} := \begin{cases} \Gamma_{s,t} & \text{if } u \notin]s,t[\\ \Gamma_{s,u} + \Gamma_{s,t} & \text{if } u \in]s,t[\end{cases}$. Thus R_u is a linear transformation on $\mathfrak{G}_{[a,b]}$.

It is easy to check that $R_a = R_b = \mathrm{Id}$, that each R_u is a projection ($R_u^2 = R_u$) and that the transformations $(R_u)_{u\in[a,b]}$ commute. Thus we may always write a product of R_us where the us are in increasing order and without repetition, i.e.

[2]Beware however that the existence of $\int_a^t(\delta\Gamma)$ and $\int_t^b(\delta\Gamma)$ for some $t \in]a,b[$ does not imply the existence of $\int_a^b(\delta\Gamma)$, see Exercise 4.1.

[3]It is plain that Chasles relation remains valid, and that the integrability conditions reduce to the one involving the retarded situation.

write the product as $R_{t_1} \cdots R_{t_{n-1}}$ for some subdivision $\Delta : a = t_0 < t_1 \cdots < t_n = b$ of $[a, b]$. The interesting point is that $S_\Delta^{\text{ret}}(\Gamma, a, b) = (R_{t_1} \cdots R_{t_{n-1}}\Gamma)_{a,b}$. Finally, $R_u\Gamma = \Gamma$ for every $u \in [a, b]$ if and only if $\Gamma_{s,t} = \check{\Gamma}_t - \check{\Gamma}_s$ for some function $\check{\Gamma}\colon [a, b] \to \mathbb{R}$. This leads to the following analogy.

We view the R_us, more generally their products, as renormalization group transformations. The general renormalization transformations are in one-to-one correspondence with subdivisions. The cutoff scale of the renormalization transformation is defined to be the mesh[4] of the associated subdivision. Then $\int_a^b(\delta\Gamma)$ exists if and only if Γ goes to a fixed point under renormalization transformations when the cutoff goes to 0.

From all viewpoints, this flow is much simpler than renormalization group flows on realistic spaces of field theories. In particular, our toy renormalization transformations are linear. One feature that is slightly unusual is that we deal with *local* transformations: nothing prevents us from looking at much smaller scales in one region than in another, the only thing that matters for the approach of a fixed point is that the overall scale gets small. A more standard approach would be to consider just a sequence of nested subdivisions, splitting $[a, b]$ in 2, then 4, then 8 and so on, pieces of equal size. This would hide the reparameterization invariance though.

There are some obvious manifestations of universality. To mention only one, suppose that Γ goes to a fixed point and let $\overline{\Gamma}$ be an arbitrary bounded member of $\mathfrak{G}_{[a,b]}$. Then $\overline{\Gamma}\cdot(\Gamma - \int(\delta\Gamma))$ goes to the trivial fixed point, an immediate consequence of the integrability criterion plus the assumed boundedness of $\overline{\Gamma}$. In particular either $\overline{\Gamma}\cdot\Gamma$ and $\overline{\Gamma}\cdot\int(\delta\Gamma)$ go to the same fixed point, or none of them goes to a fixed point.

4.4 Integrability via Hölder-type conditions

Up to now, we have given some properties of $\int(\delta\Gamma)$ when it exists, but it remains to offer some conditions on Γ sufficient to guarantee the existence of a fixed point. These conditions should be of some generality but also be verifiable in practice.

The simplest condition is of Hölder type. Let $\gamma \in \mathbb{R}$.

Triangular summands A function $\Gamma\colon [a, b]^2 \to \mathbb{R}$ is triangular with exponent γ if there is a constant K such that for every $s, t, u \in [a, b]$

$$|\Gamma_{s,u} - \Gamma_{s,t} - \Gamma_{t,u} + \Gamma_{t,t}| \leq K|s, t, u|^\gamma.$$ [5]

[4] Maybe it is more standard to take the inverse of the mesh as the cutoff, but we stick to the direct definition as the mesh itself.

[5] We denote by $|s, t, u|$ the length of the smallest interval containing s, t, u, i.e.

$$|s, t, u| := \max(|t - s|, |u - t|, |u - s|) = \max(s, t, u) - \min(s, t, u),$$

a quantity which vanishes if and only of the three points s, t, u coincide, and equals $u - s$ if $s \leq t \leq u$.

If Γ is in reduced form, i.e. vanishes on the diagonal, there are only three terms in the condition to be triangular and the structure of $\Gamma_{s,u}-\Gamma_{s,t}-\Gamma_{t,u}$ is reminiscent of the triangular condition in metric spaces, hence the name "Triangular summand".

We denote by $K_{[a,b]}(\Gamma)$ or simply $K_{[a,b]}$ the best possible constant, i.e.

$$K_{[a,b]}(\Gamma) := \sup_{s,t,u\in[a,b],\,|s,t,u|>0} \frac{\|\Gamma_{s,u}-\Gamma_{s,t}-\Gamma_{t,u}+\Gamma_{t,t}\|}{|s,t,u|^{\gamma}},$$

which makes sense only if $[a,b]$ is a nontrivial interval, i.e. if $a < b$. By convention, we set $K_{[a,b]}(\Gamma) := 0$ if $a = b$. Note that if Γ is triangular on $[a,b]^2$ and $[s,t] \subset [a,b]$ is a nontrivial interval the restriction of Γ to $[s,t]^2$ is also triangular (for the same exponent) and clearly $K_{[s,t]} \leq K_{[a,b]}$.

With this definition in mind, we can state the main result of this Chapter, which gives it his name.

The sewing lemma Suppose $\Gamma\colon [a,b]^2 \to \mathbb{R}$ is triangular with exponent $\gamma > 1$. Then $\int_a^b(\delta\Gamma)$ exists.

The Young-Loëve inequality Then $|\int_a^b(\delta\Gamma)-(\Gamma_{a,b}-\Gamma_{a,a})| \leq \frac{1}{1-2^{1-\gamma}}K_{[a,b]}(\Gamma)(b-a)^{\gamma}$ holds.

The Young-Loëve inequality has an immediate consequence: using Chasles' relation we infer that if $\Delta : a = t_0 < t_1 \cdots < t_n = b$ is a subdivision of $[a,b]$ then $|\int_a^b(\delta\Gamma) - S^{\mathrm{ret}}_{\Delta}(\Gamma,a,b)| \leq \frac{1}{1-2^{1-\gamma}}K_{[a,b]}(\Gamma)(b-a)\mathrm{mesh}(\Delta)^{\gamma-1}$. In terms of our renormalization group interpretation, $\gamma-1$ is a critical exponent that describes the approach of the fixed point. Unlike the usual field theory situation, $\gamma-1$ is only a lower bound for the speed of approach. This is because γ is also only a lower bound in the formulation of the triangular property. Loosely speaking, if γ is sharp uniformly in s,t,u in the triangular property, we expect that $\gamma-1$ will also be sharp for the approach to the fixed point.

Let us mention the main points of the strategy of proof:

– First one proves the convergence of retarded Riemann sums for a special sequence of nested subdivisions of $[a,b]$: at step k the subdivision points are the dyadic rationals of order k that belong to $]a,b[$. So up to (small) boundary effects, going from step k to step $k+1$ divides the cutoff by 2 uniformly in $[a,b]$ as in standard renormalization group transformations. One gets get via the triangular property an exponentially decreasing bound for the difference between the (retarded) Riemann sums at one step and the next and this implies convergence. The drawback of small boundary effects is compensated by the fact that Chasles' relation for the limit can be proven easily. One proves also that the limit satisfies the Young-Loëve inequality.

– The Young-Loëve inequality plus Chasles' relation and the triangular property imply easily that in fact general retarded Riemann sums converge to the same limit at small mesh.

– The same tools then allow to deal with general Riemann sums. This is the only

place where the triangular property is used with $t \leq s \leq u$ and not $s \leq t \leq u$.
– Using plain dichotomy of $[a, b]$ to construct a sequence of nested subdivisions, one improves the constant in the Young-Loëve inequality.

We are now ready to apply the sewing lemma to Young integrals.

Appendix

In this appendix $(E, \| \|_E)$ denotes a Banach space (i.e. a normed vector space in which Cauchy sequences converge). We fix $a, b \in \mathbb{R}$, $a \le b$.

If Γ is a function on $[a,b]^2$ with values in E and $\Delta : a = t_0 \le t_1 \cdots \le t_{2n} = b$ is a tagged subdivision of $[a,b]$, recall that the Γ-Riemann sum along Δ is

$$S_\Delta(\Gamma, a, b) := \sum_{m=0}^{n-1} \Gamma_{t_{2m+1},t_{2m+2}} - \Gamma_{t_{2m+1},t_{2m}}.$$

We observe that $S_\Delta(\Gamma, a, b)$ is a sum of independent contributions, one for each part of Δ. Retarded and advanced sums play an important role in the sequel. Recall that if $\Delta : a = t_0 < t_1 \cdots < t_n = b$ is a subdivision of $[a,b]$ we have defined

$$S_\Delta^{\mathrm{ret}}(\Gamma, a, b) := \sum_{m=0}^{n-1} \Gamma_{t_m,t_{m+1}} - \Gamma_{t_m,t_m} \qquad S_\Delta^{\mathrm{adv}}(\Gamma, a, b) = \sum_{m=0}^{n-1} \Gamma_{t_{m+1},t_{m+1}} - \Gamma_{t_{m+1},t_m}.$$

As usual, a Γ-Riemann sum with respect to a tagged subdivision $\Delta : a = t_0 \le t_1 \cdots \le t_{2n} = b$ such that $t_{2m+1} = t_{2m}$ (resp. $t_{2m+1} = t_{2m+2}$) for $m = 0, \cdots, n-1$ is just a retarded (resp. advanced) Riemann-Young sum for the subdivision Δ_*.

Recall that if $\lim_{\mathrm{mesh}(\Delta)\downarrow 0} S_\Delta(\Gamma, a, b)$ exists, we denote it by $\int_a^b (\delta\Gamma)$ and sometimes call it the integral of Γ on the interval $[a,b]$. More precisely the existence of the limit means that there an element of E, denoted by $\int_a^b(\delta\Gamma)$ such that for any $\varepsilon > 0$ there is $\delta > 0$ such that $\|S_\Delta(\Gamma, a, b) - \int_a^b(\delta\Gamma)\|_E \le \varepsilon$ whenever $\mathrm{mesh}(\Delta) \le \delta$. We denote by $m_\varepsilon(\Gamma, a, b) \in [0, +\infty]$ the supremum of those δs: one has $\|S_\Delta(\Gamma, a, b) - \int_a^b(\delta\Gamma)\|_E \le \varepsilon$ whenever $\mathrm{mesh}(\Delta) \le m_\varepsilon(\Gamma, a, b)$ but for every $\delta > m_\varepsilon(\Gamma, a, b)$ there is at least one subdivision Δ with mesh $\le \delta$ such that $\|S_\Delta(\Gamma, a, b) - \int_a^b(\delta\Gamma)\|_E > \varepsilon$.

We shall sometimes concentrate only on retarded or advanced subdivisions, and we define the existence condition for $\int_a^b(\delta\Gamma)^{\mathrm{ret}}$ or $\int_a^b(\delta\Gamma)^{\mathrm{adv}}$ accordingly, leading to introduce m^{ret} and m^{adv} analogously by using only retarded and advanced Γ-Riemann sums.

Obviously, the existence of $\int_a^b(\delta\Gamma)$ implies the existence of the retarded and advanced versions (and then the three are equal), and the obvious inequalities $m^{\mathrm{ret}}, m^{\mathrm{adv}} \ge m$ hold (for given $\varepsilon, \Gamma, a, b$). We shall show a converse below.

4.A Generic properties of Γ-Riemann sums: proofs

We turn to the proof of the results announced in the main text.

We start with the more than elementary:

Reduction, Summands with vanishing Riemann sums If $Y\colon [a,b] \to V$ is an arbitrary function then $\Lambda\colon [a,b]^2 \to V$, $(s,t) \mapsto Y_s$ is such that $S_\Delta(\Lambda, a, b) = 0$ for every tagged subdivision of $[a,b]$. By linearity, for any $\Lambda\colon [a,b]^2 \to V$, $S_\Delta(\Gamma + \Lambda, a, b) = S_\Delta(\Gamma, a, b)$ for every tagged subdivision of $[a,b]$.

In particular, taking $Y_s := -\Gamma_{s,s}$ for $s \in [a,b]$, we see that is is harmless to replace Γ by $\Gamma^{\mathrm{red}}_{s,t} := \Gamma_{s,t} - \Gamma_{s,s}$ and assume that Γ vanishes on the diagonal $\{(t,t),\ t \in [a,b]\}$ of $[a,b]^2$. We shall often do so.

Locality If Γ vanishes in a neighborhood of the diagonal, then $\int_a^b (\delta\Gamma)$ exists and is 0. By linearity, the existence and then the value, of $\int_a^b (\delta\Gamma)$ depend only on the germ of Γ along the diagonal, or in physical term on the short distance behavior of $\Gamma_{s,t}$, $|t-s|$ small.

Proof. If $\Gamma_{s,t} = 0$ for $|t-s| \leq \delta$ then Γ-Riemann sums along subdivisions of mesh $\leq \delta$ vanish identically. $\square$

There is an obvious analog for advanced and retarded Γ-Riemann sums.

The next results require an argument.

Subintervals Let $a \leq s \leq t \leq b$. We can restrict Γ to $[s,t]^2$. If $\int_a^b (\delta\Gamma)$ exists then so does $\int_s^t (\delta\Gamma)$.

Proof. The existence of $\int_s^t (\delta\Gamma)$ follows from the Cauchy criterion. Indeed, let $\varepsilon > 0$ and let $\delta > 0$ be such that $\|S_\Delta(\Gamma, a, b) - \int_a^b (\delta\Gamma)\|_E \leq \varepsilon/2$ if $\mathrm{mesh}(\Delta) \leq \delta$. Let Δ', Δ'' be two tagged subdivisions of $[s,t]$ both with mesh $\leq \delta$ and let Δ^a, Δ^b we any fixed tagged subdivisions of $[a,s]$ and $[t,b]$ respectively, both with mesh $\leq \delta$. Let $\overline{\Delta}'$ (resp. $\overline{\Delta}''$) be the tagged subdivision of $[a,b]$ obtained by "gluing" Δ^a Δ' (resp.Δ'') and Δ^b. Then $\mathrm{mesh}(\overline{\Delta}'), \mathrm{mesh}(\overline{\Delta}'') \leq \delta$ and $S_{\Delta'}(\Gamma, s, t) - S_{\Delta''}(\Gamma, s, t) = S_{\overline{\Delta}'}(\Gamma, a, b) - S_{\overline{\Delta}''}(\Gamma, a, b)$ (the contributions of the peripheral intervals cancel) so

$$\begin{aligned}
\|S_{\Delta'}(\Gamma, s, t) - S_{\Delta''}(\Gamma, s, t)\|_E &= \|S_{\overline{\Delta}'}(\Gamma, a, b) - S_{\overline{\Delta}''}(\Gamma, a, b)\|_E \\
&\leq \|S_{\overline{\Delta}'}(\Gamma, a, b) - \textstyle\int_a^b (\delta\Gamma)\|_E + \|\textstyle\int_a^b (\delta\Gamma) - S_{\overline{\Delta}'}(\Gamma, a, b)\|_E \\
&\leq \varepsilon/2 + \varepsilon/2 = \varepsilon.
\end{aligned}$$

By the Cauchy criterion, this shows that Γ is integrable on $[s,t]$.

Integrability on subintervals also holds for the advanced and retarded versions. the proof is the same, *mutatis mutandis*. $\square$

As a corollary we obtain:

Chasles' relation Suppose $a \leq c \in \mathbb{R}$. If $\Gamma\colon [a,b]^2 \to V$ is such that $\int_a^b (\delta\Gamma)$ is defined, then so are $\int_a^s (\delta\Gamma)$ and $\int_s^b (\delta\Gamma)$ for any $s \in [a,b]$ and Chasles' relation $\int_a^b (\delta\Gamma) = \int_a^s (\delta\Gamma) + \int_s^b (\delta\Gamma)$ holds.

Proof. The existence of $\int_a^s(\delta\Gamma)$ and $\int_s^b(\delta\Gamma)$ follows from the previous result. Then Chasles' relation follows by considering a subclass of tagged subdivisions of $[a,b]$, those for which t is a subdivision point, and letting the mesh go to 0.Note that Chasles relation implies the existence of a function $\check{\Gamma}\colon [a,b] \to E$ such that $\int_t^t(\delta\Gamma) = \check{\Gamma}_t - \check{\Gamma}_s$, and $\check{\Gamma}$ is uniquely defined if we add the condition that $\check{\Gamma}_a = 0$.

Chasles relation also holds for the advanced and retarded versions. The proof is the same, *mutatis mutandis*. □

This yields yet another corollary.

Subintervals (II) Let $a \leq s \leq t \leq b$. Suppose the $\int_a^b(\delta\Gamma)$ exists. Then $\mathfrak{m}_\varepsilon(\Gamma, s, t) \leq \mathfrak{m}_\varepsilon(\Gamma, a, b)$.

Proof. Let $\varepsilon > 0$. Fix a tagged subdivision Δ of $[s,t]$ with mesh $\leq \mathfrak{m}_\varepsilon(\Gamma, a, b)$. If Δ^a, Δ^b are any tagged subdivisions of $[a,s]$ and $[t,b]$ respectively, both with mesh $\leq \mathfrak{m}_\varepsilon(\Gamma, a, b)$ let $\overline{\Delta}$ be the tagged subdivision of $[a,b]$ obtained by "gluing" Δ^a Δ and Δ^b. Then $\mathrm{mesh}(\overline{\Delta}) \leq \mathfrak{m}_\varepsilon(\Gamma, a, b)$. As $S_{\overline{\Delta}}(\Gamma, a, b) = S_{\Delta^a}(\Gamma, a, s) + S_\Delta(\Gamma, s, t) + S_{\Delta^b}(\Gamma, t, b)$, using Chasles' relation we have

$$\|\left(S_{\Delta^a}(\Gamma, a, s) - \int_a^s(\delta\Gamma)\right) + \left(S_\Delta(\Gamma, s, t) - \int_s^t(\delta\Gamma)\right) + \left(S_{\Delta^b}(\Gamma, t, b) - \int_t^b(\delta\Gamma)\right)\|_E \leq \varepsilon.$$

Using the **Subintervals** item, we know that $S_{\Delta^a}(\Gamma, a, s) - \int_a^s(\delta\Gamma)$ can be made as small as we please by taking Δ^a fine enough, and the same holds for $S_{\Delta^b}(\Gamma, t, b) - \int_t^b(\delta\Gamma)$ by taking Δ^b fine enough. Thus $\|S_\Delta(\Gamma, s, t) - \int_s^t(\delta\Gamma)\|_E \leq \varepsilon$. As Δ is an arbitrary tagged subdivision of $[s,t]$ with mesh $\leq \mathfrak{m}_\varepsilon(\Gamma, a, b)$, we infer that $\mathfrak{m}_\varepsilon(\Gamma, s, t) \leq \mathfrak{m}_\varepsilon(\Gamma, a, b)$.

This inequality also holds for the advanced and retarded versions. The proof is the same, *mutatis mutandis*. □

Integrability conditions, locality (II) In case $\dim E < +\infty$, if $\int_a^b(\delta\Gamma)$ exists then $\sum_{m=0}^{n-1}\|\int_{t_m}^{t_{m+1}}(\delta\Gamma) - (\Gamma_{t_m,t_{m+1}} - \Gamma_{t_m,t_m})\|_E$ and $\sum_{m=0}^{n-1}\|\int_{t_m}^{t_{m+1}}(\delta\Gamma) - (\Gamma_{t_{m+1},t_{m+1}} - \Gamma_{t_{m+1},t_m})\|_E$ are small when the mesh of the subdivision $\Delta : a = t_0 < t_1 \cdots < t_n = b$ is small. Conversely, without restrictions on $\dim E$ if there is a function $\check{\Gamma}\colon [a,b] \to \mathbb{R}$ such that $\sum_{m=0}^{n-1}\|(\check{\Gamma}_{t_{m+1}} - \check{\Gamma}_{t_m}) - (\Gamma_{t_m,t_{m+1}} - \Gamma_{t_m,t_m})\|_E$ and $\sum_{m=0}^{n-1}\|(\check{\Gamma}_{t_{m+1}} - \check{\Gamma}_{t_m}) - (\Gamma_{t_{m+1},t_{m+1}} - \Gamma_{t_{m+1},t_m})\|_E$ are small when the mesh of the subdivision $\Delta : a = t_0 < t_1 \cdots < t_n = b$ is small then $\int_a^b(\delta\Gamma)$ exists and equals $\check{\Gamma}_b - \check{\Gamma}_a$.

The conditions and the conclusions only involve the reduced form of Γ and we may assume by **Reduction** that Γ vanishes on the diagonal to lighten the notations.

Proof of the direct statement. We do a bit better: under the weaker condition that $\int_a^b(\delta\Gamma)^{\mathrm{ret}}$ exists, we prove that $\sum_{m=0}^{n-1}\|\int_{t_m}^{t_{m+1}}(\delta\Gamma)-(\Gamma_{t_m,t_{m+1}}-\Gamma_{t_m,t_m})\|_E$ is small when the mesh of the subdivision $\Delta : a=t_0<t_1\cdots<t_n=b$ is small. The proof of the second inequality assuming only that $\int_a^b(\delta\Gamma)^{\mathrm{ret}}$ exists is analogous.

We first prove the statement when $E=\mathbb{R}$. Suppose thus that $\int_a^b(\delta\Gamma)^{\mathrm{ret}}$ exists and fix $\varepsilon>0$. Let Δ be a subdivision of $[a,b]$ with mesh $\le \mathfrak{m}_\varepsilon^{\mathrm{ret}}(\Gamma,a,b)$. Let $M_+:=\{m\in[\![0,n-1]\!],\ \int_{t_m}^{t_{m+1}}(\delta\Gamma)^{\mathrm{ret}}-\Gamma_{t_m,t_{m+1}}\ge 0\}$ and M_- be the complement, $M_-:=\{m\in[\![0,n-1]\!],\ \int_{t_m}^{t_{m+1}}(\delta\Gamma)^{\mathrm{ret}}-\Gamma_{t_m,t_{m+1}}<0\}$. If the $\Delta^{(m)}$s, $m\in M_-$, are arbitrary subdivisions of $[t_m,t_{m+1}]$ we may "glue" them with the intervals $[t_m,t_{m+1}]$, $m\in M_+$ (left untouched and each seen as the simplest subdivision of $[t_m,t_{m+1}]$) to get a subdivision of $[a,b]$ which we denote by $\overline{\Delta}$. Using Chasles relation we get

$$\begin{aligned}\int_a^b(\delta\Gamma)^{\mathrm{ret}}-S_{\overline{\Delta}}(\Gamma,a,b) &= \sum_{m\in M_+}\left(\int_{t_m}^{t_{m+1}}(\delta\Gamma)^{\mathrm{ret}}-\Gamma_{t_m,t_{m+1}}\right)\\ &\quad+\sum_{m\in M_-}\left(\int_{t_m}^{t_{m+1}}(\delta\Gamma)^{\mathrm{ret}}-S^{\mathrm{ret}}_{\Delta^{(m)}}(\Gamma,t_m,t_{m+1})\right).\end{aligned}$$

The absolute value[6]of the left-hand side is $\le\varepsilon$ because $\mathrm{mesh}(\overline{\Delta})\le\mathfrak{m}_\varepsilon^{\mathrm{ret}}(\Gamma,a,b)$. On the right-hand side, using the **Subintervals** item we know that that the absolute value of each term $\int_{t_m}^{t_{m+1}}(\delta\Gamma)-S^{\mathrm{ret}}_{\Delta^{(m)}}(\Gamma,t_m,t_{m+1})$, $m\in M_-$ can be made as small as we please by taking the $\Delta^{(m)}$s fine enough. This implies that in fact $\left|\sum_{m\in M_+}\left(\int_{t_m}^{t_{m+1}}(\delta\Gamma)-\Gamma_{t_m,t_{m+1}}\right)\right|\le\varepsilon$. But $\int_{t_m}^{t_{m+1}}(\delta\Gamma)-\Gamma_{t_m,t_{m+1}}\ge 0$ for $m\in M_+$ so $\sum_{m\in M_+}\left|\int_{t_m}^{t_{m+1}}(\delta\Gamma)-\Gamma_{t_m,t_{m+1}}\right|\le\varepsilon$. A similar reasoning (this time refining Δ on the intervals associated to M_+, leaving the intervals associated to M_- untouched) shows that $\sum_{m\in M_-}\left|\int_{t_m}^{t_{m+1}}(\delta\Gamma)-\Gamma_{t_m,t_{m+1}}\right|\le\varepsilon$. Putting things together, we have proven that for each subdivision Δ of $[a,b]$ with mesh $\le\mathfrak{m}_\varepsilon^{\mathrm{adv}}(\Gamma,a,b)$

$$\sum_{m=0}^{n-1}\left|\int_{t_m}^{t_{m+1}}(\delta\Gamma)^{\mathrm{ret}}-(\Gamma_{t_m,t_{m+1}}-\Gamma_{t_m,t_m})\right|\le 2\varepsilon,$$

valid if Γ vanishes on the diagonal and then in general by **Reduction**.

If $d:=\dim E<+\infty$ we proceed in two steps. First endow E with a basis $e_1,\cdots,e_d$ and prove the result for the L^1 norm with respect to that basis: $\|v\|_E=\sum_{i=1}^d|v_i|$ if $v=v^ie_i$. Then we may decompose $\Gamma=\Gamma^ie_i$. It is plain that retarded Γ-Riemann sums have a small-mesh limit if an only if this holds for each Γ^i and we may apply the previous reasoning to each Γ^i. This yields immediately that for each subdivision Δ of $[a,b]$ with mesh $\le\mathfrak{m}_\varepsilon^{\mathrm{ret}}(\Gamma,a,b)$

$$\sum_{m=0}^{n-1}\|\int_{t_m}^{t_{m+1}}(\delta\Gamma)^{\mathrm{ret}}-(\Gamma_{t_m,t_{m+1}}-\Gamma_{t_m,t_m})\|_E\le 2\varepsilon\dim E,$$

This settles the case of the L^1 norm. As all norms are equivalent in finite dimensions, the announced result holds for arbitrary norms: $\sum_{m=0}^{n-1} \| \int_{t_m}^{t_{m+1}} (\delta\Gamma)^{\text{ret}} - (\Gamma_{t_m,t_{m+1}} - \Gamma_{t_m,t_m}) \|_E$ is small if $\Delta : a = t_0 < t_1 \cdots < t_n = b$ has a small mesh.[7]

These arguments is easily transposed to advanced Γ-Riemann sums to get that if $\int_a^b (\delta\Gamma)^{\text{adv}}$ exists then for each subdivision Δ of $[a, b]$ with mesh $\le \mathfrak{m}_\varepsilon^{\text{adv}}(\Gamma, a, b)$,

$$\sum_{m=0}^{n-1} \left| \int_{t_m}^{t_{m+1}} (\delta\Gamma)^{\text{adv}} - (\Gamma_{t_{m+1},t_{m+1}} - \Gamma_{t_{m+1},t_m}) \right| \le 2\varepsilon \dim E$$

if E is endowed with the L^1 norm.

As $\mathfrak{m}^{\text{ret}}, \mathfrak{m}^{\text{adv}} \ge \mathfrak{m}$ we have proven the direct statement.

□

Proof of the reverse statement (this is simpler). Fix $\varepsilon > 0$ and chose δ such that such that $\sum_{m=0}^{n-1} \|(\check{\Gamma}_{t_{m+1}} - \check{\Gamma}_{t_m}) - (\Gamma_{t_m,t_{m+1}} - \Gamma_{t_m,t_m})\|_E$ and $\sum_{m=0}^{n-1} \|(\check{\Gamma}_{t_{m+1}} - \check{\Gamma}_{t_m}) - (\Gamma_{t_{m+1},t_{m+1}} - \Gamma_{t_{m+1},t_m})\|_E$ are $\le \varepsilon/2$ whenever $\Delta : a = t_0 < t_1 \cdots < t_n = b$ is a subdivision of $[a, b]$ with mesh $\le \delta$. If $\Delta : a = t_0 \le t_1 \cdots \le t_{2n} = b$ is a tagged subdivision of $[a, b]$ we write

$$\begin{aligned} \check{\Gamma}_b - \check{\Gamma}_a - S_\Delta(\Gamma, a, b) &= \sum_{m=0}^{n-1} ((\check{\Gamma}_{t_{2m+2}} - \check{\Gamma}_{t_{2m+1}}) - (\Gamma_{t_{2m+2},t_{2m+1}} - \Gamma_{t_{2m+1},t_{2m+1}})) \\ &+ \sum_{m=0}^{n-1} ((\check{\Gamma}_{t_{2m+1}} - \check{\Gamma}_{t_{2m}}) - (\Gamma_{t_{2m+1},t_{2m+1}} - \Gamma_{t_{2m+1},t_{2m}})). \end{aligned}$$

We claim that

$$\sum_{m=0}^{n-1} \|_{\|} E(\check{\Gamma}_{t_{2m+2}} - \check{\Gamma}_{t_{2m+1}}) - (\Gamma_{t_{2m+2},t_{2m+1}} - \Gamma_{t_{2m+1},t_{2m+1}}))_E \le \sum_{m=0}^{2n-1} \|(\check{\Gamma}_{t_{m+1}} - \check{\Gamma}_{t_m}) - (\Gamma_{t_m,t_{m+1}} - \Gamma_{t_m,t_m})\|_E \le \varepsilon/2$$

The first inequality holds because the terms in the first sum are those with odd m in the second sum. The set $\{t_0, \cdots, t_{2n}\}$ defines a subdivision of $[a, b]$ denoted by Δ_{**} (see Appendix A.1 for details) which has mesh $\le \delta$ and by the hypothesis applied to this subdivision one gets the second inequality. Observe that due to coincidences Δ_{**} may have less than $2n$ parts because, for some ms, t_m and t_{m+1} can be equal, but then they count automatically for 0 in the second sum. The same argument shows that

$$\sum_{m=0}^{n-1} \|_{\|} E(\check{\Gamma}_{t_{2m+1}} - \check{\Gamma}_{t_{2m}}) - (\Gamma_{t_{2m+1},t_{2m+1}} - \Gamma_{t_{2m+1},t_{2m}}))_E \le \sum_{m=0}^{2n-1} \|(\check{\Gamma}_{t_{m+1}} - \check{\Gamma}_{t_m}) - (\Gamma_{t_{m+1},t_{m+1}} - \Gamma_{t_{m+1},t_m})\|_E \le \varepsilon.$$

Thus we have shown that $\check{\Gamma}_b - \check{\Gamma}_a - S_\Delta(\Gamma, a, b)$ has norm $\le \varepsilon$ whenever Δ has mesh $\le \delta$, so that $\int_a^b (\delta\Gamma)$ exists and equals $\check{\Gamma}_b - \check{\Gamma}_a$. □

[6] Recall the $E = \mathbb{R}$ in this part of the argument.

[7] Beware however that to get explicit bounds one has to take into account that though we did not emphasize it in the notation, $\mathfrak{m}$ depends on the choice of norm.

We note the following corollary

From retarded and advanced to general integrals Suppose that $\dim E < +\infty$. Assume that $\int_a^b (\delta\Gamma)^{\mathrm{ret}}$, $\int_a^b (\delta\Gamma)^{\mathrm{adv}}$ both exist. If $\int_s^t (\delta\Gamma)^{\mathrm{ret}} = \int_s^t (\delta\Gamma)^{\mathrm{adv}}$ for $s \leq t \in [a,b]$ then the general integral $\int_a^b (\delta\Gamma)$ exists (and coincides with the retarded-advanced versions). Else $\int_a^b (\delta\Gamma)$ does not exist.

Proof. Start with the first statement. The first part of the previous result gives the inequalities needed in the second part. Thus part however requires a consistent $\check{\Gamma}$ – which amounts to the equality $\int (\delta\Gamma)^{\mathrm{ret}} = \int (\delta\Gamma)^{\mathrm{adv}}$– to be carried. □

If $\varphi \colon [c,d] \to [a,b]$ is an arbitrary function we can define the pullback $\varphi^*\Gamma \colon [c,d]^2 \to E$ of Γ by φ by $(\varphi^*\Gamma)_{u,v} := \Gamma_{\varphi(u),\varphi(v)}$.

Reparameterization invariance If φ is continuous and increasing, and if $\int_a^b (\delta\Gamma)$ exists then $\int_c^d (\delta(\varphi^*\Gamma))$ exists as well and the two are equal.

Proof. Being increasing φ maps tagged subdivisions : if $\Delta : c = u_0 \leq u_1 \cdots \leq u_{2n} = d$ is a tagged subdivision of $[c,d]$, we set $\varphi(\Delta) : a = \varphi(u_0) \leq \varphi(u_1) \cdots \leq \varphi(u_{2n}) = b$. Let $\varepsilon > 0$ and let $\tilde{\delta} > 0$ be such that $\|S_{\tilde{\Delta}}(\Gamma, a, b) - \int_a^b (\delta\Gamma)\|_E \leq \varepsilon$ for $\mathrm{mesh}(\tilde{\Delta}) \leq \varepsilon$. Being continuous on the compact interval $[c,d]$, φ is uniformly continuous there: there is a $\delta > 0$ such that $\mathrm{mesh}(\varphi(\Delta)) \leq \tilde{\delta}$ whenever Δ is a tagged subdivision of $[c,d]$ with $\mathrm{mesh}(\Delta) \leq \delta$. If this is the case, as $S_\Delta(\varphi^*\Gamma, c, d) = S_{varphi(\Delta)}(\Gamma, a, b)$ we obtain $\|S_\Delta(\varphi^*\Gamma, c, d) - \int_a^b (\delta\Gamma)\|_E \leq \varepsilon$. Thus the summand $\varphi^*\Gamma$ is integrable on $[c,d]$ and $\int_c^d (\delta(\varphi^*\Gamma)) = \int_a^b (\delta\Gamma)$. □

4.B The sewing lemma: proofs

We now turn to a useful setting that guaranties the existence of $\int (\delta\Gamma)$. This setting is not optimal because the conditions are not reparameterization invariant

To study the small mesh behavior of Γ-Riemann sums, a first step is to control what happens to, say, a retarded sum, when a single subdivision point is added to a subdivision. Suppose that $[s,u]$ is a part of a subdivision Δ and point $t \in]s,u[$ is inserted. Then

$$S_\Delta^{\mathrm{ret}}(\Gamma, a, b) - S_{\Delta \cup \{t\}}^{\mathrm{ret}}(\Gamma, a, b) = \Gamma_{s,u} - \Gamma_{s,t} - \Gamma_{t,u} + \Gamma_{t,t}.$$

This motivates the definition of a triangular summand, which we repeat here.

Triangular summand Let $\gamma \in \mathbb{R}$. A summand $\Gamma \colon [a,b]^2 \to E$ is triangular with exponent γ if there is a constant K such that for every $s,t,u \in [a,b]$

$$\|\Gamma_{s,u} - \Gamma_{s,t} - \Gamma_{t,u} + \Gamma_{t,t}\|_E \leq K|s,t,u|^\gamma,$$

where $|s,t,u|$ denotes the length of the smallest interval containing s,t,u, i.e.

$$|s,t,u| := \max(|t-s|, |u-t|, |u-s|) = \max(s,t,u) - \min(s,t,u),$$

a quantity which vanishes if and only of the three points s,t,u coincide, and equals $u - s$ if $s \leq t \leq u$.

We denote by $K_{[a,b]}(\Gamma)$ or simply $K_{[a,b]}$ the best possible constant, i.e.

$$K_{[a,b]}(\Gamma) := \sup_{s,t,u \in [a,b],\, |s,t,u|>0} \frac{\|\Gamma_{s,u} - \Gamma_{s,t} - \Gamma_{t,u} + \Gamma_{t,t}\|}{|s,t,u|^{\gamma}},$$

which makes sense only if $[a,b]$ is a nontrivial interval, i.e. if $a < b$. By convention, we set $K_{[a,b]}(\Gamma) := 0$ if $a = b$.

Now for our main result of this section:

The sewing lemma Suppose $\Gamma\colon [a,b]^2 \to E$ is triangular with exponent $\gamma > 1$. Then the Γ-Riemann sums $S_\Delta(\Gamma, a, b)$ have a limit when $\mathrm{mesh}(\Delta) \downarrow 0$. We denote this limit by $\int_a^b (\delta\Gamma)$.

Chasles' relation If $a \le s \le t \le u \le b$, the restriction of Γ to $[s,t]^2$, $[t,u]^2$ and $[s,u]^2$ is also triangular and Chasles relation

$$\int_s^u (\delta\Gamma) = \int_s^t (\delta\Gamma) + \int_t^u (\delta\Gamma)$$

holds.

The Young-Loëve inequality Then

$$\|\int_a^b (\delta\Gamma) - \Gamma_{a,b}\|_E \le \frac{1}{1-2^{1-\gamma}} K_{[a,b]}(\Gamma)(b-a)^{\gamma}$$

holds.

If Γ is triangular on $[a,b]^2$ and $[s,t] \subset [a,b]$ is a nontrivial interval the restriction of Γ to $[s,t]^2$ is plainly also triangular (for the same exponent) and clearly $K_{[c,d]} \le K_{[a,b]}$. Moreover, we have already proven Chasles' relation under the sole assumption that $\int_a^b (\delta\Gamma)$ exists. So we concentrate on the first and third assertions.

Most of the proof relies on the use of retarded sums S_Δ^{ret}. It is split into a number of intermediate results. We often write simply K for $K_{[a,b]}$ remembering that this constant decreases if $[a,b]$ is replaced by $[s,t] \subset [a,b]$.

From now on, taking advantage of **Reduction**, we assume that Γ vanishes on the diagonal. Then retarded sums simply read $S_\Delta^{\mathrm{ret}}(\Gamma, a, b) := \sum_{m=0}^{n-1} \Gamma_{t_m, t_{m+1}}$, and being triangular reads $\|\Gamma_{s,u} - \Gamma_{s,t} - \Gamma_{t,u}\| \le K|s,t,u|^{\gamma}$, which amusingly implies in turn that Γ vanishes on the diagonal, as seen by taking $s = t = u$.

We shall use dyadic subdivisions of $[a,b]$ to define refinements. This is a bit cumbersome because of boundary effects when a, b are not themselves dyadic numbers, but there are some advantages. In particular we construct the "integral" simultaneously for all intervals on which Γ is defined and has the required properties, i.e. we obtain the integral as a function of a and b.

We may (and shall) assume that $a < b$. Recall that, for $k \in \mathbb{Z}$, $\mathbb{D}^k := \{l/2^k, l \in \mathbb{Z}\}$ is the set of dyadic rationals of order k.

Dyadic subdivisions of $[\mathbf{a},\mathbf{b}]$, $\mathbf{a}<\mathbf{b}$ Let $\Delta_k := \Delta_k([a,b])$ be the subdivision defined by ordering the set $\{a,b\}\cup(]a,b[\cap\mathbb{D}^k)$. Let $n_k \geq 1$ be the number of parts of Δ_k and write $\Delta_k : a = t_{k,0} < \cdots < t_{k,n_k} = b$. Let $S_k^{\mathrm{ret}} := S_{\Delta_k}^{\mathrm{ret}}(\Gamma)$ be the retarded sum associated to Δ_k.

Claim 1 The difference $S_{k+1}^{\mathrm{ret}} - S_k^{\mathrm{ret}}$ is bounded by $Kn_k2^{-k\gamma}$.

Proof. The parts of Δ_k have length at most 2^{-k}. When going from Δ_k to Δ_{k+1} two things may happen to a part of Δ_k. The first possibility is that this part is also a part of Δ_{k+1}, and then it does not contribute to $S_{k+1}^{\mathrm{ret}} - S_k^{\mathrm{ret}}$. The second possibility is that it splits in two parts, but then by the triangular property of Γ the corresponding contribution to $S_{k+1}^{\mathrm{ret}} - S_k^{\mathrm{ret}}$ is at most $K2^{-k\gamma}$. The total number of parts of Δ_k is n_k, leading to the announced bound. □

Claim 2 There is a largest $l \in \mathbb{Z}$ such that $]a,b[\cap D_l$ is a singleton. The inequality $2^{-l} < 2(b-a)$ holds.

Proof. Obviously the cardinal of $]a,b[\cap D_k$ is an increasing function of k. Define $j\in\mathbb{Z}$ by the inequality $2^{-j-1} < b-a \leq 2^{-j}$. Then $]a,b[$ contains at most 1 point of D_j, at least 1 point of D_{j+1} and at least 2 points of D_{j+2}. If $]a,b[$ contains exactly 1 point of D_{j+1} take $l=j+1$. If $]a,b[$ contains more than 1 point of D_{j+1}, it contains two consecutive points of D_{j+1}, and one of them is in D_j so that $]a,b[$ contains exactly 1 point of D_j and then take $l=j$. The announced inequality follows. □

Claim 3 The sequence S_k^{ret} has a limit, which we denote by $\int_a^b(\delta\Gamma)^{\mathrm{dyad}}$, as $k\to+\infty$ and
$$\|\int_a^b(\delta\Gamma)^{\mathrm{dyad}} - \Gamma_{a,b}\|_E \leq C(\gamma)K(b-a)^\gamma$$
for some "constant" C which depends only on γ.

Proof. First note that $n_k - 2 \leq (b-a)2^k$ because appart from the parts containing a and b, all others parts of Δ_k have length 2^{-k}. Thus, using **Claim 1**
$$\|S_{k+1}^{\mathrm{ret}} - S_k^{\mathrm{ret}}\|_E \leq K\left((b-a)2^k+2\right)2^{-k\gamma}.$$
Hence $\|S_{k+1}^{\mathrm{ret}} - S_k^{\mathrm{ret}}\|_E$ is bounded by a sequence decreasing exponentially to 0 when $k\to+\infty$ (because $\gamma>1$). As $(E,\|\ \|_E)$ is complete, S_k^{ret} has a limit when $k\to+\infty$, which we denote by $\int_a^b(\delta\Gamma)^{\mathrm{dyad}}$. Taking l as in **Claim 2** and $k>l$ we obtain
$$\|\Gamma_{a,b} - S_k^{\mathrm{ret}}\|_E \leq \|\Gamma_{a,b} - S_l^{\mathrm{ret}}\|_E + \|S_l^{\mathrm{ret}} - S_{l+1}^{\mathrm{ret}}\|_E + \cdots + \|S_{k-1}^{\mathrm{ret}} - S_k^{\mathrm{ret}}\|_E.$$
Taking the limit $k\to+\infty$ we find
$$\|\int_a^b(\delta\Gamma)^{\mathrm{dyad}} - \Gamma_{a,b}\|_E \leq \|\Gamma_{a,b} - S_l^{\mathrm{ret}}\|_E + K\sum_{j\geq 0}\left((b-a)2^{l+j}+2\right)2^{-(l+j)\gamma}.$$

The triangular property of Γ says precisely that the first term on the right is bounded by $K(b-a)^{\gamma}$ because Δ_l contains a single subdivision point in $]a,b[$. The sum is $(b-a)2^{-l(\gamma-1)}\frac{1}{1-2^{1-\gamma}} + 2^{-l\gamma}\frac{2}{1-2^{-\gamma}}$. As $\gamma > 1$ we may bound $2^{-l(\gamma-1)}$ and $2^{-l\gamma}$ using the inequality $2^{-l} < 2(b-a)$ established in **Claim 2** This leads to the announced bound with $C(\gamma) := 1 + \frac{2^{\gamma-1}}{1-2^{1-\gamma}} + \frac{2^{\gamma+1}}{1-2^{-\gamma}}$. This explicit value of C is of no interest but the bound on $\|\int_a^b(\delta\Gamma)^{\text{dyad}} - \Gamma_{a,b}\|_E$, especially the factor $(b-a)^{\gamma}$ it involves, is decisive for the end of the argument. □

Of course, if $a \le s \le t \le b$ we can restrict Γ to $[s,t]^2$ and the bound

$$\|\int_s^t(\delta\Gamma)^{\text{dyad}} - \Gamma_{s,t}\|_E \le C(\gamma)K(t-s)^{\Gamma}.$$

holds. Call it the general dyadic Young-Loëve inequality. The constant $K = K_{[a,b]}$ on the right-hand side could be replaced by the possibly smaller $K_{[s,t]}$.

Claim 4 The integral $\int_a^b(\delta\Gamma)^{\text{dyad}}$ satisfies Chasles relation.

Proof. Take $t \in [a,b]$. The restriction of Γ to $[a,t]^2$ and $[t,b]^2$ is triangular and $K_{[a,t]}, K_{[t,b]} \le K_{[a,b]}$, so that $\int_a^t(\delta\Gamma)^{\text{dyad}}$ and $\int_t^b(\delta\Gamma)^{\text{dyad}}$ are well-defined. Insert point t to refine the subdivision $\Delta_k([a,b])$ into another subdivision which we call $\Delta_k([a,b],t)$: it is nothing but the union of $\Delta_k([a,t])$ and $\Delta_k([t,b])$. The retarded sum for $\Delta_k([a,b],t)$ defines the dyadic approximation of order k to (and thus at large k is close to) $\int_a^t(\delta\Gamma)^{\text{dyad}} + \int_t^b(\delta\Gamma)^{\text{dyad}}$. But at most one part of Δ_k is split in $\Delta_k([a,b],t)$, and this part has length at most 2^{-k} so, by the triangular property of Γ, $\|\Delta_k - \Delta_k([a,b],t)\|_E \le K2^{-k\gamma}$. Taking the limit $k \to +\infty$,

$$\int_a^t(\delta\Gamma)^{\text{dyad}} + \int_t^b(\delta\Gamma)^{\text{dyad}} = \int_a^b(\delta\Gamma)^{\text{dyad}} \text{ for } a \le t \le b$$

as announced. □

This leads us to

Claim 5 Let Δ be a subdivision of $[a,b]$. The corresponding retarded Riemann-Young sum $S_\Delta^{\text{ret}}(\Gamma)$ satisfies

$$\|\int_a^b(\delta\Gamma)^{\text{dyad}} - S_\Delta^{\text{ret}}(\Gamma)\|_E \le C(\gamma)K(b-a)\text{mesh}(\Delta)^{\gamma-1}.$$

In particular, retarded sums for arbitrary subdivisions converge to $\int_a^b(\delta\Gamma)^{\text{dyad}}$ when their mesh goes to 0.

Proof. Write $\Delta : a = t_0 < t_1 \cdots < t_n = b$ so that $S_\Delta^{\text{ret}}(\Gamma) := \sum_{m=0}^{n-1}\Gamma_{t_m,t_{m+1}}$ and use Chasles relation to get

$$\|\int_a^b(\delta\Gamma)^{\text{dyad}} - S_\Delta^{\text{ret}}(\Gamma)\|_E \le \sum_{m=0}^{n-1}\|\int_{t_m}^{t_{m+1}}(\delta\Gamma)^{\text{dyad}} - \Gamma_{t_m,t_{m+1}}\|_E.$$

By the general dyadic Young-Loëve inequality, $\|\int_{t_m}^{t_{m+1}}(\delta\Gamma)^{\mathrm{dyad}}-\Gamma_{t_m,t_{m+1}}\|_E \leq C(\gamma)K(t_{m+1}-t_m)^\gamma$. Using $(t_{m+1}-t_m)^\gamma \leq (t_{m+1}-t_m)\mathrm{mesh}(\Delta)^{\gamma-1}$ and summing over m yields the announced formula. □

To stress the fact that retarded sums for *arbitrary* subdivisions are close to the same limit when their mesh is small, we denote this limit by $\int_a^b(\delta\Gamma)^{\mathrm{ret}}$ instead of $\int_a^b(\delta\Gamma)^{\mathrm{dyad}}$ from now on.

The convergence of sums S_Δ in general is now easily settled.

Proof of the sewing lemma. We show that Riemann-Young sums for arbitrary tagged subdivisions converge to $\int_a^b(\delta\Gamma)^{\mathrm{ret}}$ when their mesh goes to 0. Let $\Delta : a = t_0 \leq t_1 \leq \cdots \leq t_{2n} = b$ be a tagged subdivision of $[a,b]$ in n parts, and Δ_* its associated subdivision. Then

$$\begin{aligned} S_\Delta(\Gamma) &= \sum_{m=0}^{n-1} \Gamma_{t_{2m+1},t_{2m+2}} - \Gamma_{t_{2m+1},t_{2m}} \\ &= \sum_{m=0}^{n-1} \Gamma_{t_{2m},t_{2m+2}} + \sum_{m=0}^{n-1}\left(\Gamma_{t_{2m+1},t_{2m+2}} - \Gamma_{t_{2m+1},t_{2m}} - \Gamma_{t_{2m},t_{2m+2}}\right) \\ &= S^{\mathrm{ret}}_{\Delta_*}(\Gamma) + \sum_{m=0}^{n-1}\left(\Gamma_{t_{2m+1},t_{2m+2}} - \Gamma_{t_{2m+1},t_{2m}} - \Gamma_{t_{2m},t_{2m+2}}\right). \end{aligned}$$

Note that this is the only place in the whole argument where we face the expression $\Gamma_{s,u}-\Gamma_{s,t}-\Gamma_{s,u}$ in a case when the order is $t \leq s \leq u$ (in all other instances, the order is $s \leq t \leq u$). The triangular property yields, for $m = 0, \cdots, n-1$, that $\|\Gamma_{t_{2m+1},t_{2m+2}} - \Gamma_{t_{2m+1},t_{2m}} - \Gamma_{t_{2m},t_{2m+2}}\|_E \leq K(t_{2m+2}-t_{2m})^\gamma$, so that

$$\|S_\Delta(\Gamma) - S^{\mathrm{ret}}_{\Delta_*}(\Gamma)\|_E \leq K\sum_{m=0}^{n-1}(t_{2m+2}-t_{2m})^\gamma \leq K(b-a)\mathrm{mesh}(\Delta)^{\gamma-1}.$$

Combining with the bound for retarded sums in **Claim 5** we obtain

$$\|\int_a^b(\delta\Gamma)^{\mathrm{ret}} - S_\Delta(\Gamma)\|_E \leq (C(\gamma)+1)K(b-a)\mathrm{mesh}(\Delta)^{\gamma-1},$$

and in particular sums $S_\Delta(\Gamma)$ for arbitrary subdivisions are close to one and the same limit, $\int_a^b(\delta\Gamma)^{\mathrm{ret}}$, when their mesh is small. This common limit is denoted $\int_a^b(\delta\Gamma)$ from now on.

Chasles relation has been proven in **Claim 5** But is it also a consequence of the mere existence of $\int_a^b(\delta\Gamma)$ □

Proof of the Young-Loëve inequality. We prove the improved integral estimate

$$\|\int_a^b(\delta\Gamma) - \Gamma_{a,b}\|_E \leq \frac{1}{1-2^{1-\gamma}}K_{[a,b]}(\Gamma)(b-a)^\gamma.$$

For $k = 0, 1, 2, \cdots$ define Δ^k as the regular subdivision of $[a, b]$ in 2^k parts. We denote by $t_{k,m} := a + m\frac{b-a}{2^k}$ for $m = 0, \cdots, 2^k$ the subdivision points of Δ^k. Each part of Δ^k is made of two parts of Δ^{k+1}, and

$$S^{\mathrm{ret}}_{\Delta^k}(\Gamma) - S^{\mathrm{ret}}_{\Delta^{k+1}}(\Gamma) = \sum_{m=0}^{2^k-1} \left(\Gamma_{t_{k+1,2m},t_{k+1,2m+2}} - \Gamma_{t_{k+1,2m},t_{k+1,2m+1}} - \Gamma_{t_{k+1,2m+1},t_{k+1,2m+2}}\right).$$

Thus, using the triangular property of Γ for each term in the sum , we infer

$$\|S^{\mathrm{ret}}_{\Delta^k}(\Gamma) - S^{\mathrm{ret}}_{\Delta^{k+1}}(\Gamma)\|_E \leq K2^k(2^{-k}(b-a))^\gamma.$$

This is enough to show that $S^{\mathrm{ret}}_{\Delta^k}$ has a limit when $k \to +\infty$. Of course we already know from **Claim 5**that it converges, and that the limit is $\int_a^b(\delta\Gamma)$. But what we get is a better bound. Indeed, summing the series over k we obtain:

$$\|\int_a^b(\delta\Gamma) - \Gamma_{a,b}\|_E \leq \frac{K}{1-2^{1-\gamma}}(b-a)^\gamma$$

which is the integral estimate. □

This completes the proofs.

One may wonder why the subdivisions Δ^k are not used from the beginning to define the integral. The problem is that the subdivision points depend on the interval $[a, b]$ that is under consideration, and it is hard to compare the integrals in different intervals. In particular Chasles relation is quite hard to derive from this definition, and Chasles relation is crucial to show that arbitrary subdivisions can be used to compute the integral.

Though this is of limited practical value, we may improve the integral estimate into

$$\|\int_a^b(\delta\Gamma) - \Gamma_{a,b}\|_E \leq \frac{\tilde{K}}{1-2^{1-\gamma}}(b-a)^\gamma,$$

where $\tilde{K} = \tilde{K}_{[a,b]}(\Gamma) := \sup_{s,u\in[a,b],\, s<u} \frac{\|\Gamma_{s,u}-\Gamma_{s,(s+u)/2}-\Gamma_{(s+u)/2,u}\|_E}{|s-u|^\gamma}$ (the only combination that is relevant to compute with the subdivisions Δ^k) is smaller than K in general.

Uniqueness The function $s \leq t \in [a, b] \mapsto \int_s^t(\delta\Gamma) \in E$ is the only one satisfying Chasles relation and an integral estimate with respect to Γ.

Proof. Indeed, if I_s^t is another candidate, we infer that $\|\int_s^t(\delta\Gamma)-\Gamma_{s,t}\|_E = O((t-s)^\gamma)$ and $\|I_s^t - \Gamma_{s,t}\|_E = O((t-s)^\gamma)$ so that $\|I_s^t - \int_s^t(\delta\Gamma)\|_E = O((t-s)^\gamma)$, and then by Chasles relation that as a function of t the function $I_a^t - \int_a^t(\delta\Gamma)$ is γ-Hölder, i.e. constant because $\gamma > 1$ i.e 0 because 0 at $t = a$. □

We end the section with a simple observation. The interval $[a, b]$ and the number $\gamma > 1$ being given, we say that a function $\Gamma\colon [a, b]^2 \to E$ is negligible if $\|\Gamma_{s,t}\|_E \leq K|t-s|^\gamma$ for some unspecified constant K and for $s, t \in [a, b]$.

Clearly, the sum of two negligible functions is negligible, as is the product of a negligible function by a bounded scalar function, and negligible functions are bounded. Moreover, the following result holds:

Negligible summands Suppose that Γ is negligible. Then Γ is triangular with the same exponent γ and as $\gamma > 1$ by assumption, $\int(\delta\Gamma) \equiv 0$.

Proof. That Γ is triangular is clear, and then the corresponding Γ-Riemann sums for a subdivision Δ are $O(\mathrm{mesh}(\Delta)^{\gamma-1})$, hence go to 0 when $\mathrm{mesh}(\Delta) \downarrow 0$. □

This elementary fact can be seen as a generalization of **Locality**, p. 46: if Γ vanishes in a neighborhood of the diagonal, it is automatically negligible as soon as it is bounded on $[a, b]^2$.

CHAPTER 5

Young integrals

Suppose that $X, Y : [a, b] \to \mathbb{R}$ are two functions. In 1936 the mathematician Laurence C. Young asked about the possibility to make sense of a notion of integral $\int_a^b Y_t\,dX_t$:
– When $X_t = t$ and Y is Riemann integrable we have $\int_a^b Y_t\,dt$ at our disposal.
– More generally when X is differentiable with derivative $\dot{X}$ and $Y\dot{X}$ is Riemann integrable we can interpret $\int_a^b Y_t\,dX_t$ as $\int_a^b (Y\dot{X})_t\,dt$.

Young studied the possibility to define the integral in cases not covered by the above two, concretely as a limit of discrete sums.

5.1 Riemann-Young sums

Young integrals fit into the general discussion of Chapter 4. Indeed, the discrete sums introduced by Young amount to work with $\Gamma\colon [a, b]^2 \to \mathbb{R}$, $(s, t) \mapsto Y_s X_t$, or with a reduced version (which we use in the sequel) $\Gamma\colon [a, b]^2 \to \mathbb{R}$, $(, s, t) \mapsto Y_s(X_t - X_s)$ which leads to the same Γ-Riemann sums. If $\Delta : a = t_0 \leq t_1 \cdots \leq t_{2n} = b$ is a tagged subdivision of $[a, b]$, we define $A_\Delta(X, Y) := S_\Delta(\Gamma, a, b)$ which rearranges to yield

$$A_\Delta(X, Y) = \sum_{m=0}^{n-1} Y_{t_{2m+1}}(X_{t_{2m+2}} - X_{t_{2m}}).$$

Such sums are called Riemann-Young sums in the sequel.

In the same spirit, if $\Delta : a = t_0 < t_1 < \cdots < t_n = b$ is a subdivision of $[a, b]$ the retarded and advanced Riemann-Young sums are defined by

$$A^{\mathrm{ret}}_\Delta(X, Y) := \sum_{m=0}^{n-1} Y_{t_m}(X_{t_{m+1}} - X_{t_m}) \text{ and } A^{\mathrm{adv}}_\Delta(X, Y) := \sum_{m=0}^{n-1} Y_{t_{m+1}}(X_{t_{m+1}} - X_{t_m})$$

respectively, and these are again special cases of general Riemann-Young sums for tagged subdivisions.

We say that Y is integrable along X if the Riemann-Young sums have a limit at small mesh, and then the limit is denoted by $\int_a^b Y\,dX$ or $\int_a^b Y_t\,dX_t$. In that case, we use the term Young-integrable pair for the pair (X, Y). We say that X is the integrator and Y is the integrand.

Note that we use the word *along* in "Riemann-Young sum of Y along X" or "Y is integrable along X". This is because we view X as a path instead of a function. This is of course just a question of terminology. It becomes more natural if X and possibly Y have several components.

If $X_t = t$ for $t \in [a, b]$, Riemann-Young sums for (X, Y) reduce to Riemann sums for Y. Thus in that case $\int_a^b Y_t\,dX_t$ exists in the sense of Young if and only if Y is Riemann integrable on $[a, b]$ and then the two integrals are equal. We shall see later more (and more interesting) examples of Young integrability.

It is plain that for given X the space of Ys for which $\int_a^b Y\,dX$ exists is a vector space and the integral is linear in Y. Also, for fixed Y the the space of Xs for which $\int_a^b Y\,dX$ exists is a vector space and the integral is linear in X. Thus morally (why only "morally"?) $\int_a^b Y\,dX$ is bilinear in (X, Y).

There is one important result for Young integrals that follows from the definition but has no counterpart for the case of a general Γ.

Anti-symmetry, aka Integration by parts If Y is Young-integrable along the path X, then X is Young-integrable along the path Y and $\int_a^b Y_t\,dX_t + \int_a^b X_t\,dY_t = X_bY_b - X_aY_a$.

The basic nature of this result is combinatorial: if $\Delta : a = t_0 < t_1 < \cdots < t_n = b$ is any subdivision of $[a, b]$ then

$$A_\Delta^{\text{ret}}(X, Y) + A_\Delta^{\text{adv}}(Y, X) = A_\Delta^{\text{adv}}(X, Y) + A_\Delta^{\text{ret}}(Y, X) = X_bY_b - X_aY_a$$

by a telescopic sum argument. Thus if retarded sums for Y along X converge at small mesh, then so do advanced sums for X along Y. The case of arbitrary tagged subdivisions is easy. The details can be found in Appendix D

5.2 Integrals of the type $\int_a^b f(X)dX$

We start with a remarkably simple characterization for the existence of $\int_a^b X\,dX$. When is X integrable along itself? The answer is: if and only if X has vanishing 2-variation on $[a, b]$.

We insist on this elementary result for two main reasons:

– First, 2-variation played an important role in the discussion of the Euler scheme for $dY_t = Y_t\,dX_t$ in the motivation section, see Example 1.1 to 1.6. This may come as no surprise because the first order in the Born expansion leads to retarded Riemann-Young sum of X along itself.

– This gives another opportunity to stress the difference between pointwise integrals and stochastic integrals via the example of Brownian motion.

The point is that, if $\Delta : a = t_0 < t_1 < \cdots < t_n = b$ is a subdivision of $[a, b]$, the sum $A_\Delta^{\text{adv}}(X, X) + A_\Delta^{\text{ret}}(X, X)$ is telescopic (this is again the combinatorics behind integration by parts) while the difference is a sum of squares, namely

$$A_\Delta^{adv}(X,X) + A_\Delta^{ret}(X,X) = X_b^2 - X_a^2 \quad A_\Delta^{ret}(X,X) - A_\Delta^{adv}(X,X) = \sum_{m=0}^{n-1}(X_{t_{m+1}} - X_{t_m})^2.$$

Thus $A_\Delta^{ret}(X,X) - A_\Delta^{adv}(X,X)$ is precisely the 2-variation Q_Δ of X with respect to Δ as introduced in Example 1.1. Hence if X does not have vanishing 2-variation on $[a,b]$ (i.e.if the 2-variation of X with respect to Δ is not small when $\mathrm{mesh}(\Delta)$ is small) there is no hope for the existence of $\int_a^b X\,dX$. Conversely, if X has vanishing 2-variation on $[a,b]$ both advanced and retarded sums go to $(X_b^2 - X_a^2)/2$ at small mesh by the telescopic formula for $A_\Delta^{adv}(X,X) + A_\Delta^{ret}(X,X)$. Then a simple argument shows that the same limit is approached for general Riemann-Young sums.

There are many reasons why Brownian motion cannot have vanishing 2-variation on any interval. One relies on the well-known fact that for Brownian motion $\mathbb{E}(Q_\Delta) = b - a$. With a small effort we compute that $\mathbb{E}\left((b-a-Q_\Delta)^2\right) = 2\sum_{m=0}^{n-1}(t_{m+1}-t_m)^2 \leq 2(b-a)\mathrm{mesh}(\Delta)$ so that at small mesh Q_Δ goes to $b-a$ in mean square, and then almost surely along appropriate sequences. By the way, the above estimates are enough to justify the computation of the simplest non-trivial Itō integral[1]: when X is a Brownian motion $\int_a^b X_t\,dX_t \stackrel{\text{Itō}}{=} (X_b^2 - X_a^2)/2 - (b-a)/2$.

Returning to functions with vanishing 2-variation, the next step is to integrate more general functions of X along X. Not surprisingly, there is a naive fundamental theorem of calculus and change of variable formula. To summarize:

Integrability conditions The integral $\int_a^b X_t\,dX_t$ exists in the Young sense if and only if X has vanishing 2-variation on $[a,b]$ and then $\int_a^b X_t\,dX_t = (X_b^2 - X_a^2)/2$.

Fundamental theorem of calculus If X has vanishing 2-variation on $[a,b]$, f is a real function defined on the range of X with Lipschitz first derivative (in particular if f is C^2 on the range of X) then $f(X)$ and $f'(X)$ have vanishing 2-variation on $[a,b]$, $f'(X)$ is Young-integrable along X and

$$\int_a^b f'(X_u)dX_u = f(X_b) - f(X_a).$$

Change of variable formula If X has vanishing 2-variation on $[a,b]$, f is a real function defined on the range of X with Lipschitz first derivative and g is a Lipschitz function on the range of $Y := f(X)$ (in particular if g is C^1 on the range of X) then Y, $f'(X)$, $g(Y)$ and $f'(X)g(f(X))$ have vanishing 2-variation on $[a,b]$ and

$$\int_a^b g(Y_u)dY_u = \int_a^b g(f(X_u))f'(X_u)dX_u.$$

[1] Itō's stochastic integral relies on mean square convergence of retarded sums.

The proofs are not difficult. They are left here as exercises, but can be found in Appendix D

This simple theory is nice, but grossly insufficient applications: it is restricted to the one dimensional setting (X is a path in $\mathbb{R}$). In the multidimensional setting, there are no simple necessary and sufficient conditions for Young integrability, but there are useful classes of paths and functions leading to Young integrable pairs. We turn to this in the next section.

5.3 Young integrability via Hölder-type conditions

Recall that a function $X : [a, b] \to \mathbb{R}$ is α-Hölder if there is a constant K such that, $\forall s, t \in [a, b]$, $|X_t - X_s| \leq K|t - s|^\alpha$. Equivalently, X is α-Hölder if

$$|X|_{\alpha\text{-Hölder},[a,b]} := \sup_{s,t\in[a,b],s\neq y} \frac{|X_t - X_s|}{|t - s|^\alpha} < +\infty.$$

.

We have seen in Chapter 4 that $\int(\delta\Gamma)$, if it exists, is reparameterization invariant. The condition of being α-Hölder is not, so Hölder spaces are not the most natural places to look Young integrable pairs. However, this is a simple and convenient setting. Moreover, and though we shall not elaborate on that construction, there is a nice trick to transport results obtained via Hölder spaces to reparameterization invariant spaces.

The main results are the following:

Young integrability via Hölder-type conditions Let X and Y be two real paths defined on $[a, b]$. Assume that X is α-Hölder and Y is β-Hölder on $[a, b]$, with $\alpha + \beta > 1$.

Then the Young integral $\int_a^b Y_u dX_u$ exists (as the limit of the Riemann-Young sums $A_\Delta(X, Y)$ when mesh(Δ) goes to 0). The integral is bilinear in (X, Y).

Young-Loëve estimate

$$\left|\int_a^b Y_u dX_u - Y_a(X_b - X_a)\right| \leq \frac{1}{2^{\alpha+\beta} - 2}|X|_{\alpha\text{-Hölder},[a,b]}|Y|_{\beta\text{-Hölder},[a,b]}(b - a)^{\alpha+\beta}$$

holds.

Regularity The function $t \in [a, b] \mapsto \int_a^t Y_u dX_u$ is α-Hölder.

Integrability is an elementary consequence of the general discussion in Chapter 4. Indeed, a simple computation shows that $\Gamma\colon [a, b]^2 \to \mathbb{R},\ (s, t) \mapsto Y_s(X_t - X_s)$ (which vanishes automatically on the diagonal) is triangular with exponent $\gamma := \alpha + \beta$ because such that $\Gamma_{s,t} := Y_s(X_t - X_s)$ is such that

$$\Gamma_{s,t} + \Gamma_{t,u} - \Gamma_{s,u} = Y_s(X_t - X_s) + Y_t(X_u - X_t) - Y_s(X_u - X_s) = (Y_t - Y_s)(X_u - X_t).$$ [2]

[2] A kind of primitive Yang-Baxter relation!

The constant in the Young-Loëve estimate is obtained by explicit optimization, and the regularity then follows from the regularity of X and the triangular inequality $|\int_s^t Y_u dX_u| \leq |Y_s(X_t - X_s)| + |\int_s^t Y_u dX_u - Y_s(X_t - X_s)|$, see p. 67.

The **Regularity** property has an immediate usefulness for controlled differential equations. Recall our initial aim, i.e. solving

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

or its integrated version.

$$Y_t = y_a + \int_a^t V(Y_s)\, dX_s.$$

The meaning of such an integral equation is not yet clear, but we explain how Picard iteration can be defined.

Picard iteration Suppose that X is α-Hölder on $[a, b]$ and f is γ-Hölder on $\mathbb{R}$. If $(\gamma+1)\alpha > 1$ then the Picard iteration sequence for the controlled differential equation $Y_t = a + \int_a^t f(Y_s)\, dX_s$, namely

$$Y_t^{(0)} := y_a \text{ for } t \in [a, b] \qquad Y_t^{(n+1)} := y_a + \int_0^t V(Y_s^{(n)})\, dX_s \text{ for } t \in [0, T] \text{ and } n \in \mathbb{N},$$

where all integrals are in the Young sense, is well-defined.

That $(\gamma + 1)\alpha > 1$ is *the* natural condition is easy to understand: as we have seen, by **Regularity** $\int_0^t f(Y_s)\, dX_s$, if defined, is expected to have the same regularity in t as X does, so that if Y solves $Y_t =_a + \int_0^t V(Y_s)\, dX_s$ then Y must be α-Hölder, and then $f(Y)$ is $\beta := \gamma\alpha$-Hölder and then the integral $\int f(Y_s)\, dX$ is well-defined if $\alpha + \beta > 1$. Of course, the conditions we have given for the Young integral to exist are only sufficient conditions, so this remark is only heuristic. But it is the essence of the proof of the lemma.

The detailed argument is by recursion. The zeroth approximation $Y^{(0)}$ is clearly α-Hölder on $[a, b]$. Let $n \in \mathbb{N}$. If $Y^{(n)}$ is α-Hölder on $[a, b]$ then $f(Y^{(n)})$ is $\gamma\alpha$-Hölder on $[a, b]$, so that, as $(\gamma + 1)\alpha > 1$, the integral $\int_0^t f(Y_s^{(n)})\, dX_s$ is well defined as a Young integral for $t \in [a, b]$, an then $Y_t^{(n+1)} := y + \int_0^t f(Y_s^{(n)})\, dX_s$ is α-Hölder on $[a, b]$ by **Regularity**, closing the recursion step.

The question of the convergence of the sequence of Picard approximations $Y^{(n)}$ towards a solution of the controlled differential equation $Y_t = y + \int_0^t f(Y_s)\, dX_s$ is more delicate than, but in some sense paralel to, its classical counterpart for differential equation. Naive counting for this simple case is that the integrator $X_s = s$ is 1-Hölder ($\alpha = 1$) and the condition put on f to ensure convergence of the Picard Scheme is that f should be 1-Hölder as well ($\gamma = 1$), so the classical version gives uniqueness for $(\gamma + 1)\alpha = 2$, a condition much more stringent than the mere continuity of f which is enough to have a well-defined Picard iteration sequence, enough to have the existence of solutions (by the Peano scheme of approximations) but insufficient in general to have convergence of the Picard scheme and uniqueness of the solution. For the case of a controlled differential equation, uniqueness of the solution and convergence of the Picard scheme hold if f is differentiable and f' (not f!) is γ-Hölder on $\mathbb{R}$ when $(\gamma+1)\alpha > 1$. With the additional hypothesis that f, f' are bounded, the proof, that we shall not even sketch, relies on a careful use of the local inequality

$$\frac{\left|\int_s^t f(Y_u)\, dX_u\right|}{|t-s|^\alpha} \leq |X|_{\alpha\text{-Hölder},[a,b]} \left(\|f\|_\infty + \frac{1}{2^{\alpha+\beta}-2} \|f'\|_\infty |Y|_{\beta\text{-Hölder},[a,b]} |t-s|^\beta \right).$$

which is a direct consequence of the proof of **Regularity**, see p. 69.

Appendix

5.A The Young integral: proofs

5.A.1 Basics

Before going to the heart of the results announced in the main text, we quote a useful set of identities. Recall that if $\Delta : a = t_0 \leq t_1 \cdots \leq t_{2n} = b$ is a tagged subdivision of $[a, b]$ the $\Delta_* : a = t_0 < t_2 \cdots \leq t_{2n} = b$ is a subdivision of $[ab]$. The set $\{t_0, t_1 \cdots, t_{2n}\}$ defines a subdivision of $[a, b]$ denoted by Δ_{**}: the interpolation points of Δ, if not already present in Δ_*, are promoted to subdivision points in Δ_{**}. (see Appendix A.1 for details).

Comparison Let X, Y be real functions on the interval $[a, b]$.

i) For $t \in [a, b]$

$$Y_a(X_t - X_a) + Y_t(X_b - X_t) - Y_a(X_b - X_a) = (Y_t - Y_a)(X_b - X_t).$$

ii) Let $\Delta : a = t_0 \leq t_1 \cdots \leq t_{2n} = b$ be a tagged subdivision of $[a, b]$. Then

$$A_\Delta(X, Y) - A^{\text{ret}}_{\Delta_{**}}(X, Y) = \sum_{m=0}^{n-1} (Y_{t_{2m+1}} - Y_{t_{2m}})(X_{t_{2m+1}} - X_{t_{2m}}) = A_\Delta(Y, X) - A^{\text{ret}}_{\Delta_{**}}(Y, X)$$

and

$$A^{\text{adv}}_{\Delta_{**}}(X, Y) - A_\Delta(X, Y) = \sum_{m=1}^{n} (Y_{t_{2m}} - Y_{t_{2m-1}})(X_{t_{2m}} - X_{t_{2m-1}}) = A^{\text{adv}}_{\Delta_{**}}(Y, X) - A_\Delta(Y, X).$$

[2] This restriction is only to comply with our definition of subdivisions as made of distinct points. But if we extend the definition to allow clusters of coinciding points, the corresponding Riemann-Young sums give the same result as the Riemann-Young sums when only one point in each cluster is kept, because $X_{t''}(X_{t'''} - X_{t'})$ vanishes if $t' = t'''$.

Proof. Point i) really requires no explanation (but is surprisingly powerful). It is the one used to show that $\Gamma_{s,t} := Y_s(X_t - X_s)$ is triangular with exponent $\alpha + \beta$ is X is α-Hölder and X is β-Hölder.

For point ii) we rewrite

$$A_\Delta(X,Y) = \sum_{m=0}^{n-1} Y_{t_{2m+1}}(X_{t_{2m+2}} - X_{t_{2m}}) = \sum_{m=0}^{n-1} Y_{t_{2m+1}}(X_{t_{2m+1}} - X_{t_{2m}}) + \sum_{m=0}^{n-1} Y_{t_{2m+1}}(X_{t_{2m+2}} - X_{t_{2m+1}})$$

and

$$A^{\mathrm{ret}}_{\Delta_{**}}(X,Y) = \sum_{m=0}^{2n-1} Y_{t_m}(X_{t_{m+1}} - X_{t_m}) = \sum_{m=0}^{n-1} Y_{t_{2m}}(X_{t_{2m+1}} - X_{t_{2m}}) + \sum_{m=0}^{n-1} Y_{t_{2m+1}}(X_{t_{2m+2}} - X_{t_{2m+1}})$$

The second sums in each expression cancel each other in $A_\Delta(X,Y) - A^{\mathrm{ret}}_{\Delta_{**}}(X,Y)$ to yield

$$A_\Delta(X,Y) - A^{\mathrm{ret}}_{\Delta_{**}}(X,Y) = \sum_{m=0}^{n-1} (Y_{t_{2m+1}} - Y_{t_{2m}})(X_{t_{2m+1}} - X_{t_{2m}}),$$

an expression which is symmetric in (X, Y). The proof of the comparison formula with advanced sums is analogous. □

Integration by parts Let X, Y be real functions on the interval $[a, b]$. If Y is Young-integrable along the path X, then X is Young-integrable along the path Y and $\int_a^b Y_t\, dX_t + \int_a^b X_t\, dY_t = X_b Y_b - X_a Y_a$.

Proof. The essence of the proof is combinatorial.

First, if $\Delta : a = t_0 < t_1 < \cdots < t_n = b$ is any subdivision of $[a, b]$ then

$$A^{\mathrm{ret}}_\Delta(X,Y) + A^{\mathrm{adv}}_\Delta(Y,X) = A^{\mathrm{adv}}_\Delta(X,Y) + A^{\mathrm{ret}}_\Delta(Y,X) = X_b Y_b - X_a Y_a$$

by a telescopic sum argument.

Second, let $\Delta : a = t_0 \le t_1 \cdots \le t_{2n} = b$ be a tagged subdivision of $[a, b]$. By symmetry (which follows from **Comparison**)

$$\begin{aligned} A_\Delta(Y,X) - X_b Y_b + X_a Y_a &= (A_\Delta(Y,X) - A^{\mathrm{ret}}_{\Delta_{**}}(Y,X)) + A^{\mathrm{ret}}_{\Delta_{**}}(Y,X) - X_b Y_b + X_a Y_a \\ &= (A_\Delta(X,Y) - A^{\mathrm{ret}}_{\Delta_{**}}(X,Y)) - A^{\mathrm{adv}}_{\Delta_{**}}(X,Y) \end{aligned}$$

Suppose now that Y is Young-integrable along the path X. Let $\varepsilon > 0$. As $\mathrm{mesh}(\Delta_{**}) \le \mathrm{mesh}(\Delta)$ there is $\delta > 0$ such that $\left|A_\Delta(X,Y) - \int_a^b Y_t\, dX_t\right|$, $\left|A^{\mathrm{ret}}_{\Delta_{**}}(X,Y) - \int_a^b Y_t\, dX_t\right|$ and $\left|A^{\mathrm{adv}}_{\Delta_{**}}(X,Y) - \int_a^b Y_t\, dX_t\right|$ are all $\le \varepsilon/3$ whenever $\mathrm{mesh}(\Delta) \le \delta$, so that

$$\left|A_\Delta(Y,X) + \int_a^b Y_t\, dX_t - X_b Y_b + X_a Y_a\right| = \left|A_\Delta(X,Y) - A^{\mathrm{ret}}_{\Delta_{**}}(X,Y) - A^{\mathrm{adv}}_{\Delta_{**}}(Y,X) + \int_a^b Y_t\, dX_t\right| \le \varepsilon$$

whenever $\mathrm{mesh}(\Delta) \le \delta$, leading to the announced result: X is Young-integrable along the path Y and $\int_a^b X_t\, dY_t = X_b Y_b - X_a Y_a - \int_a^b Y_t\, dX_t$. □

5.A.2 Integrals of the type $\int_a^b f(X)dX$

Recall the definition of 2-variation. Let $X\colon [a,b] \to \mathbb{R}$ be a function.

2-variation Let $\Delta : a = t_0 < t_1 < \cdots < t_n = b$ be a subdivision of $[a,b]$. The 2-variation of X along Δ is $Q_\Delta(X) := \sum_{m=0}^{n-1}(X_{t_{m+1}} - X_{t_m})^2$

Functions with vanishing 2-variation The function X has vanishing 2-variation on $[a,b]$ if $\lim_{\text{mesh}(\Delta)\downarrow 0} Q_\Delta(X) = 0$.

This definition generalizes in an obvious way to $p \geq 1$ and $X\colon [a,b] \to E$ where (E,d) is a metric space, replacing $(X_{t_{m+1}} - X_{t_m})^2$ by $d(X_{t_{m+1}}, X_{t_m})^p$in the definition, leading to the concept of vanishing p-variation.

Exercise 5.1. Show that a function $X\colon [a,b] \to E$ with vanishing p-variation is continuous.

Composition with Lipschitz functions Let X have vanishing 2-variation on $[a,b]$ and let f be a real Lipshitz function defined on the range of X (this holds in particular if f has bounded derivative on the range[3]of X). Then $f \circ X$ has vanishing 2-variation on $[a,b]$.

Proof. Let k be the Lipschitz modulus of f (if f is differentiable $k := \sup_{x\in X([a,b])} |f'(x)|$, the upper bound for the absolute value of the derivative of f on the range of X). Then $|f(X_t) - f(X_s)| \leq k|X_t - X_s|$ for $s,t \in [a,b]$ so that, for any subdivision Δ of $[a,b]$, $Q_\Delta(f \circ X) \leq k Q_\Delta(X)$. □

Reparameterization invariance Let φ be an increasing homeomorphism from $[c,d]$ to $[a,b]$. Then the pullback map $\varphi^*\colon X \mapsto X \circ \varphi$ defines a one-to-one map from functions with vanishing 2-variation on $[a,b]$ to functions with vanishing 2-variation on $[c,d]$.

Proof. The proof is similar to the proof of the reparameterization invariance of Γ-Riemann sums and their limit if any, see p. 50.Borrowing the same notation, if Δ is a subdivision of $[c,d]$ and $\varphi(\Delta)$ the image subdivision of $[a,b]$ (a one-to-one correspondence between subdivisions of $[c,d]$ and of $[a,b]$), then, for X a path on $[a,b]$, $Q_{\varphi(\Delta)}(X) = Q_\Delta(X \circ \varphi)$, and by the uniform continuity of φ on the compact interval $[c,d]$, the small mesh limit for Δ and for $\varphi(\Delta)$ are the same. □

Let $X\colon [a,b] \to \mathbb{R}$ be a function. The proof that the Young integral $\int_a^b X\, dX$ if and only if X has vanishing 2-variation on $[a,b]$ was given in the main text, and we make this assumption in what follows. We prove the results announced in the main text.

[3]A closed interval by continuity.

Fundamental theorem of calculus If X has vanishing 2-variation on $[a, b]$, f is a real function defined on the range of X with Lipschitz first derivative (in particular if f is C^2 on the range of X) then $f(X)$ and $f'(X)$ have vanishing 2-variation on $[a, b]$, $f'(X)$ is Young-integrable along X and

$$\int_a^b f'(X_u)dX_u = f(X_b) - f(X_a).$$

Proof. As f' is Lipschitz on the range of X (a compact interval by continuity), f' is bounded on the range of X as well, so that f is C^1 with bounded derivative. By **Composition**, $f(X)$ and $f'(X)$ have vanishing 2-variation. Let k be the Lipschitz modulus of f'. Then for u, v in the interval of definition of f, $f(v)-f(u)-f'(u)(v-u) = \int_u^v (f'(w) - f'(u))\, dw$ (a Riemann integral!) so $|f(v) - f(u) - f'(u)(v - u)| \leq \int_u^v k|w-u|\, dw = \frac{k}{2}(v-u)^2$ leading to $|f(X_t)-f(X_s)-f'(X_s)(X_t-X_s)| \leq \frac{k}{2}(X_t-X_s)^2$. Take a subdivision $\Delta : a = t_0 < t_1 \cdots < t_n = b$ of $[a, b]$, and write

$$f(X_b) - f(X_a) - A^{\text{ret}}_\Delta(X, f'(X)) = \sum_{m=0}^{n-1} f(X_{t_{m+1}}) - f(X_{t_m}) - f'(X_{t_m})(X_{t_{m+1}} - X_{t_m}).$$

Apply the previous inequality to each part of Δ to get

$$|f(X_b) - f(X_a) - A^{\text{ret}}_\Delta(X, f'(X))| \leq \frac{k}{2}|X|^2_{2\text{-var},\Delta}.$$

Thus $\lim_{\text{mesh}(\Delta)\downarrow 0} A^{\text{ret}}_\Delta(X, f'(X))$ exists and its value is $f(X_b) - f(X_a)$. Finally, if $\Delta : a = t_0 < t_1 \cdots < t_{2n} = b$ is a tagged subdivision of $[a, b]$, by **Comparison** p. 63:

$$A_\Delta(X, f'(X)) - A^{\text{ret}}_{\Delta_{**}}(X, f'(X)) = \sum_{m=0}^{n-1}(f'(X_{t_{2m+1}}) - f'(X_{t_{2m}}))(X_{t_{2m+1}} - X_{t_{2m}}),$$

so

$$|A_\Delta(X, f'(X)) - A^{\text{ret}}_{\Delta_{**}}(X, f'(X))| \leq \sum_{m=0}^{n-1} k(X_{t_{2m+1}} - X_{t_{2m}})^2 \leq k|X|^2_{2\text{-var},\Delta_{**}}.$$

Thus general Riemann-Young sums are close to retarded Riemann Young sums and approach $f(X_b) - f(X_a)$ at small mesh. Hence $f'(X)$ is Young-integrable along X and

$$\int_a^b f'(X_u)dX_u = f(X_b) - f(X_a).$$

$\square$

This leads to the change of variable formula.

Change of variable for the Young integral Let $X\colon [a,b] \to \mathbb{R}$ have vanishing 2-variation on $[ab]$. Let f be a real function defined on the range of X with Lipschitz first derivative. Let Y be the path $Y := f(X)$ on $[a,b]$. Let g be a Lipschitz function on the range of Y. Then $Y = f(X), f'(X), g(Y)$ and $f'(X)g(f(X))$ have vanishing 2-variation on $[ab]$ and

$$\int_a^b g(Y_u)dY_u = \int_a^b g(f(X_u))f'(X_u)dX_u.$$

Proof. This follows via the standard pattern from the fundamental theorem of calculus. The function g, being Lipschitz, is Riemann integrable on the range of Y. Let G be a primitive of g on that range. The function $Y = f(X)$ has vanishing 2-variation on $[ab]$ by the **Fundamental theorem of calculus**. The function G has Lipschitz first derivative. Thus we may apply the **Fundamental theorem of calculus** to G and Y in place of f and X to get

$$\int_a^b g(Y_u)dY_u = \int_a^b G'(Y_u)dY_u = G(Y_b) - G(Y_a).$$

The composition (when defined) of Lipschitz functions is Lipschitz. The product of bounded Lipschitz functions is bounded Lipschitz. Hence, $(g \circ f)$ and $(g \circ f)f'$ are Lipschitz. Thus $g(f(X))$ and $g(f(X))f'(X)$ are in $\mathcal{C}^{0,2\text{-var}}([a,b],\mathbb{R})$ by **Composition**, p. 65. We infer that $\int_a^b g(f(X_u))f'(X_u)dX_u$ is defined. As $(g\circ f)f' = (G\circ f)'$ the function $G \circ f$ has Lipschitz first derivative and another use of the **Fundamental theorem of calculus** yields

$$\int_a^b g(f(X_u))f'(X_u)dX_u = \int_a^b (G\circ f)'(X_u)dX_u = (G\circ f)(X_b) - (G\circ f)(X_a) = G(Y_b) - G(Y_a),$$

establishing the desired equality. □

Though this integration theory is quite satisfactory from a formal viewpoint, it is insufficient for applications. The Hölder theory via the triangular lemma as summarized in p. 60 has a much broader scope.

5.A.3 Complements

We give a detailed proof of two properties of the Young integral announced on p. 60: an improved constant for the Young-Löeve estimate and the fact that he Young integral as a function of the upper bound has the regularity of the integrator.

We start with

Improved triangular lemma Let X and Y be two real paths defined on $[a,b]$. Assume that X is α-Hölder and Y is β-Hölder on $[a,b]$ and let $\Gamma\colon [a,b]^2 \to \mathbb{R}, (s,t) \to Y_s(X_t - X_s)$. Set $\gamma := \alpha + \beta$. Then Γ is triangular with exponent γ and

$$\sup_{s,t,u\in[a,b],\, |s,t,u|>0} \frac{|\Gamma_{s,u} - \Gamma_{s,t} - \Gamma_{t,u}|}{|s,t,u|^\gamma} \le |X|_{\alpha\text{-Hölder},[a,b]} |Y|_{\beta\text{-Hölder},[a,b]}.$$

Moreover

$$\sup_{s,t,u\in[a,b],\, s\le t\le u, s<u} \frac{|\Gamma_{s,u}-\Gamma_{s,t}-\Gamma_{t,u}|}{|s-u|^{\gamma}} \le \frac{\alpha^{\alpha}\beta^{\beta}}{(\alpha+\beta)^{\alpha+\beta}}|X|_{\alpha\text{-Hölder},[a,b]}|Y|_{\beta\text{-Hölder},[a,b]},$$

and

$$\sup_{s,u\in[a,b],\, s<u} \frac{|\Gamma_{s,u}-\Gamma_{s,(s+u)/2}-\Gamma_{(s+u)/2,u}|}{|s-u|^{\gamma}} \le \frac{1}{2^{\gamma}}|X|_{\alpha\text{-Hölder},[a,b]}|Y|_{\beta\text{-Hölder},[a,b]}.$$

Proof. Observe that

$$\Gamma_{s,t}+\Gamma_{t,u}-\Gamma_{s,u} = Y_s(X_t-X_s)+Y_t(X_u-X_t)-Y_s(X_u-X_s) = (Y_t-Y_s)(X_u-X_t)$$

(a nice rearrangement) and

$$|(Y_t-Y_s)(X_u-X_t)| \le |X|_{\alpha\text{-Hölder},[a,b]}|Y|_{\beta\text{-Hölder},[a,b]}|t-s|^{\beta}|u-t|^{\alpha}$$

by the Hölder properties of X and Y. It is immediate that $|t-s|^{\beta}|u-t|^{\alpha} \le |s,t,u|^{\gamma}$ which yields the first inequality and is enough to get that Γ is triangular with exponent γ. Doing some optimization when the ordering is $s \le t \le u$ improves the bound and yields

$$\sup_{t,s\le t\le u} (t-s)^{\beta}(u-t)^{\alpha} \le \frac{\alpha^{\alpha}\beta^{\beta}}{(\alpha+\beta)^{\alpha+\beta}}(u-s)^{\gamma}$$

(the extremizing t is $\frac{\alpha s+\beta u}{\alpha+\beta}$), yielding the second inequality. The last inequality is clear.[4] □

The last point yields as a corollary

Young-Löeve constant

$$\left|\int_a^b Y_u dX_u - Y_a(X_b-X_a)\right| \le \frac{1}{2^{\alpha+\beta}-2}|X|_{\alpha\text{-Hölder},[a,b]}|Y|_{\beta\text{-Hölder},[a,b]}(b-a)^{\alpha+\beta}$$

Proof. As explained in the general theory of the sewing lemma (see p. 55) we know that

$$\|\int_a^b (\delta\Gamma) - \Gamma_{a,b}\|_E \le \frac{\tilde{K}}{1-2^{1-\gamma}}(b-a)^{\gamma},$$

where $\tilde{K} = \tilde{K}_{[a,b]}(\Gamma) := \sup_{s,u\in[a,b],\, s<u} \frac{\|\Gamma_{s,u}-\Gamma_{s,(s+u)/2}-\Gamma_{(s+u)/2,u}+\Gamma_{(s+u)/2,(s+u)/2}\|_E}{|s-u|^{\gamma}}$ for a general Γ. The above computation shows that for $\Gamma_{s,t} = Y_s(X_t-X_s)$ we have $\tilde{K} \le \frac{1}{2^{\gamma}}|X|_{\alpha\text{-Hölder},[a,b]}|Y|_{\beta\text{-Hölder},[a,b]}$ which leads to the announced bound. □

[4] From the definition via an extremization problem, $\frac{\alpha^{\alpha}\beta^{\beta}}{(\alpha+\beta)^{\alpha+\beta}} \ge 1/2^{\alpha+\beta}$, an inequality which also expresses that fair coin tossing has higher entropy than biased coin tossing.

Regularity Let X and Y be two real paths defined on $[a, b]$. Assume that X is α-Hölder and Y is β-Hölder on $[a, b]$, with $\alpha + \beta > 1$. The Young integral as a function of the upper bound is α-Hölder, i.e.

$Z\colon [a, b] \to \mathbb{R},\ t \mapsto Z_t := \int_a^t Y_u dX_u$ is α-Hölder.

Proof. We need to show that

$$\sup_{a \leq s < t \leq b} \frac{|Z_t - Z_s|}{|t-s|^\alpha} < +\infty.$$

This is an easy consequence of the Young-Loëve estimate. Using it for $Z_t - Z_s = \int_s^t Y_u dX_u$ on $[s, t]$ yields

$$|Z_t - Z_s| \leq |Y_s(X_t - X_s)| + \frac{1}{2^{\alpha+\beta} - 2} |X|_{\alpha\text{-Hölder},[s,t]} |Y|_{\beta\text{-Hölder},[s,t]} |t-s|^{\alpha+\beta}$$

Thus

$$\frac{|Z_t - Z_s|}{|t-s|^\alpha} \leq |Y_s| \frac{|X_t - X_s|}{|t-s|^\alpha} + \frac{1}{2^{\alpha+\beta} - 2} |X|_{\alpha\text{-Hölder},[s,t]} |Y|_{\beta\text{-Hölder},[s,t]} |t-s|^{\beta}.$$

But $|X|_{\alpha\text{-Hölder},[s,t]} \leq |X|_{\alpha\text{-Hölder},[a,b]}$, $|Y|_{\beta\text{-Hölder},[s,t]} \leq |Y|_{\beta\text{-Hölder},[a,b]}$ so we infer a crude bound

$$|Z|_{\alpha\text{-Hölder},[a,b]} \leq |X|_{\alpha\text{-Hölder},[a,b]} \left(\|Y\|_\infty + \frac{1}{2^{\alpha+\beta} - 2} |Y|_{\beta\text{-Hölder},[a,b]} |b-a|^\beta \right) < +\infty.$$

$\square$