Stochastic Processes

Volker Betz

Motivation and contents

Many laws of nature are encoded by differential equations. For example, the position $x(t) \in \mathbb{R}$ of a moving particle as a function of time can be described by

$$\partial_t^2 x(t) = -a(x(t))\partial_t x(t) + F(t, x(t)), \tag{0.1}$$

where $a(y)$ is the friction coefficient at the location $y \in \mathbb{R}$, and $F(y)$ is the external force at position $y \in \mathbb{R}$. Other laws of nature (like e.g. fluid dynamics or electrodynamics) are encoded by partial differential equations.

In many cases, it is useful to model the external force F as a random quantity. For example, the particle might be travelling through a gas of other particles and get frequent pushes from these other particles. Of course, one could try to model all of the particles at the same time, but there are typically many millions of them, and this would give millions of coupled ordinary differential equations, impossible to solve in practice. So in this cases, it is better to focus on the one particle that we are interested in, and model the pushes by other particles by a random external force. There are two questions:

(1) What type of random force can we use? Is it suitable for the given physical situation? Cn we prove that the randomly forced system somehow approximates the relevant aspects (e.g. the movement of the single particle) of the full, non-random system?
(2) How can we mathematically describe the most relevant random forces?
(3) Given a certain random force, can we actually solve the resulting differential equation? This means: does a solution exist? Is it unique? Can we calculate it in good cases, or at least find properties of it?

Point 1) is usually quite difficult and either part of physics, or sometimes can be solved via statistical mechanics. We will not discuss it in this lecture. Instead, we will start with item 2) and present a very important case of random force (called „noise“ in the following), leading us to the study of Brownian motion. Brownian motion is easily the most important, and also among the most beautiful objects of probability theory, and we will study many aspects of it. Then, we will turn to item 3) and talk about stochastic differential equations. Here we will take a modern point of view, and use the theory of rough paths that has been established since the early 2000’s.

We will be content with a simpler version of (0.1), namely the first order differential equation

$$\partial_t X_t = b(t, X_t) + \sigma(t, X_t) \cdot \xi_t, \tag{0.2}$$

where $X_t \in \mathbb{R}$ is the quantity of interest (e.g. our particle), $b \in \mathbb{R}$ is a deterministic force that may depend on the time t and on the current state X_t of the system, $\xi_t \in \mathbb{R}^d$ is the random noise (more about this just below), and $\sigma(t, x)$ is another factor that modulates how strongly this noise influences the dynamics of the particle when the particle is at position x at time t.

So how should the random function $t \mapsto \xi_t$ look like? If we think of the particle getting random and very chaotic pushes from other particles, we want to make $t \mapsto \xi_t$ „as random as possible“. An interesting idea for this is the following: let us restrict our attention to $0 \leqslant t \leqslant 1$. Then a large class of functions on this interval can be represented by a Fourier series:

$$\xi_t := \sum_{k=0}^{\infty} Y_k \cos(2\pi k t) + \sum_{k=1}^{\infty} Z_k \sin(2\pi k t),$$

with coefficients Y_k, Z_k.

If we already know the function ξ that we want to approximate, then Fourier theory tells us that we can compute the coefficients by the usual Fourier inversion formula, for example $Z_k = 2 \int_0^1 \xi_t \sin(2\pi k t) \, \mathrm{d}t$ for $k \geqslant 1$. But since we *do not* know the function ξ but instead want to produce a random function, it is reasonable to just use random Fourier coefficients. Our aim to be „as random as possible“ is achieved by choosing $Y_k, Z_k \sim \mathcal{N}(0, 1)$ and iid, i.e. standard normally distributed and independent. A more physical justification for this choice is that the energy spectrum is iid, i.e. if ξ_t were a wave composed of elementary waves (the sin and cos functions), then all elementary waves occur with the same strength on average, and do not influence each other.

The problem with this definition is that the limit $N \to \infty$ of the functions

$$\xi_t^{(N)} := \sum_{k=1}^{N} Y_k \cos(2\pi k t) + \sum_{k=1}^{N} Z_k \sin(2\pi k t),$$

with Y_k, Z_k iid $\mathcal{N}(0, 1)$, does not exist for any $t \in [0, 1]$ almost surely. This can be seen easily for $t = 0$, because $\xi_t^{(N)}(0) = \sum_{k=0}^{N} Y_k$, and the variance of this expression is equal to $N - 1/2$. It is also true for all other t. Nevertheless, the limit $\xi_t = \lim_{N\to\infty} \xi_t^{(N)}$ is very important and is called *white noise*. This might sound strange, but what happens is that while the limit does not exist as a random function, it *does* exist as a random distribution. We will however not follow this route, but rather use a trick to improve convergence, namely we integrate. This means that we consider

$$B_t^{(N)} := \int_0^t \xi_s^{(N)} \, \mathrm{d}s = tY_0 + \sum_{k=1}^{N} \frac{1}{2\pi k} \big(Y_k \sin(2\pi k t) + Z_k (\cos(2\pi k t) - 1) \big).$$

You will soon prove as an exercise that this limit does exist as a random function. An easy special case is the case $t = 1$, when $B_1^{(N)} = Y_0$. The random function $t \mapsto B_t = lim_{N\to\infty} B_t^{(N)}$ is *Brownian motion* (in our case: on the unit interval), which will be the main topic for the first half of the lecture. We will give a quite different definition of it in the next chapter.

Coming back to the equation (0.2), we see that it still has the problematic quantity ξ_t in it, instead of the (allegedly) unproblematic B_t. At least of $\sigma(t,x) = \sigma_0 =$ constant, we can integrate this equation on both sides, and get

$$X_t - X_0 = \int_0^t \partial_s X_s \, \mathrm{d}s = \int_0^t b(s, X_s) \, \mathrm{d}s + \sigma_0 B_t.$$

This equation now makes sense provided we can prove that the function $t \mapsto B_t$ makes sense. A different question is how to find a solution to it, this is not clear at the moment. In the second half of the lecture we will learn how to find solutions to this equation, and also how to deal with non-constant $\sigma(t,x)$. We will introduce the famous Itô-integral, but also the more modern theory of rough paths to deal with these problems.

Some literature

There are many books about Brownian motion and stoachstic differential equations, and not very many about rough paths. Here I give a few books that I have read (at least in part) and used for this lecture, along with some comments on them.

Books on Brownian Motion and SDE

[SP] R. Schilling and L. Partzsch: Brownian Motion: An Introduction to Stochastic Processes. Berlin, Boston: De Gruyter, 2012. https://doi.org/10.1515/9783110278989.

The strength of this book is that it does all the proofs very carefully and in great detail. The price to pay for this is that in some places, things appear more complicated than necessary. Overall a great reference.

[MP] P. Mörters and Y. Peres: Brownian Motion. Cambridge University Press, 2016.

A very beautiful book concentrating on the geometric and sample path properties of Brownian motion, less on SDE. It is written in „american style“ which means it is engaging and fun to read, but can skip over some details that may be a bit hard to fill in in some places. In this respect it is quite the opposite of the book of Schilling and Partzsch.

[Li] T. Liggett: Continuous Time Markov Processes: An Introduction. AMS publishing, 2010.

Another fantastic book. It has the best treatment of the theory of Markov processes that I am aware of, has a very nice chapter on Brownian motion and several other things that are not done in this lecture. Very clear and careful presentation, great choice of notation and terminology.

[RY] D. Revuz and M. Yor: Continuous martingales and Brownian Motion. Springer 1999

Still „the bible“ on the topic of Brownian motion and SDE. Contains everything you want to know about continuous martingales and Brownian motion, and probably much more. Written in a concise, exact style, that can be a bit demanding, but very thorough and complete.

Book on Rough paths

[FH] P. Friz and M. Hairer: a course on rough paths, Springer Universitext, 2014

The only textbook on the subject that I am familiar with. Its advantage is that it treats the topics in a very concise way and does not try to achieve maximal generality. The proofs,

however, are often so concise that I find it very hard to follow them. Several proofs in these lecture notes are mine, in part because they were so short in the book that I had to fill 90 percent of the steps, in part because I simply did not understand them and had to do my own ones. So please check them very carefully! If you work hard enough, the book can be very inspiring, but it can also be a bit frustrating at times. Of note, there is a chapter on regularity structures (which I do not cover). The only other book I am aware of is a monograph by Friz and Victoir, which is a few years older, seems very thorough, and has almost 600 pages.

Other sources

[Kl] A. Klenke: Probability Theory, Springer Universitext, 2014

Typical german style text book: does not try very hard to entertain you, but does a very good job in laying out and explaining the basic building blocks in probability, and it does have some very nice examples. Very solid reference volume to have around.

[Ka] O. Kallenberg: Foundations of modern probability, Third edition, 2021

In my opinion, the best reference on probability theory in general, although not completely easy to read. The notation may be a bit hard to follow at first, because it is rather minimalistic, but the proofs are highly elegant (and often very short). The whole book has a very „pure mathematics" feeling to it. Highly recommended.

[Be] V. Betz: Probability theory lecture notes, 2020; available on the moodle page.

I refer to these notes in several places, mostly because I know them rather well. I do like them, but anything els would be quite sad I suppose. Whether you like them too is up to you - but do try some of the other literature as well in any case!

1. Brownian Motion

(1.1) Definition

Let $(E, \mathcal{E})$ be a measurable space, T a set. A collection (which is another word for „a set") of $(E, \mathcal{E})$-valued random variables $\boldsymbol{X} = (X_t)_{t \in T}$ is called **E-valued stochastic process with index set** T. For each $\omega \in \Omega$, the function $T \to E, t \mapsto X_t(\omega)$ is called the **sample path** (or simply: path) of the stochastic process $\boldsymbol{X}$ belonging to $\omega \in \Omega$ if T is a subset of $\mathbb{R}$, otherwise it is sometimes called the **configuration** corresponding to ω.

(1.2) Examples

The definition of stochastic process is very broad - for example, for any random variable one can take $T = \{1\}$ and $X_1 = Y$, so every random variable is a (trivial) stochastic process. Like some other definitions in probability theory, the value of the concept of a stochastic process lies less in the mathematical structure that it implies (because it is too broad to have much structure) and more in the intuition it tries to evoke. See also next item. For now, here are some examples:

a) If $(E, \mathcal{E}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and $T = \mathbb{R}$, then (X_t) is called *d-dimensional, real time stochastic process.*

b) If $(E, \mathcal{E}) = (\{-1,1\}, \mathcal{P}(\{-1,1\})$ and $T = \mathbb{Z}^d$, then (X_t) is called *spin system* with lattice $\mathbb{Z}^d$ and binary spins.

c) If $T = \mathbb{N}_0$, then (X_t) is called a discrete time stochastic process; for example, a simple random walk would have $E = \mathbb{Z}$.

(1.3) Remark

There are two fundamental points of view on a stochastic process, both of them are useful in some ways and less useful in others. The first is the *dynamical point of view*: (X_t) is thought of as a random quantity and t as a time, and the random quantity changes with time. This is natural for examples a) and c) above.

The other point of view is the *global point of view*: a stochastic process can also be viewed as *one single* random variable that takes values in the space

$$\Omega = E^T \equiv \{f : T \mapsto E\}$$

of sample paths. This amounts to considering the „random quantity“ (X_t) at all times at once! In other words, instead of looking at each X_t separately as t evolves, one looks at the whole path $t \mapsto X_t$ as one single random object. This is natural e.g. for example b) above, where T anyway does not look like a time, and $(X_t)_{t \in T}$ can be visualized as an infinitely large pattern of -1's and 1's located at the vertices of $\mathbb{Z}^d$. But it is also useful in example c). When we apply it, we imagine that instead of discovering $(X_n)_{n \in \mathbb{N}}$ separately for each n, each $\omega \in \Omega = E^{\mathbb{N}}$ represents one possible evolution of the random events into the infinite future, i.e. for all n at once.

An important question is what σ-algebra should be used on Ω, but we will come to that.

(1.4) Definition

Let $\boldsymbol{X} = (X_t)_{t \in T}$ be a stochastic process, and assume that the state space E is a group (the most important examples are $E = \mathbb{R}$ and $E = \mathbb{R}^n$), and that $T \subset \mathbb{R}$. The set of random variables $(X_{s,t})_{s,t \in T}$ with

$$X_{s,t} := X_t - X_s$$

is called the **set of increments** of the process $\boldsymbol{X}$.

A stochastic process is said to have **independent increments** if for all $n \in \mathbb{N}$ and for all $s_1 < t_1 \leqslant s_2 < t_2 \leqslant \ldots \leqslant s_n < t_n$ with $s_i, t_i \in T$, the random variables $(X_{s_i,t_i})_{1 \leqslant i \leqslant n}$ are independent.

In the same situation as above, the stochastic process is said to have **stationary increments** if for all $r \in \mathbb{R}$ so that $s_i + r \in T$ and $t_i + r \in T$ for all i, we have $(X_{s_i,t_i})_{i=1,\ldots,n} \sim (X_{s_i+r,t_i+r})_{i=1,\ldots,n}$.

Remark: a)$(X_{s,t})_{s,t \in T}$ is a stochastic process with index set $\{(s,t) \in T \times T : s < t\}$ and state space E, called the increment process.

b) Importantly, independence of increments is only required when the intervals $[s_i, t_i)$ do not overlap for different i. This is important because otherwise we could take $s_1 = s_2, t_1 = t_2$, and then there would be only very boring stochastic processes with independent increments.

c) For shorter presentation, the disjointness property is also required in the definition of stationary increments. But here it is not necessary, and it actually implies the stationary increment property for all sets of increments. It is a useful exercise to convince yourself that this is so.

(1.5) Definition

A $\mathbb{R}^d$-valued stochastic process $\boldsymbol{B} = (B_t)_{t\in\mathbb{R}_0^+}$ is called a **Brownian motion** if it has the following five properties:

(B0): $B_0(\omega) = 0$ for almost all ω, i.e. $B_0 = 0$ almost surely.

(B1): $\boldsymbol{B}$ has independent increments.

(B2): $\boldsymbol{B}$ has stationary increments.

(B3): $B_t - B_s \sim B_{t-s} \sim \mathcal{N}(0,(t-s)\mathrm{id}_{\mathbb{R}^d})$, where $\mathcal{N}$ is the d-dimensional normal distribution[1)] and $\mathrm{id}_{\mathbb{R}^d}$ is the d-dimensional identity matrix.

(B4): The map $t \mapsto B_t(\omega)$ is continuous for *all* (note: *not only* almost all) $\omega \in \Omega$.

(1.6) Remark

Let us interpret the properties (B0)-(B4) in the light of the formula $B_t = \int_0^t \xi_s \,\mathrm{d}s$, where ξ_s is the white noise from the introduction. We ignore the fact that ξ does not exist as a function, and just work heuristically. We have:

(B0) is necessary since $\int_0^0 \xi_s \,\mathrm{d}s = 0$ whatever ξ might be.

(B1) represents the complete lack of memory („whiteness“) in the white noise: We have $B_{s,t} = \int_s^t \xi_r \,\mathrm{d}r$ for all s,t, and for $s_1 < t_1 \leqslant s_2 < t_2$ the sets $(\xi_r)_{s_1 \leqslant r<t}$ and $(\xi_r)_{s_2 \leqslant s<t}$ are sets of independent random variables. So also the integrals B_{s_1,t_1} and B_{s_1,t_2} are independent.

(B2) just means that white noise is not changing with time: the integral $\int_{s_1}^{t_1} \xi_s \,\mathrm{d}s$ only depends on the length $t_1 - s_1$ of the integration interval.

(B3) comes from the central limit theorem: if we believe that all the ξ_s are independent of each other, then $B_{t-s} = \int_0^{t-2} \xi_r \,\mathrm{d}r$ is a sum of (infinitely many!) iid random variables, and must therefore be a Gaussian.

(B4) is again very reasonable since $t \mapsto \int_0^t f(s)\,\mathrm{d}s$ should be continuous for any function f, even a non-existent one like ξ. More seriously, and only for those who might have already heard about these things or are interested in them: ξ as a distribution is regular enough to guarantee the continuity of the map $t \mapsto (\mathbb{1}_{[0,t]}, \xi)$, where the bracket is the dual pairing via the ordinary L^2-scalar product.

Since Gaussian measures play a central role in most of what follows, we next give a short but complete (for our purposes) and systematic account of their most important properties.

(1.7) Definition

The Gaussian measure (or: normal distribution) with mean $m \in \mathbb{R}$ and variance $\sigma^2 \geqslant 0$ is the

[1)] We will recall (of learn about) the most important facts about multi-dimensional Gaussians (or normal distributions) just below, see in particular (1.13) probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with Lebesgue density

$$g_{m,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(- \frac{(x-m)^2}{2\sigma^2}\Big).$$

(1.8) Proposition

Let $X \sim \mathcal{N}(m, \sigma^2)$. Then

a) $\mathbb{E}(X) = m$, $\mathbb{V}(X) = \sigma^2$.

b) For $\sigma > 0$ and all $C > 0$, we have

$$\frac{1}{\sqrt{2\pi}} \frac{C}{C^2+1} \, \mathrm{e}^{-C^2/2} \leqslant \mathbb{P}(X - m \geqslant C\sigma) \leqslant \frac{1}{\sqrt{2\pi}} \frac{1}{C} \, \mathrm{e}^{-C^2/2}.$$

c) For real sequences $(m_k)_{k\in\mathbb{N}}$ and $(\sigma_k)_{k\in\mathbb{N}}$ and $m, \sigma \in \mathbb{R}$ we have

$$\Big(\lim_{k\to\infty} m_k = m \text{ and } \lim_{k\to\infty} \sigma_k = \sigma \Big) \quad \Leftrightarrow \quad \mathcal{N}(m_k, \sigma_k^2) \stackrel{k\to\infty}{\longrightarrow} \mathcal{N}(m, \sigma^2) \text{ in distribution.}$$

Proof: exercise.

(1.9) Definition

A $\mathbb{R}^d$-valued random variable X is called d**-dimensional Gaussian** if for all linear maps $L : \mathbb{R}^d \to \mathbb{R}$, there exist $m, \sigma \in \mathbb{R}$ with $LX \sim \mathcal{N}(m, \sigma^2)$.

Remark: In words: a random variable is d-dimensional Gaussian if and only if the image measure of its distribution under all linear maps $L : \mathbb{R}^d \to \mathbb{R}$ is one-dimensional Gaussian. Explicitly, this means that with $X = (X^1, \ldots, X^d)$, we need that for any choice $a_1, \ldots, a_d \in \mathbb{R}$, the real random variable $\sum_{i=1}^d a_i X^i$ is Gaussian.

(1.10) Example and warning

If $X^1, \ldots, X^d$ are *independent* one-dimensional Gaussians, then $(X^1, \ldots, X^d)$ is a d-dimensional Gaussian. But without independence, this is not necessarily true. As an example, let

$$X^1 \sim \mathcal{N}(0,1), \quad X^2(\omega) = \begin{cases} X^1(\omega) & \text{for those } \omega \text{ that give } |X^1(\omega)| \leqslant 1, \\ -X^1(\omega) & \text{for those } \omega \text{ that give } |X^1(\omega)| > 1. \end{cases}$$

Then $X^2 \sim \mathcal{N}(0,1)$ (to see this calculate $\mathbb{P}(X^2 < c)$ for all $c \in \mathbb{R}$), but the pair (X^1, X^2) is *not* a Gaussian random variable. One can see this by noticing that $|X^2(\omega) + X^1(\omega)| \leqslant 2$ for all $\omega \in \Omega$, and $\mathbb{P}(X^1 + X^2 \neq 0) > 0$. So with $L(x,y) = x + y$ the image measure of the distribution of (X^1, X^2) under L is a nonzero distribution with compact support, and thus can not be Gaussian.

From a systematic point of view, requiring that X^i is Gaussian for all i means that the image measures under all coordinate projections are Gaussian, or in the phrasing of the remark above, that all a_i except one are zero. But since there are many more linear maps from $\mathbb{R}^d$ to $\mathbb{R}$ than there are coordinate projections (for example the map L above), this is not sufficient to fulfil Definition (1.9). For independent Gaussians, on the other hand, it is enough, since we know (e.g. by studying the characteristic function, see below) that the sum of *independent* Gaussian random variables is Gaussian, and so again in the language of the remark above, it is clear that $\sum_{i=1}^{d} a_i X^i$ is Gaussian if all the X^i are independent Gaussians.

(1.11) Proposition

Let X be a real random variable. $X \sim \mathcal{N}(m, \sigma^2)$ if and only if its characteristic function is given by

$$\varphi_X(u) = \mathrm{e}^{\mathrm{i}um}\, \mathrm{e}^{-\frac{1}{2}u^2\sigma^2} \qquad (*)$$

Proof: Recall that $\varphi_X(u) := \mathbb{E}(\mathrm{e}^{\mathrm{i}uX})$ uniquely determines the distribution of X. Therefore we only need to show that $(*)$ holds when $X \sim \mathcal{N}(m, \sigma^2)$. Since $\varphi_{X+m}(u) = \mathrm{e}^{\mathrm{i}mu}\varphi_X(u)$, we can restrict to the case $m = 0$. For this case, we calculate

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}u}\varphi_X(u) &= \frac{1}{\sqrt{2\pi\sigma^2}}\int \mathrm{i}x\, \mathrm{e}^{\mathrm{i}ux}\, \mathrm{e}^{-\frac{x^2}{2\sigma^2}}\, \mathrm{d}x == \frac{1}{\sqrt{2\pi\sigma^2}}\int \mathrm{i}\, \mathrm{e}^{\mathrm{i}ux}\Big(x\, \mathrm{e}^{-\frac{x^2}{2\sigma^2}}\Big)\, \mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}\int \mathrm{i}\Big(\mathrm{i}u\, \mathrm{e}^{\mathrm{i}ux}\Big)\sigma^2\, \mathrm{e}^{-\frac{x^2}{2\sigma^2}}\, \mathrm{d}x = -u\sigma^2\varphi_X(u).\end{aligned}$$

For the first equality, integration under the integral is justified by the integrability of $x \mapsto |x|\, \mathrm{e}^{-\frac{x^2}{2\sigma^2}}$, and in the third equality we integrated by part using the fact that $\partial_x \mathrm{e}^{-\frac{x^2}{2\sigma^2}} = \frac{x}{\sigma^2}\mathrm{e}^{-\frac{x^2}{2\sigma^2}}$. Since we have $\varphi_X(0) = 1$, for the function $h(u) = \ln\varphi_X(u)$ we get

$$h'(u) = \frac{\varphi_X'(u)}{\varphi_X(u)} = -u\sigma^2, \qquad h(0) = 0.$$

Integrating gives $h(u) = -\frac{1}{2}u^2\sigma^2$, and this shows the claim. □

(1.12) Corollary

For $X \sim \mathcal{N}(0, \sigma^2)$ and $\zeta \in \mathbb{C}$ we have

$$\mathbb{E}(\mathrm{e}^{\zeta X}) = \mathrm{e}^{+\sigma^2\zeta^2/2}.$$

Proof: The function $\zeta \mapsto \mathbb{E}(\mathrm{e}^{\zeta X})$ is analytic on $\mathbb{C}$ (e.g. estimate its Taylor coefficients using the integral with Gaussian density). It is given by $\mathrm{e}^{+\sigma^2\zeta^2/2}$ for $\zeta = \mathrm{i}u$, $u \in \mathbb{R}$. Therefore by the uniqueness of analytic continuation, the claim follows. □

(1.13) Theorem and Definition

Let $\boldsymbol{X} = (X^i)_{1 \leqslant i \leqslant d}$ be a d-dimensional Gaussian.

a) The distribution of X is uniquely determined by its **mean vector**

$$\boldsymbol{m} := \mathbb{E}(\boldsymbol{X}) = (\mathbb{E}(X^i))_{1 \leqslant i \leqslant d} \in \mathbb{R}^d,$$

and its **covariance matrix**

$$C = (C_{i,j})_{1 \leqslant i,k \leqslant d} \in \mathbb{R}^{d\times d} \quad \text{with} \quad C_{i,j} := \mathrm{Cov}(X^i, X^j) = \mathbb{E}(X^iX^j) - \mathbb{E}(X^i)\mathbb{E}(X^j).$$

We write $\mathcal{N}(\boldsymbol{X}, C)$ for such a Gaussian.

b) $\boldsymbol{X}$ has a Lenesgue-density on $\mathbb{R}^d$ if and only if C is invertible. In this case,

$$\mathbb{P}(\boldsymbol{X} \in \mathrm{d}\boldsymbol{x}) = \underbrace{\frac{1}{(2\pi)^{d/2}(\det C)^{1/2}} \exp\Big(-\frac{1}{2}\big(\boldsymbol{x} - \boldsymbol{m}, C^{-1}(\boldsymbol{x} - \boldsymbol{m})\big)\Big)}_{=:\psi(\boldsymbol{x})} \mathrm{d}\boldsymbol{x},$$

where $(.,.)$ is the scalar product in $\mathbb{R}^d$, and the above formula is a useful and suggestive way to write that $\psi(\boldsymbol{x})$ is the Lebesgue density of the distribution of $\boldsymbol{X}$.

Proof: Assume that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are d-dimensional Gaussians, both with mean $\boldsymbol{m}$ and covariance matrix C. Then for all $\boldsymbol{a} \in \mathbb{R}^d$, the random variables $Z = \sum_{i=1}^d a_i X^i$ and $W = \sum_{i=1}^d a_i Y^i$ are one-dimensional Gaussians with $\mathbb{E}(Z) = \mathbb{E}(W) = (\boldsymbol{a}, \boldsymbol{m})$, and

$$\mathbb{V}(Z) = \mathbb{V}(W) = (a, Ca). \qquad (*)$$

Therefore,

$$\varphi_{\boldsymbol{X}}(\boldsymbol{a}) = \mathbb{E}(\,\mathrm{e}^{\mathrm{i}(\boldsymbol{a},\boldsymbol{X})}\,) \overset{(1.11)}{=} \mathrm{e}^{\mathrm{i}(\boldsymbol{a},\boldsymbol{m})}\ \mathrm{e}^{-\frac{1}{2}(\boldsymbol{a},C\boldsymbol{a})} = \varphi_{\boldsymbol{Y}}(\boldsymbol{a})$$

for all $\boldsymbol{a} \in \mathbb{R}^d$, and thus the characteristic functions of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are equal. This implies that the distributions of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are equal, which shows the claim.

b) Assume first that C is not invertible. Then there exists $0 \neq \boldsymbol{y} \in \mathbb{R}^d$ with $C\boldsymbol{y} = 0$, and so

$$\mathbb{V}\big((\boldsymbol{y}, \boldsymbol{X})^2\big) = (\boldsymbol{y}, C\boldsymbol{y}) = 0.$$

This means that $(\boldsymbol{y}, \boldsymbol{X}) = (\boldsymbol{y}, \boldsymbol{m})$ almost surely, and so the distribution of $\boldsymbol{X}$ is concentrated on the hyperplane. Therefore it can note have a Lebesgue-denisty.

Assume now that C is invertible. This ensures that ψ as defined in the claimed equation is well-defined and nonnegative. What remains is to compute the characteristic function of $\psi(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$ and verify that it matches the function φ_X given in the first part of the proof. This is left as an exercise. □

(1.14) Proposition

Let $\boldsymbol{X}$ be a d-dimensional Gaussian, $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{m}, C)$ and $A \in \mathbb{R}^{n \times d}$. Then

$$A\boldsymbol{X} \sim \mathcal{N}(A\boldsymbol{m}, ACA^*),$$

where A is the Hermitian conjugate of A.

Proof: exercise.

(1.15) Proposition

Let $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{m}, C)$. Then $X^1, \ldots, X^d$ are uncorrelated if and only if they are independent.

Proof: Independent random variables are always uncorrelated, wo the one implication is trivial. For the other implication, note that uncorrelatedness of the X^i implies that C is a diagonal matrix, with diagonal elements $C_{i,i}$. Let $Y^1, \ldots, Y^d$ be independent Gaussians with $\mathbb{E}(Y_i) = \mathbb{E}(X_i)$ and with $\mathbb{V}(X_i) = C_{i,i}$ for all i. Then $\varphi_{\boldsymbol{Y}}(\boldsymbol{a}) = \varphi_{\boldsymbol{X}}(\boldsymbol{a})$ can be easily verified for all $\boldsymbol{a} \in \mathbb{R}^d$, and thus by Theorem (1.13) the distributions of $\boldsymbol{Y}$ and $\boldsymbol{X}$ are equal. □

Remark: Be careful: it is *not* true that uncorrelated Gaussians are always independent. They need to be *jointly* Gaussian in the sense of Definition (1.9). The reason is similar to the one given in (1.10).

It is time to come back to the topic of stochastic processes and Brownian motion!

(1.16) Definition

Let (X_t) be a stochastic processes with state space $(E, \mathcal{E})$, defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The set of **finite dimensional distributions** (short: fidis) of $\boldsymbol{X}$ is the family of probability measures

$$\{p_{t_1,\ldots,t_n} : t_1, \ldots, t_n \in T, t_i \neq t_j \text{ if } i \neq j, n \in \mathbb{N}\}, \quad \text{where} \quad p_{t_1,\ldots,t_n} = \mathbb{P} \circ (X_{t_1}, \ldots X_{t_n})^{-1}$$

is the image measure of $\mathbb{P}$ under the measurable map $\Omega \to E, \omega \mapsto (X_{t_1}(\omega), \ldots, X_{t_n}(\omega))$. Put differently, $p_{t_1,\ldots,t_n}$ is the unique probability measure on $(E^n, \mathcal{E}^{\otimes n})$ with $p_{t_1,\ldots,t_n}(A) = \mathbb{P}((X_{t_1}, \ldots, X_{t_n}) \in A)$ for all $A \in \mathcal{E}^{\otimes n}$.

(1.17) Example

Let $T = \mathbb{N}$, $(X_n)_{n\in\mathbb{N}}$ be a simple random walk, i.e. $X_n = \sum_{i=1}^n X_i$ where (X_i) are iid random variables it $\mathbb{P}(X_i = 1) = \mathbb{P}(X = -1) = 1/2$. Then, for example, for $A, B, C \subset \mathbb{Z}$,

$$p_{1,4,9}(A \times B \times C) = \mathbb{P}(X_1 \in A, X_4 \in B, X_9 \in C).$$

So, the $p_{t_1,\ldots,t_n}$ are used to measure what the process does at the precise times $t_1, \ldots, t_n$, but are blind to everything that it does at any other times. Each finite dimensional distributions thus only controls a finite number of points in time of the process, hence the name.

(1.18) Proposition

Let $\boldsymbol{X}$ be as in (1.16). Then the finite dimensional distributions fulfil the **consistency conditions**: for all $t_1, \ldots, t_n \in T$, all $C_1, \ldots, C_n \in \mathcal{E}$, and all permutations $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$, we have

$$\text{(C1)} \qquad p_{t_1,\ldots,t_n}(C_1, \ldots, C_n) = p_{t_{\sigma(1)},\ldots,t_{\sigma(n)}}(C_{\sigma(1)} \times \cdots \times C_{\sigma(n)}),$$

$$\text{(C2)} \qquad p_{t_1,\ldots,t_n}(C_1 \times \cdots \times C_{n-1} \times E) = p_{t_1,\ldots,t_{n-1}}(C_1 \times \cdots \times C_{n-1}).$$

Proof: easy exercise.

Remark: By (C1), we may always assume that $t_1 < t_2 < \cdots < t_n$ when talking about fidis.

(1.19) Definition

A $\mathbb{R}^d$-valued stochastic process is called **Gaussian process** if all its fidis are Gaussian measures.

(1.20) Remark

a) Explicitly, this means that each $p_{t_1,\ldots,t_n}$ is the distribution of a $\mathbb{R}^{dn}$-dimensional (jointly) Gaussian random variable. In fact, the relevant random variable is simply given by $\omega \mapsto (X_{t_1}(\omega), \ldots, X_{t_n}(\omega))$.
b) Example (1.10) shows that there are stochastic processes that are not Gaussian processes, but where the *one-dimensional* distributions p_t are Gaussian vor all t. In fact, the example given in (1.10) is such a process if we take $T = \{1, 2\}$, $E = \mathbb{R}$ and $X_1 = X^1$, $X_2 = X^2$.
c) If $\boldsymbol{X}$ is a Gaussian process, then its fidis are fully determined by the two functions

$$T \to \mathbb{R}^d, \qquad t \mapsto \mathbb{E}(X_t) \qquad \text{(the mean)},$$
$$T^2 \to \mathbb{R}^{d\times d}, \qquad (s,t) \mapsto \mathrm{Cov}(X_s, X_t) \qquad \text{(the covariance function)}.$$

This follows from Theorem (1.13).

(1.21) Theorem

a) A $\mathbb{R}^d$-valued Brownian Motion (short: BM^d) is a Gaussian process $\mathbf{B}$ with mean $\mathbb{E}(B_t) = 0$ for all t and (matrix-valued!) covariance function

$$\mathrm{Cov}(B_s, B_t) := \mathbb{E}(B_s^* B_t) = \min\{s,t\}\mathrm{id}_{\mathbb{R}^d} \equiv (s \wedge t)\mathrm{id}_{\mathbb{R}^d}.$$

Here B_s^* is the column vector $(B_s^1, \ldots, B_s^d)^{\mathrm{t}}$, i.e. $B_s^* B_t$ is the rank one matrix A with entries $A_{i,j} = B_s^i B_t^j$.
b) Conversely, a Gaussian process with the above mean and covariance is a BM^d if it also fulfils the continuous paths condition (B4).

Proof: Let $0 < t_1 < t_2 < \ldots < t_n$. Define the matrix

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & 1 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 \\ 1 & 1 & 1 & 1 & \cdots & 1 \end{pmatrix}.$$

Then

$$(B_{t_1}(\omega), \ldots, B_{t_n}(\omega)) = A\big(B_{t_1} - B_0, B_{t_2}(\omega) - B_{t_1}(\omega), \ldots, B_{t_n}(\omega) - B_{t_{n-1}}(\omega)\big)^t.$$

By (B1), (B3) and (1.10), we have $(B_{t_i} - B_{t_{i-1}})_{1 \leqslant i \leqslant n} \sim \mathcal{N}(0, C)$ with $C_{i,j} = \delta_{i,j}(t_i - t_{i-1})$. By (1.14), this implies $p_{t_1,\ldots,t_n} \sim \mathcal{N}(0, ACA^*)$, so $\boldsymbol{B}$ is a Gaussian process. Clearly its mean is zero, and for the covariance we compute

$$\mathrm{Cov}(B_s, B_t) = \mathbb{E}((B_s - 0)(B_t - 0)) = \underbrace{\mathbb{E}((B_t - B_s)B_s)}_{=0,\text{independent increments}} + \mathbb{E}(B_s^2) = s = s \wedge t.$$

b) We need to check (B0)-(B4). (B4) is explicitly assumed to hold, and (B0) follows from $\mathbb{E}(B_0) = \mathbb{V}(B_0) = \mathbb{E}(B_0^2) = 0 \wedge 0 = 0$. For (B1), (B2) and (B3), we note that the assumed form of the covariance function implies that for $t_1 < \cdots < t_n$, the covariance matrix of $(B_{t_1}, \ldots, B_{t_n})$ is given my

$$M = \begin{pmatrix} t_1 & t_1 & t_1 & t_1 & \cdots & t_1 \\ t_1 & t_2 & t_2 & t_2 & \cdots & t_2 \\ t_1 & t_2 & t_3 & t_3 & \cdots & t_3 \\ t_1 & t_2 & t_3 & t_4 & \cdots & t_4 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ t_1 & t_2 & t_3 & t_4 & \cdots & t_n \end{pmatrix}$$

The matrix A from the first part of the proof has inverse

$$A^{-1} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}.$$

Therefore, $(B_{t_1} - B_0, B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}})$ has covariance matrix

$$M' = A^{-1}M(A^{-1})^* = \begin{pmatrix} t_1 & 0 & \cdots & 0 \\ 0 & t_2 - t_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & t_n - t_{n-1} \end{pmatrix}.$$

This implies (B1)-(B3). □

(1.22) Proposition

Let $B^1, \ldots, B^d$ be independent 1-dimensional Brownian motions. Then $\boldsymbol{B} := (B_t^1, \ldots, B_t^d)_{t \geqslant 0}$ is a d-dimensional BM. Conversely, if we are given a d-dimensional BM $\tilde{\boldsymbol{B}}$, then the coordinate processes $(\tilde{B}_s^i)_{s \geqslant 0}$ are one-dimensional Brownian Motions for all $i \leqslant d$.

Proof: exercise.

(1.23) Proposition

Let $\boldsymbol{B}$ be a one-dimensional Brownian motion. Then its fidis have the following Lebesgue-density: for all $0 = t_0 < t_1 < \ldots < t_n$, we have

$$p_{t_1,\ldots,t_n}(\mathrm{d}\boldsymbol{x}) \equiv \mathbb{P}(B_{t_1} \in \mathrm{d}x_1, \ldots, B_{t_n} \in \mathrm{d}x_n) = \phi_{t_1,\ldots,t_n}(\boldsymbol{x})\mathrm{d}\boldsymbol{x}, \qquad (*)$$

with

$$\phi_{t_1,\ldots,t_n}(x_1, \ldots x_n) = \prod_{i=1}^{n} \left(\frac{\exp\left(\frac{-(x_j - x_{j-1})^2}{t_j - t_{j-1}}\right)}{2\pi(t_j - t_{j-1})} \right)^{1/2}$$

with the convention $x_0 = 0$. An equivalent way to write this fact is to state that for all $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$, we have

$$\mathbb{P}(B_{t_1} \in A_1, \dots, B_{t_n} \in A_n) = \frac{1}{(2\pi)^{n/2}\Big(\prod_{i=1}^n (t_j - t_{j-1})\Big)^{1/2}} \int_{A_1\times\cdots\times A_n} \mathrm{e}^{-\frac{1}{2}\sum_{j=1}^n \frac{(x_j - x_{j-1})^2}{t_j - t_{j-1}}} \,\mathrm{d}\boldsymbol{x}.$$

Proof: exercise - use Theorem (1.21).

We have now defined Brownian Motion axiomatically and connected it with the theory of Gaussian processes, but we do not yet know that there even exists a stochastic process fulfilling (B0)-(B4). Our next task is to see that this is the case. This will take a while. We start by making sure that the fidis we found in (1.23) have the chance of belonging to a bona fide stochastic process, i.e. that they fulfil the consistency conditions from (1.18).

(1.24) Proposition

Consider the family of time-ordered fidis $\{p_{t_1,\dots,t_n} : 0 < t_1 < t_2 \dots < t_n\}$, where the $p_{t_1,\dots t_n}$ are given by the right hand side of equation $(*)$ in (1.23). Extend this family in the unique way that satisfies condition (C1) of (1.18) to a family of non-time-ordered fidis. Then this family is consistent, i.e. it also fulfils condition (C2) of (1.18).

Proof: a direct, but somewhat tedious calculation that involves some integration arithmetics. Left as an exercise.

(1.25) Definition

Let $(E, \mathcal{E})$ be a measurable space and T a set.

a) The maps $\pi_t : E^T \to E, (e_s)_{s\in T} \to e_t$ are called the **coordinate projections** to the t-th coordinate. When we identify E^T with the set $\{f : T \to E\}$ of all functions from T to E, then $\pi_t(\boldsymbol{e}) = \boldsymbol{e}(t) = \boldsymbol{e}_t$ is the point evaluation of the function $\boldsymbol{e}$ at the point t.

b) The σ-algebra $\mathcal{E}^{\otimes T}$ is the smallest σ-algebra on E^T so that all the maps $\pi_t : E^T \to E$ are $\mathcal{E}^{\otimes T}$-$\mathcal{E}$-measurable.

c) The measurable space $(E^T, \mathcal{E}^{\otimes T})$ is called the **canonical measurable space** for E-valued stochastic processes with index space T.

d) If $\Omega_0 \subset E^T$ is any (not necessarily measurable) subset of E^T, then the σ-algebra (!)

$$\mathcal{E}^{\otimes T} \cap \Omega_0 := \{A \cap \Omega_0 : A \in \mathcal{E}^{\otimes T}\}$$

is called the **trace** of $\mathcal{E}^{\otimes T}$ on Ω_0. The measurable space $(\Omega_0, \mathcal{E}^{\otimes T} \cap \Omega_0)$ is the canonical measurable space for E-valued stochastic processes with index set T and sample paths in Ω_0.

(1.26) Example

(i): For $E = \mathbb{R}^d$, $T = \mathbb{R}_0^+$ and $\Omega = E^T = \{\omega : \mathbb{R}_0^+ \to \mathbb{R}^d\}$, $\pi_t(\omega) = \omega(t)$, we see that Ω is the space of all paths $t \mapsto \omega(t)$ of the process $\boldsymbol{X}$ with $X_t(\omega) = \omega(t)$.

(ii): With

$$\Omega_{\mathrm{c}} = C_0(\mathbb{R}_0^+, \mathbb{R}^d),$$

the measurable space $(\Omega_c, \mathcal{B}(\mathbb{R}^d)^{\otimes \mathbb{R}_0^+} \cap \Omega_c)$ is the canonical space for stochastic processes with continuous paths.
(iii):

$$\Omega_0 = C_0(\mathbb{R}_0^+, \mathbb{R}^d) = \{\omega \in C(\mathbb{R}_0^+, \mathbb{R}^d) : \omega(0) = 0\},$$

the measurable space $(\Omega_0, \mathcal{B}(\mathbb{R}^d)^{\otimes \mathbb{R}_0^+} \cap \Omega_0)$ is the canonical space for stochastic processes with continuous paths starting in 0.

(1.27) Remark

The metric of **local uniform convergence** on $C_0 \equiv C_0(\mathbb{R}_0^+, \mathbb{R}^d)$ is given by

$$\rho : C_0 \times C_0 \to \mathbb{R}_0^+, \quad \rho(f,g) = \sum_{n=1}^{\infty} \big(1 \wedge \sup_{0 \leqslant t \leqslant n} |f(t) - g(t)|\big) 2^{-n}.$$

The Borel-σ-algebra $\mathcal{B}(C_0)$ on C_0 is the smallest σ-algebra on C_0 such that all ρ-open subsets of C_0 are $\mathcal{B}(C_0)$-measurable. We have

$$\mathcal{B}(C_0) = \mathcal{B}(\mathbb{R})^{\otimes \mathbb{R}_0^+} \cap C_0.$$

Proof: exercise.

(1.28) Lemma

Let $(E, \mathcal{E})$ be a measurable space and T a set. Then for $A \subset E^T$, we have

$$A \in \mathcal{E}^{\otimes T} \qquad \Longleftrightarrow \qquad \exists I \subset T, I \text{ countable, with } A \in \sigma(\pi_t : t \in I).$$

Proof: exercise.

We recall a theorem from the probability theory course. For the purpose of this lecture, we add a Definition that is a bit more general than what we had before. In my lecture notes on probability theory, part c) below is proved as Theorem 3.29.

(1.29) Definition and Theorem

a) Let $(S, \mathcal{S})$ and $(T, \mathcal{T})$ be measurable spaces. A map $\mu : S \times \mathcal{T} \to [0,1]$ is called a **probability kernel** from S to T if
(i): $s \mapsto \mu(s, A)$ is $\mathcal{S}$-measurable for all $A \in \mathcal{T}$,
(ii): $A \mapsto \mu(s, A)$ is a probability measure on $(T, \mathcal{T})$ for all $s \in S$.

b) Let $(\Omega, \mathcal{F})$ be a measurable space, $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra. A probability kernel[2)] $\mu : \Omega \times \mathcal{F} \to [0,1]$ is called a **regular conditional probability** with respect to the σ-algebra $\mathcal{G}$ if for almost all $\omega \in \Omega$, we have

$$\mu(\omega, A) = \mathbb{P}(A \,|\, \mathcal{G})(\omega).$$

c) If $(\Omega, \mathcal{F})$ is a Borel-space, then for each $\mathcal{G} \subset \mathcal{F}$ a regular conditional probability exists.

[2)] in the notation of a) we have $(S, \mathcal{S}) = (\Omega, \mathcal{G})$ and $(T, \mathcal{T}) = (\Omega, \mathcal{F})$

We will also need the following technical but useful fact:

(1.30) Lemma

Let $(S, \mathcal{S})$ be a measurable space, $(T, \mathcal{T})$ a Borel space, μ a probability kernel from S to T, and Y a $\mathcal{U}([0,1])$-distributed[3)] random variable. Then there exists a $\mathcal{S} \otimes \mathcal{B}([0,1])$-measurable function $f : S \times [0,1] \to T$ such that

$$\mu(s, A) = \mathbb{P}(f(s, Y) \in A) = \int_0^1 \mathbb{1}_{\{f(s,.) \in A\}}(u) \, \mathrm{d}u$$

for all $s \in S$, $A \in \mathcal{T}$.

Proof: A Borel space is by definition isomorphic to a Borel subset of $[0,1]$ as a measurable space, i.e. there exists a Borel subset $U \subset [0,1]$ and a bijective map $h : U \to T$ so that h and h^{-1} are measurable. We first show the claim for the case $T = U$ and $\mathcal{T} = \mathcal{B}([0,1]) \cap U$. In this case, we define the function $f : S \times U \to [0,1]$ by

$$f(s,t) := \sup\{x \in [0,1] : \mu(s, U \cap [0,x]) < t\}, \qquad \text{for all } s \in S, t \in U.$$

We claim that for all $c \in [0,1]$ we have

$$f(s,t) \leqslant c \quad \Leftrightarrow \quad \sup\{x \in [0,1] : \mu(s, U \cap [0,x]) < t\} \leqslant c \quad \Leftrightarrow \quad \mu(s, U \cap [0,c]) \geqslant t.$$

Indeed, the first equivalence is by definition, and for the second we look at the two possibilities for the right hand side:

if $\mu(s, U \cap [0,c]) \geqslant t$ is true, then $c \notin A_t := \{x \in [0,1] : \mu(s, U \cap [0,x]) < t\}$. By the monotonicity of the map $x \mapsto \mu(s, U \cap [0,x])$, A_t is an interval starting from 0. Therefore, its rightmost point $f(s,t)$ can be at most equal to c, in other words $f(s,t) \leqslant c$.

If, on the other hand, $\mu(s, U \cap [0,c]) \geqslant t$ is not true, then by the continuity from above of the measure $\mu(s,.)$ we have $\lim_{n\to\infty} \mu(s, U \cap [0, c + 1/n]) < t$, which means that there must be $\tilde{c} > c$ with $\mu(s, U \cap [0, \tilde{c}]) < t$. So $\tilde{c} \in A_t$, which means that $f(s,t) \geqslant \tilde{c} > c$. This shows the claimed equivalences.

Next we show that the function f is $\mathcal{S} \otimes \mathcal{B}([0,1])$-measurable. For this we first note that the function $F_c : S \times U \to \mathbb{R} \cup \{\infty\}, (s,t) \mapsto \mu(s, U \cap [0,x])/t$ is jointly measurable as the quotient of two (trivially) jointly measurable functions. Then the measurability of f follows from the equality

$$f^{-1}([0,c]) = F_c^{-1}([1,\infty]) \in \mathcal{S} \times (\mathcal{B}([0,1]) \cap U)$$

for all c.

Since $Y \sim \mathcal{U}([0,1])$, for all $x \in [0,1]$ we have

$$\mathbb{P}(f(s,Y) \leqslant c) = \mathbb{P}(\mu(s,[0,c]) \geqslant Y) = \mu(s,[0,c]),$$

and since probability measures on $[0,1]$ are determined by their values on intervals, this shows the claim for the case $T = U$.

For the case of a general Borel space T, let U and $h : T \to U$ be as indicated in the beginning of the proof. We define the probability kernel $\tilde{\mu}(s,B) = \mu(s, h^{-1}(B))$ for all $B \in \mathcal{B}([0,1]) \cap U$, and find a function $\tilde{f}$ and a $\mathcal{U}([0,1])$ random variable Y with $\tilde{\mu}(s,B) = \mathbb{P}(\tilde{f}(s,Y) \in B)$ by

[3)] this means: uniformly distributed on $[0,1]$

the first part of our proof. Now define $f(s,t) = h^{-1} \circ f(s,t)$, check measurability (easy), and compute

$$\mu(s,A) = \tilde{\mu}(s,h(A)) = \mathbb{P}(\tilde{f}(s,Y) \in h(A)) = \mathbb{P}(f(s,Y) \in A),$$

which shows the claim. □

The following theorem is the central step in the proof of existence os stochastic processes.

(1.31) Theorem

Let $(E,\mathcal{E})$ be a Borel space. For each $n \in \mathbb{N}$, let $\mathbb{P}_n$ be a probability measure on $(E^n, \mathcal{E}^{\otimes n})$, and assume that the consistency equations

$$\mathbb{P}_{n+1}(A \times E) = \mathbb{P}_n(A)$$

hold for all $A \in \mathcal{E}^{\otimes n}$ and all $n \in \mathbb{N}$. Then on the probability space

$$(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1]^{\mathbb{N}}, \mathcal{B}([0,1])^{\otimes \mathbb{N}}, \mathcal{U}([0,1])^{\otimes \mathbb{N}}),$$

there exists a family $(X_i)_{i\in\mathbb{N}}$ of E-valued random variables so that for all $n \in \mathbb{N}$ and all $A \in \mathcal{E}^{\otimes n}$ we have

$$\mathbb{P}_n(A) = \mathbb{P}((X_1, \ldots, X_n) \in A). \qquad (*)$$

In other words, a stochastic process with the fidis given by the $\mathbb{P}_n$ exists.

Proof: We construct the X_i recursively. For this purpose, we write $\omega \in \Omega$ as $\omega = (\omega_1, \omega_2, \ldots)$ and define $Y_i(\omega) = \omega_i$. Then the (Y_i) are iid, $\mathcal{U}([0,1])$-distributed random variables.

For the first step we are looking for a random variable X_1 so that $\mathbb{P}_1(A) = \mathbb{P}(X_1 \in A)$. We did this in Theorem (2.22) of the course „Einführung in die Stochastik“, where we constructed X_1 as the generalized inverse of the distribution function of $\mathbb{P}_1$. Alternatively, we can use the trivial probability kernel $\mu_0(\boldsymbol{x}, A) = \mathbb{P}_1(A)$ from $\{0\}$ to E in the context of Lemma (1.30), and find a function f with $\mathbb{P}_1(A) = \mathbb{E}(f(Y_1) \in A)$, so $X_1 = f(Y_1)$.

Assume now that we have already constructed $X_1, \ldots, X_n$ so that $(*)$ holds up to n, and that in addition each X_i is $\sigma(Y_1, \ldots, Y_i)$-measurable. Let

$$\mathcal{G}_n := \sigma(\{A_1 \times \cdots \times A_n \times E : A_i \in \mathcal{E}\ \forall i\}),$$

be the sub-σ-algebra of $\mathcal{E}^{\otimes(n+1)}$ that only depends on the information contained in the first n coordinates. The regular conditional probability

$$\mu_n : E^{n+1} \times \mathcal{E}^{\otimes(n+1)} \to [0,1] \text{ with } \mu_n(\boldsymbol{x}, A) = \mathbb{P}_{n+1}(A \,|\, \mathcal{G}_n)(\boldsymbol{x}) \text{ almost surely}$$

exists by (1.29). Since $\boldsymbol{x} \mapsto \mu_n(\boldsymbol{x}, A)$ is $\mathcal{G}_n$-measurable, it depends only on $x_1, \ldots, x_n$ but not on x_{n+1}. In other words, there exists a function $\tilde{\mu} : E^n \times \mathcal{E}^{\otimes(n+1)} \to [0,1]$ with $\mu(\boldsymbol{x}, A) =$

$\tilde{\mu}((x_1, \dots, x_n), A)$. We then have

$$\mathbb{P}_{n+1}(A_1 \times \cdots \times A_{n+1}) = \mathbb{E}_{n+1}\big(\mathbb{P}_{n+1}(A_1 \times \cdots \times A_{n+1} \,|\, \mathcal{G}_n)\big) = \mathbb{E}_{n+1}\Big(\mathbb{P}_{n+1}(A_{n+1} \,|\, \mathcal{G}_n) \prod_{i=1}^n \mathbb{1}_{A_i}\Big) =$$

$$= \int \tilde{\mu}((x_1, \dots, x_n), A_{n+1}) \prod_{i=1}^n \mathbb{1}_{A_i}(x_i) \mathbb{P}_{n+1}(\mathrm{d}\boldsymbol{x})$$

$$= \int \tilde{\mu}((x_1, \dots, x_n), A_{n+1}) \prod_{i=1}^n \mathbb{1}_{A_i}(x_i) \mathbb{P}_n(\mathrm{d}\boldsymbol{x}) = \mathbb{E}\Big(\tilde{\mu}((X_1, \dots, X_n), A_{n+1}) \prod_{i=1}^n \mathbb{1}_{A_i}(X_i)\Big) =: (**).$$

In the penultimate step, we used the assumed projectivity property, and in the last we used the induction hypothesis.

By Lemma (1.30) there exists[4] $f : E^n \times [0,1] \to E$ so that

$$\tilde{\mu}((x_1, \dots, x_n), A_{n+1}) = \mathbb{P}(f(x_1, \dots, x_n, Y_{n+1}) \in A_{n+1})$$

for all $x_1, \dots, x_n \in E$. Let $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n) \subset \mathcal{F}$. Since Y_{n+1} is independent of $X_1, \dots, X_n$ by our induction hypothesis, Proposition (3.23) of the probability theory lecture notes applies and states that

$$\mathbb{P}(f(X_1, \dots, X_n, Y_{n+1}) \in A_{n+1} \,|\, \mathcal{F}_n)(\bar{\omega}) = \mathbb{P}(f(X_1(\bar{\omega}), \dots, X_n(\bar{\omega}), Y_{n+1}) \in A_{n+1}).$$

Taking these facts together, we obtain

$$\begin{aligned}
(**) &= \int \mathbb{P}\big(f(X_1(\bar{\omega}), \dots, X_n(\bar{\omega}), Y_{n+1}\big) \in A_{n+1}) \prod_{i=1}^n \mathbb{1}_{A_i}(X_i(\bar{\omega})) \mathbb{P}(\mathrm{d}\bar{\omega}) \\
&= \mathbb{E}\Big(\mathbb{P}(f(X_1, \dots, X_n, Y_{n+1}) \in A_{n+1} \,|\, \mathcal{F}_n) \prod_{i=1}^n \mathbb{1}_{A_i}(X_i)\Big) \\
&= \mathbb{P}\Big(f(X_1, \dots, X_n, Y_{n+1}) \in A_{n+1}, X_1 \in A_1, \dots, X_n \in A_n).
\end{aligned}$$

Setting $X_{n+1}(\omega) = f(X_1(\omega), \dots, X_n(\omega), Y_{n+1}(\omega))$, we have constructed X_{n+1} that fulfils $(*)$ and is measurable with respect to $Y_i, i \leqslant n+1$. This concludes the proof. □

(1.32) Theorem (Kolmogorov 1932)

Let $(E, \mathcal{E})$ be a Borel space, T a set. Let $\{p_{t_1,\dots,t_n} : t_1, \dots, t_n \in T, n \in \mathbb{N}\}$ be a family of probability measures that fulfil the consistency conditions (1.18), (C1) and (C2). Then there exists a probabiltiy measure $\mathbb{P}$ on $(E^T, \mathcal{E}^{\otimes T})$ with

$$p_{t_1,\dots,t_n}(A) = \mathbb{P}((\pi_{t_1}, \dots, \pi_{t_n}) \in A) \qquad \forall A \in \mathcal{E}^{\otimes n}, n \in \mathbb{N}, t_1, \dots, t_n \in T.$$

(Recall Definition (1.25)).

Proof: Let $A \in \mathcal{E}^{\otimes T}$. By Lemma (1.28),

$$\exists I \subset T, I \text{ countable with } A \in \sigma(\pi_t : t \in I).$$

[4] here we need that products of Borel spaces are Borel spaces, which is left as an exercise

We can therefore write $A = B \times E^{T\setminus I}$ for some $B \in \mathcal{E}^{\otimes I}$. By Theorem (1.31), there exists a (unique (!)) probability measure $\mathbb{P}_I$ on $\mathcal{E}^{\otimes I}$ with

$$p_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n) = \mathbb{P}_I(\pi_{t_i} \in A_i \ \forall i \leqslant n), \qquad \forall n \in \mathbb{N}, A_1, \dots, A_n \in \mathcal{E}, t_1, \dots, t_n \in I.$$

Now for each $A \in \mathcal{E}^{\otimes T}$, pick I and B with $A = B \times E^{T\setminus I}$, define $\mathbb{P}_I$ as above, and set $\mathbb{P}(A) = \mathbb{P}_I(B)$. The choice of I is not unique, but consistency guarantees that the value of $\mathbb{P}(A)$ does not depend on the choice of I, thus $\mathbb{P}$ is a well-defined map from $\mathcal{E}^{\otimes T}$ to $[0, 1]$. It remains to show its σ-additivity. So let $(A_j)_{j\in\mathbb{N}}$ be disjoint elements of $\mathcal{E}^{\otimes T}$, choose I_j, B_j for each of them so that $A_j = B_{I_j} \times E^{T\setminus I_j}$, and put $I = \bigcup_{i=1}^{\infty} I_j$. Then $A_j = \tilde{B}_j \times E^{T\setminus I}$ with $\tilde{B}_j = B_{I_j} \times E^{I\setminus I_j}$. Then the $\tilde{B}_j$ form a disjoint family of elements from $\mathcal{E}^{\otimes I}$, I is still countable, and thus the σ-additivity of $\mathbb{P}_I$ gives

$$\mathbb{P}(\bigcup_{j\in\mathbb{N}} A_j) = \mathbb{P}_I(\bigcup_{j\in\mathbb{N}} \tilde{B}_j) = \sum_{j\in\mathbb{N}} \mathbb{P}_I(\tilde{B}_j) = \sum_{j\in\mathbb{N}} \mathbb{P}(A_j).$$

□

(1.33) Corollary and Definition

A $\mathbb{R}^d$-valued stochastic process fulfilling (B0)-(B3) from (1.5) exists. Explicitly, there exists a (unique) probability measure $\mathcal{W}_0$ on $(\Omega, \mathcal{F}) = ((\mathbb{R}^d)^{\mathbb{R}_0^+}, (\mathcal{B}(\mathbb{R}^d))^{\mathbb{R}_0^+})$ such that the random variables

$$B_t : \Omega \to \mathbb{R}^d, \qquad \omega \mapsto B_t(\omega) = \pi_t(\omega) = \omega_t$$

have the properties (B0)-(B3). $\mathcal{W}$ is called the **pre-Wiener measure**.

It remains to show that property (B4) also holds. The slight difficulty here is that the statement „$\mathcal{W}_0(C(\mathbb{R}_0^+, \mathbb{R}^d)) = 1$“ makes no sense, because $C(\mathbb{R}_0^+, \mathbb{R}^d)$ is not an element of $\mathcal{F}$ (exercise). Therefore we need to be slightly more careful.

(1.34) Definition

Let $D \subset \mathbb{R}_0^+$, $\alpha > 0$. A function $f : \mathbb{R}_0^+ \to \mathbb{R}^d$ is called **α-Hölder continuous** on D if

$$\|f\|_{D,\alpha} := \sup\{\frac{|f(t) - f(s)|}{|t - s|^\alpha} : s, t \in D, s \neq t\} < \infty.$$

In this case we write $f \in C^\alpha(D)$.

f is **locally α-Hölder continuous** on D if $\|f\|_{D\cap[0,n],\alpha} < \infty$ for all $n \in \mathbb{N}$. We then write $f \in C^\alpha_{\text{loc}}(D)$.

(1.35) Remarks

a) If D has no cluster points, then $f \in C^\alpha(D)$ for all functions f.

b) Usually we will take D to be a dense subset of some interval $[a, b] \subset \mathbb{R}_0^+$.

c) In the case described in b), $f \in C^\alpha(D)$ can be uniquely extended to a function $f \in C^\alpha([a, b])$: for $x \in [a, b]$, we define $f(x) = \lim_{n\to\infty} f(x_n)$ for any sequence $(x_n) \subset D$ with $x_n \to x$. You should check that $\|f\|_{D,\alpha} < \infty$ guarantees the independence of this limit from the chosen

sequence.
d) If D is dense and $f \in C^{\alpha}_{\mathrm{loc}}(D)$ for some $\alpha > 1$, then f is constant
e) $f \in C^1_{\mathrm{loc}}$ means that f is locally Lipschitz continuous.
f) $\|.\|_{D,\alpha}$ is only a semi-norm: we have $\|f\|_{D,\alpha} = 0$ for all constant f.
g) The map $\|.\|_{D,\alpha} : C(\mathbb{R}^+_0, \mathbb{R}^d) \to [0,\infty]$ is $(\mathcal{B}(\mathbb{R}^d))^{\otimes \mathbb{R}^+_0}$-$\mathcal{B}([0,\infty])$-measurable if D is countable (exercise!).

(1.36) Theorem

Let $(X_t)_{t\in[0,T]}$ be a $\mathbb{R}^d$ valued stochastic process. Assume that there exist $q \geqslant 2$, $\beta > \frac{1}{q}$ and $C < \infty$ such that

$$\forall s,t \in [0,T] \text{ with } |t-s| < \tfrac{1}{2}: \qquad \mathbb{E}(|X_{s,t}|^q) \leqslant C|t-s|^{\beta q} \qquad (*)$$

(or, equivalently, $|X_{s,t}|_{L^q} \equiv \mathbb{E}(|X_{s,t}|^q)^{1/q} \leqslant C|t-s|^{\beta}$). Then for the choice

$$D = \{2^{-n}k : k \in \mathbb{N}, n \in \mathbb{N}\} \cap [0,T] \qquad \text{(the dyadic rational numbers)}$$

we have

$$\mathbb{E}(\|X\|^q_{D,\alpha}) < \infty \qquad \forall \alpha \in \big[0, \beta - \tfrac{1}{q}\big).$$

Proof: Let $D_n := \{2^{-n}k : k \in \mathbb{N}\} \cap [0,T]$. Then $D = \bigcup_{n\in\mathbb{N}} D_n$. We define

$$K_n(\omega) := \max\{|X_{t,t+2^{-n}}(\omega)| : t \in D_n\} = \max\{|X_{s,t}(\omega)| : s,t \text{ are neighbous in } D_n\}.$$

We begin by estimating the q-th moment of this random variable, something we will need later. We have

$$\mathbb{E}(K_n^q) \leqslant \mathbb{E}\Big(\sum_{t\in D_n} |X_{t,t+2^{-n}}|^q\Big) \overset{(*)}{\leqslant} |D_n| C (2^{-n})^{\beta q} \leqslant T 2^n C (2^{-n})^{\beta q} = CT2^{-n(\beta q - 1)}. \quad (**)$$

The next task is to find a way to estimate $\sup_{s,t\in D} \frac{|X_{s,t}(\omega)|}{|t-s|^\alpha}$ in terms of the $K_n(\omega)$, for all ω. First we note that we may restrict to the supremum to $|t-s| < \frac{1}{2}$, because if $|t-s| > \frac{1}{2}$, we can choose $m \leqslant 2T$ and $t_0, \dots t_m \in D$ with $t_0 = s$, $t_m = t$, and $|t_i - t_{i-1}| < 1/2$ for all i. Since $|X_{s,t}| \leqslant \sum_{i=1}^m |X_{t_{i-1},t_i}|$, we have

$$\frac{|X_{s,t}|}{|t-s|^\alpha} \leqslant \sum_{i=1}^m \frac{|X_{t_{i-1},t_i}|}{|t_i - t_{i-1}|^\alpha} \leqslant 2T \sup\Big\{\frac{X_{s,t}}{|t-s|^\alpha} : s,t \in D, |t-s| < \frac{1}{2}\Big\}.$$

Now we can take the supremum over $s,t \in D$ on the left hand side and indeed find that there is only a constant factor $2T$ difference between it and the supremum over s,t with $|t-s| < 1/2$, for which we have made assumption $(*)$.
Under this assumption, for each $s < t \in D$ there exists $j \in \mathbb{N}$ with

$$2^{-j} < t - s \leqslant 2^{-j+1}.$$

Below, we will connect s and t by a chain $(s = t_0, t_1, \dots, t_{n-1}, t_n = t)$ of points from D with the properties that two consecutive points always have distance 2^{-m} from each other for some

$m \geqslant j$, and that each of the possible distances occurs at most twice in the chain. Let us assume that we have already constructed such a chain, then we can estimate

$$|X_{s,t}(\omega)| \leqslant \sum_{i=1}^{n} |X_{t_{i-1},t_i}(\omega)| \leqslant 2 \sum_{\ell=j}^{\infty} K_\ell(\omega).$$

Since $|t-s| > 2^{-j}$, we then have

$$\frac{|X_{s,t}(\omega)|}{|t-s|^\alpha} \leqslant 2^{j\alpha} \cdot 2 \sum_{\ell=j}^{\infty} K_\ell(\omega) \leqslant 2 \sum_{\ell=j}^{\infty} 2^{\ell\alpha} K_\ell(\omega) \leqslant 2 \sum_{\ell=0}^{\infty} 2^{\ell\alpha} K_\ell(\omega).$$

Since the right hand side now no longer depends on s and t, we obtain

$$\mathbb{E}\big(\|X\|_{D,\alpha}^q\big)^{1/q} = \mathbb{E}\Big(\sup\big\{\frac{|X_{s,t}|}{|t-s|^\alpha} : s,t \in D\big\}^q\Big)^{1/q} \leqslant \mathbb{E}\Big((2T)^q\big(\sum_{\ell=0}^{\infty} 2^{\ell\alpha} K_\ell\big)^q\Big)^{1/q} \leqslant$$

$$\leqslant 4T\Big|\sum_{\ell=0}^{\infty} 2^{\ell\alpha} K_\ell\Big|_{L^q} \leqslant 4T \sum_{\ell=0}^{\infty} 2^{\ell\alpha} |K_\ell|_{L^q} \overset{(**)}{\leqslant} 4CT^2 \sum_{\ell=0}^{\infty} 2^{\ell\alpha} 2^{-\ell(\beta q-1)/q}.$$

The sum on the right hand side is finite when $\alpha < \beta - 1/q$, showing the claim.

It remains to construct the chain $t_0, \ldots, t_n$. To follow the construction, it is very useful to generate a drawing as you read the steps. We start by observing that since $2^{-j} < t-s \leqslant 2^{-j+1}$, the set $(s,t) \cap D_j$ contains either one or two points. We call them $T_j^{\pm}$, with $t_j^- \leqslant t_j^+$. Now from t_j^-, we look to the left: the interval $[s, t_{j^-})$ contains at most one point from D_{j+1}. If it does, we define this point to be t_{j+1}^-, if not we take $t_{j+1}^- = t_j^-$. Now we look left again from t_{j+1}^-, the interval $[a, t_{j+1}^-)$ again contains at most one point from D_{j+2}. As before, we define this point to be t_{j+2}^- if it exists, otherwise we take $t_{j+2}^- = t_{j+1}^-$. This now continues until we hit s, which will happen at the latest when $j = N$. The same procedure is then repeated with t_j^+, this time looking the the right. It is easy to see that we obtain a chain of ordered points, each distance occurs at most twice, and that there are at most $n = 2(N-j)$ such points. This finishes the construction of the $t_0, \ldots, t_n$ and thus the proof. □

(1.37) Definition

Let $\boldsymbol{X} = (X_t)_{t \in \mathbb{R}_0^+}$ be a stochastic process with values in $\mathbb{R}^d$. A stochastic process $\tilde{\boldsymbol{X}} = (\tilde{X}_t)_{t \in \mathbb{R}_0^+}$ on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ as $\boldsymbol{X}$ is called a **continuous version** of $\boldsymbol{X}$ if

(i): $\tilde{\boldsymbol{X}}$ has continuous paths *for all* ω, i.e. for all $\omega \in \Omega$ the map $t \mapsto \tilde{X}_t(\omega)$ is continuous.

(ii): For each $t \in \mathbb{R}_0^+$, we have $\mathbb{P}(X_t = \tilde{X}_t) = 1$.

This definition circumvents the problem that the set of continuous functions might not be measurable by simply requiring continuity for all paths of $\tilde{\boldsymbol{X}}$. The second requirement ensures that we cannot distinguish $\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$ for all practical purposes, in particular their fidis are the same.

(1.38) Theorem

Assume that a stochastic process X satisfies condition $(*)$ of (1.36).

a) **Kolmogorov-Chentsov-Theorem:** Then there exists a continuous version $\tilde{X}$ of X.

b) $\tilde{X}$ can be realized as a random variable on the canonical probability space Ω_c described in Example (1.26), i.e. there exists a probability measure $\mathbb{P}_c$ on this space so that when $Y_t(\omega) := \pi_t(\omega) = \omega(t)$ for all $\omega \in \Omega_c$, then the process (Y_t) has the same fidis as $(\tilde{X}_t)$ (and also the same as (X_t)).

Proof: a) By Theorem (1.36), $\mathbb{E}(\|X\|^q_{D_{T,\alpha}}) < \infty$ for some $\alpha > 0, q > 0$ and for all T, and so

$$\mathbb{P}(\|X\|_{D_T,\alpha} < \infty) = 1 \qquad \Longrightarrow \qquad \mathbb{P}(X \in C^\alpha_{\text{loc}}(D)) = 1.$$

Note in particular that the set $\Omega_0 := \{\omega \in \Omega : \|X(\omega)\|_{D_T,\alpha} < \infty\}$ depends only on the values of $X_t(\omega)$ for t in a countable set and is therefore measurable. We define

$$\tilde{X}_t(\omega) := \begin{cases} \lim_{t_n \to t} X_{t_n}(\omega) \text{ for some } (t_n) \subset D, t_n \to t & \text{if } \omega \in \Omega_0, \\ 0 & \text{otherwise,} \end{cases}$$

and note that $\tilde{X}_t(\omega)$ is independent of the approximating sequence for all $\omega \in \Omega_0$ and all $t \in \mathbb{R}^+_0$. Also, $t \mapsto 1\tilde{X}_t(\omega)$ is continuous for all $\omega \in \Omega$. So requirement (i) of Definition (1.37) is fulfilled. To verify requirement (ii), we fix $t \in \mathbb{R}^+_0$ and define

$$Z_t(\omega) := X_t(\omega) - \tilde{X}_t(\omega), \qquad Z_{t_n}(\omega) := X_t(\omega) - X_{t_n}(\omega)$$

for $(t_n) \subset D$, $t_n \to t$. Then $Z_{t_n} \to Z_t$ on Ω_0, i.e. almost surely, and Chebyshevs inequality gives

$$\mathbb{P}(|Z_{t_n}| > \varepsilon) \leqslant \frac{1}{\varepsilon^q}\mathbb{E}(|X_{t_n} - X_t|^q) \overset{(*) \text{ in } (1.36)}{\leqslant} C|t_n - t|^{\beta q} \overset{n\to\infty}{\longrightarrow} 0.$$

Thus Z_{t_n} converges to zero in probability, and thus its almost sure limit Z_t is equal to zero almost surely. This concludes the proof of a).

b) The map $F : \Omega \to \Omega_c, \omega \mapsto (\tilde{X}_t(\omega))_{t \in \mathbb{R}^+_0}$ is measurable - you can easily check this for sets generated by the inverse images under coordinate projections, and then it extends to the full σ-algebra. Now simply take $\mathbb{P}_c = \mathbb{P} \circ F^{-1}$. □

(1.39) Theorem

a) There exists a stochastic process satisfying (B0)-(B4) from Definition (1.5), i.e. a Brownian motion.

b) There exists a probability measure $\mathcal{W}_0$ on $(C_0, \mathcal{F})$ so that under $\mathcal{W}$, the coordinate projections $B_t : C_0 \to \mathbb{R}^d, \omega \mapsto \omega(t)$ form a BM^d. The space $(C_0, \mathcal{F}, \mathcal{W})$ is called **Wiener space** and $\mathcal{W}$ ist called the **Wiener measure**.

Proof: By (1.22), we only need to prove the theorem for $d = 1$. By (1.33), a process $(\tilde{B}_t)$ fulfilling (B0)-(B3) exists. (B3) implies

$$\mathbb{E}(|\tilde{B}_{s,t}|^q) = \mathbb{E}_{\mathcal{N}(0,t-s)}(|X|^q) \overset{\text{exercise H3}}{=} c_q|t-s|^{q/2}.$$

Thus (1.36($*$)) holds for all $\beta < 1/2$, all $q > 1$ and all $|t - s| < 1$. By Theorem (1.38), the claims follow. □

Remark: By taking $q \to \infty$ in the proof above, we see that Brownian paths are α-Hölder continuous almost surely for all $\alpha < 1/2$ (remember that we need $\alpha < \beta - 1/q$). $\alpha = 1/2$ is where the proof fails to provide information, and we will see below that this is for good reason: it will turn out that Brownian paths are *not* α-Hölder-continuous for $\alpha = 1/2$ (and thus not for any larger α.).

2. Properties of Brownian Motion

A: Invariance properties

(2.1) Orthogonal invariance

Let B be a BM^d and U an orthogonal matrix. Then $(UB_t)_{t\in T}$ is a BM^d. In particular, $-B$ is a BM^d.

Proof: $UU^* = \mathrm{id}_{\mathbb{R}^d}$, so Proposition (1.14) shows that $(UB)_{s,t} \sim \mathcal{N}(0,(t-s)\mathrm{id}_{\mathbb{R}^d})$, and that $\mathbb{E}(UB_t) = 0$ for all t. Since the map U is continuous from $\mathbb{R}^d$ to itself, the paths $t \mapsto UB_t(\omega)$ are continuous for all ω. Then Theorem (1.21) shows the claim. □

(2.2) Time shift invariance

Let $(B_t)_{t\geqslant 0}$ be a BM^d, $a \in \mathbb{R}_0^+$. Then the stochastic process $(B_{t+a} - B_a)_{t\geqslant 0}$ is a BM^d.

Proof: exercise.

(2.3) Elementary Markov property (aka memorylessness)

Let $(B_t)_{t\geqslant 0}$ be a BM^d, $a \in \mathbb{R}_0^+$. Then the stochastic processes $(B_t)_{0\leqslant t\leqslant a}$ and $(W_t)_{t\geqslant 0} = (B_{t+a} - B_a)_{t\geqslant 0}$ are independent (this means: $\sigma(B_t : 0 \leqslant t \leqslant a) \perp\!\!\!\perp \sigma(W_t : t \geqslant 0)$). In particular,

$$\mathbb{E}\Big(F\big((W_t)_{t\geqslant 0}\big)\,\Big|\,\sigma\big(B_s : s \leqslant a\big)\Big) = \mathbb{E}\Big(F\big((B_t)_{t\geqslant 0}\big)\Big)$$

for all bounded, measurable $F : C_0 \to \mathbb{R}$.

Proof: The sets

$$\Big\{\bigcap_{i=1}^n \{B_{t_i} \in A_i\} : n \in \mathbb{N}, t_1 < \cdots < t_n \in [0,a], A_1,\ldots,A_n \in \mathcal{B}(\mathbb{R}^d)\Big\}$$

are a $\cap$-stable generator of $\sigma((B_t)_{0\leqslant t\leqslant a})$, and since the matrix A from the proof of Theorem (1.21) is invertible, the same is true for the family

$$\mathcal{A}_1 := \Big\{\bigcap_{i=1}^n \{B_{t_i} - B_{t_{i-1}} \in C_i\} : n \in \mathbb{N}, 0 = t_0 < t_1 \cdots < t_n \in [0,a], C_1,\ldots,C_n \in \mathcal{B}(\mathbb{R}^d)\Big\}.$$

Similarly,

$$\mathcal{A}_2 := \Big\{\bigcap_{i=1}^n \{W_{t_i} - W_{t_{i-1}} \in C_i\} : n \in \mathbb{N}, 0 = t_0 < t_1 \cdots < t_n \in [0,a], C_1,\ldots,C_n \in \mathcal{B}(\mathbb{R}^d)\Big\}$$

generates $\sigma((W_t)_{t\geqslant 0}) = \sigma((B_{a,t})_{t\geqslant a})$. By (B1), $A_1 \perp\!\!\!\perp A_2$ whenever $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$, which (by definition) means that $\mathcal{A}_1 \perp\!\!\!\perp \mathcal{A}_2$. Since σ-algebras are independent if they have independent $\cap$-stable generators, this proves the claim. □

(2.4) Invariance under diffusive rescaling

Let $(B_t)_{t\geqslant 0}$ be a BM^d. Then for all $c>0$, the process $(\frac{1}{\sqrt{c}}B_{ct})_{t\geqslant 0}$ is a BM^d.

Proof: exercise.

(2.5) Time reversal invariance

Let $(B_t)_{0\leqslant t\leqslant T}$ be a BM^{d}on $[0,T]$. Then $(B_{T-t}-B_T)_{0\leqslant t\leqslant T}$ is a BM^{d}on $[0,T]$.

Proof: exercise.

(2.6) Time involution invariance

Let $(B_t)_{t\geqslant 0}$ be a BM^d. Define

$$W_t(\omega)=\begin{cases} tB_{1/t}(\omega) & \text{if } t>0\\ 0 & \text{if } t=0.\end{cases}$$

Then $(W_t)_{t\geqslant 0}$ is a BM^d.

Proof: Clearly, W is a Gaussian process and $W_0=0$ almost surely and $\mathbb{E}(W_t)=0$ for all t. Its covariance is, fr $s<t$,

$$\mathrm{Cov}(W_s,W_t)=\mathrm{Cov}(sB_{1/s},tB_{1/t})=st\mathrm{Cov}(B_{1/s},B_{1/t})=st(\tfrac{1}{s}\wedge\tfrac{1}{t})=st\tfrac{1}{t}=s=s\wedge t.$$

By Theorem (1.21) (B0) - (B3) are now guaranteed, and it remains to check (B4). The map $t\mapsto 1/t$ is continuous for all $t>0$, and so the map $t\mapsto tB_{1/t}$ is continuous on $(0,\infty)$. What remains to be seen is the continuity at $t=0$. We have

$$\Omega_0:=\{\omega\in\Omega:\lim_{t\to 0}|W_t(\omega)|=0\}=\{\omega\in\Omega:\forall n\geqslant 1\ \exists m\geqslant 1\ \forall r\in\mathbb{Q}\cap(0,1/m]:|W_r(\omega)|\leqslant 1/n\}$$
$$=\bigcap_{n\geqslant 1}\bigcup_{m\geqslant 1}\bigcap_{r\in\mathbb{Q}\cap(0,1/m]}\{\omega\in\Omega:|W_r(\omega)|\leqslant 1/n\}.$$

The restriction to $\mathbb{Q}\cap(0,1/m]$ is possible because we already know that W is continuous away from $t=0$. We also know that B and W have the same fidis, and thus

$$\mathbb{P}\Big(\bigcap_{r\in A}\{|W_r|\leqslant\tfrac{1}{n}\}\Big)=\mathbb{P}\Big(\bigcap_{r\in A}\{|B_r|\leqslant\tfrac{1}{n}\}\Big)$$

for all finite subsets $A\subset\mathbb{Q}\cap(0,1/m]$. Since measures are continuous from above, we can take $A\nearrow\mathbb{Q}$ and get the same equality for $A=\mathbb{Q}\cap(0,1/m]$, all n and all m. Now by first using continuity from above and then from below, we find

$$\mathbb{P}(\Omega_0)=\lim_{n\to\infty}\mathbb{P}\Big(\bigcup_{m\geqslant 1}\bigcap_{r\in\mathbb{Q}\cap(0,1/m]}\{|W_r|\leqslant 1/n\}\Big)=\lim_{n\to\infty}\lim_{m\to\infty}\mathbb{P}\Big(\bigcap_{r\in\mathbb{Q}\cap(0,1/m]}\{|W_r|\leqslant 1/n\}\Big)$$
$$=\lim_{n\to\infty}\lim_{m\to\infty}\mathbb{P}\Big(\bigcap_{r\in\mathbb{Q}\cap(0,1/m]}\{|B_r|\leqslant 1/n\}\Big)=\mathbb{P}(\lim_{t\to 0}B_t(\omega)=0)=1.$$

This means that (W_t) is a BM^{d}on the restricted probability space $(\Omega_0,\mathcal{F}\cap\Omega_0,\mathbb{P}|_{\mathcal{F}\cap\Omega_0})$. $\square$

Time involution invariance can be used to give a very quick proof of one of the more intriguing properties of Brownian motion: although its paths are all continuous, they are still quite irregular. More precisely, we have seen in the remark after Theorem (1.39) that Brownian paths are α-Hölder continuous for all $\alpha < 1/2$. We will now show that they are *not* $1/2$-Hölder continuous, and thus in particular not differentiable at $t = 0$. By the time shift invariance (2.2), this then immediately implies the same statement for any $t \geqslant 0$.

(2.7) Proposition

Let B be one-dimensional Brownian motion. Then all of the four statements below happen with probability 1:

$$\limsup_{t\to\infty} \frac{1}{\sqrt{t}} B_t = \infty, \quad \liminf_{t\to\infty} \frac{1}{\sqrt{t}} B_t = -\infty, \quad \limsup_{t\to 0} \frac{1}{\sqrt{t}} B_t = \infty, \quad \liminf_{t\to 0} \frac{1}{\sqrt{t}} B_t = -\infty.$$

In particular, for any $t_0 \geqslant 0$, $t \mapsto B_t$ is almost surely not differentiable at t_0.

Proof: We have

$$\limsup_{t\to\infty} \frac{1}{\sqrt{t}} B_t(\omega) \geqslant \limsup_{n\in\mathbb{N}} \frac{1}{\sqrt{n}} B_n(\omega) = \limsup_{n\to\infty} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} B_{k,k-1}(\omega).$$

The $(B_{k,k-1})$ are iid random variables with $\mathbb{V}(B_{k,k-1}) = 1$ for all k. By remark (2.50 a) of the Probability Theory course, we have $\mathbb{P}(\limsup_{n\to\infty} \frac{1}{\sqrt{n}} B_n) = 1$, which gives the first two claims. For the other two claims, note that

$$\mathbb{P}(\limsup_{t\to 0} \frac{1}{\sqrt{t}} B_t = \infty) = \mathbb{P}(\limsup_{t\to\infty} \sqrt{t} B_{1/t} = \infty) = \mathbb{P}(\limsup_{t\to\infty} \frac{1}{\sqrt{t}} \underbrace{t B_{1/t}}_{=W_t} = \infty) = 1.$$

The last equality follows from the first two claims because by (2.6), (W_t) is a Brownian motion. This shows the third and fourth claim. Now non-differentiablilty at 0 is clear, and (2.2) easily transfers this to any point $t_0 \geqslant 0$. □

Remark:

By taking unions of countably many sets of measure zero, it now follows immediately that

$$\mathbb{P}(t \mapsto B_t \text{ is differentiable for some } t \in \mathbb{Q}_0^+) = 0.$$

With some more work, we can show

$$\mathbb{P}\Big(\lambda\big(\{t \in \mathbb{R}_0^+ : B \text{ is differentiable at } t\}\big) = 0\Big) = 1,$$

where λ is the Lebesgue measure on $\mathbb{R}_0^+$. We will do this in an exercise. With significantly more work, one can even show

$$\mathbb{P}(\exists t \geqslant 0 : B \text{ is differentiable at } t) = 0.$$

We refer to the literature (e.g. the book by Mörters and Peres) for the proof of this statement.

(2.8) Remark: invariances of the Wiener measure

An important reason for the significance of the Lebesgue-measure on $\mathbb{R}^d$ are its many invariances: it remains unchanged under translations and rotations. For exactly the same reason, the Wiener measure $\mathcal{W}$ on C_0 (i.e. the probability measure for Brownian motion on its canonical probability space) is so important. We have seen in the previous few items that $\mathcal{W}$ is invariant under the following maps $C_0 \to C_0$:

a) $f_1 : C_0 \to C_0, (\omega_t)_{t \geqslant 0} \mapsto (U\omega_t)_{t \geqslant t_0}$,
i.e. orthogonal transformations in the „target space“ of the function ω, see (2.1).

b) $f_2 : C_0 \to C_0, (\omega_t)_{t \geqslant 0} \mapsto (\omega_{t+a} - \omega_t)_{t \geqslant 0}$,
i.e. the operation of cutting off the first bit of the path ω and starting with time 0 at the cut point, see (2.2).

c) $f_3 : C_0 \to C_0, (\omega_t)_{t \geqslant 0} \mapsto (\frac{1}{\sqrt{c}}\omega_{ct})_{t \geqslant 0}$,
i.e. speeding up the flow of time by a factor c, but at the same time squeezing the values of ω_t by a factor $1/\sqrt{c}$, see (2.4).

d) $f_4 : C_0([0,T]) \to C_0([0,T]), (\omega_t)_{t \leqslant T} \mapsto (\omega_{T-t} - \omega_T)_{t \leqslant T}$,
i.e. running time backwards and re-shifting the values so they all start at zero.

e) $f_5 : C_0 \to C_0, (\omega_t)_{t \geqslant 0} \mapsto (t\omega_{1/t})_{t \geqslant 0}$,
i.e. reversing time in a strange way (in fact, applying what is called an involution to it), and correcting by a factor of t.

B: Martingale properties of Brownian Motion

(2.9) Definition

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.
a) A **filtration** $(\mathcal{F}_t)_{t \geqslant 0}$ is a family of σ-algebras such that for all $s < t$, we have $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$.
b) A stochastic process $(X_t)_{t \geqslant 0}$ is **adapted** to the filtration $(\mathcal{F}_t)$ if $X_t \in \mathrm{m}\mathcal{F}_t$ for all t.
c) A $\mathbb{R}^d$-valued or $\mathbb{C}$-valued stochastic process $(X_t)_{t \geqslant 0}$ is a $(\mathcal{F}_t)$**-martingale** if

(i): (X_t) adapted to $(\mathcal{F}_t)$.
(ii): $\mathbb{E}(X_t) < \infty$ for all t.
(iii): $\mathbb{E}(X_t \,|\, \mathcal{F}_s) = X_s$ $\mathbb{P}$-almost surely for all $s \leqslant t$.

d) An $\mathbb{R}$-valued stochastic process fulfilling points (i) and (ii) is called a

submartingale if $\mathbb{E}(X_t \,|\, \mathcal{F}_s) \geqslant X_s$ for all $s \leqslant t$,

and is called a

supermartingale if $\mathbb{E}(X_t \,|\, \mathcal{F}_s) \leqslant X_s$ for all $s \leqslant t$.

(2.10) Remark

For a stochastic process X, let $\mathcal{F}_t^X := \sigma(X_s : s \leqslant t)$. Then $(\mathcal{F}_X^t)_{t \geqslant 0}$ is the smallest filtration so that X is adapted.

(2.11) Proposition

BM^{d}is an $(\mathcal{F}_t^B)$-martingale.

Proof:

$$\mathbb{E}(B_t \,|\, \mathcal{F}_s^B) = \mathbb{E}(B_t - B_s \,|\, \mathcal{F}_s^B) + \underbrace{\mathbb{E}(B_s \,|\, \mathcal{F}_s^B)}_{=B_s} \overset{(2.3)}{=} \mathbb{E}(B_{t-s}) + B_s = B_s$$

□

Remark: As we will see soon, sometimes the filtration $(\mathcal{F}_t^B)$ is too small, and we want to replace it with a larger one. This is not completely harmless, however, as it can destroy the martingale property. The reason is that a $(\mathcal{F}_s)$ martingale needs to fulfil, for all $A \in \mathcal{F}_s$ and all $t \geqslant s$, the equality

$$\mathbb{E}(B_t 1\!\!1_A) = \mathbb{E}(\mathbb{E}(B_t \,|\, \mathcal{F}_s) 1\!\!1_A) \overset{(!)}{=} \mathbb{E}(B_s 1\!\!1_A),$$

where the first equality always holds by the properties of conditional expectation, but where the second one is required to be true for B to be a martingale. Making the filtration larger means that this equality needs to hold for a larger class of sets A, and making it too large may lead to the inclusion of some A where the equality no longer holds. Then B is no longer a (F_t)-martingale. This motivates the next definition.

(2.12) Definition

A filtration $(\mathcal{F}_t)$ is **admissible** for BM^d if

a) $\mathcal{F}_t^B \subset \mathcal{F}_t$ for all t.

b) $(B_{s,t}) \perp\!\!\!\perp \mathcal{F}_s$ for all $s \leqslant t$.

(2.13) Proposition

Let (B_t) be a BM^d, $(\mathcal{F}_t)$ admissible. Then (B_t) is a $(\mathcal{F}_t)$-martingale.

Proof: the same as for (2.11). □

There are many ways to get other martingales from Brownian motion. Here are some of them.

(2.14) Proposition

Let (B_t) be a BM^d, $(\mathcal{F}_t)$ admissible. Then the following stochastic processes are $(\mathcal{F}_t)$-martingales:

a) $M_t := |B_t|^2 - dt$,

b) $M_t^v := \mathrm{e}^{(v,B_t) - \frac{t}{2}\sum_{j=1}^d v_j^2}$ for all $v \in \mathbb{C}^d$
In particular $M_t^{\mathrm{i}\xi} := \mathrm{e}^{\mathrm{i}(\xi,B_t) + \frac{t}{2}|\xi|^2}$ and $M_t^{\xi} := \mathrm{e}^{(\xi,B_t) - \frac{t}{2}|\xi|^2}$ are martingales for $\xi \in \mathbb{R}^d$.

c) $M_t^n := t^{n/2} H_n(t^{-1/2} B_t)$, in case $d = 1$,
where $H_n(x) = (-1)^n \exp(x^2/2) \partial_x^n \, \mathrm{e}^{-\frac{x^2}{2}}$ is the n-th Hermite polynomial.

Proof: a) $d = 1$ is enough since $|B_t|^2 - dt = \sum_{j=1}^d ((B_t^j)^2 - t)$. Then

$$\mathbb{E}(B_t^2 \,|\, \mathcal{F}_s) = \mathbb{E}\big((B_t - B_s)^2 + 2B_t B_s - B_s^2 \,|\, \mathcal{F}_s\big) \overset{(B_t - B_s) \perp\!\!\!\perp \mathcal{F}_s}{=} \\ = \mathbb{E}(B_{t-s}^2) + 2B_s \mathbb{E}(B_s \,|\, \mathcal{F}_s) - B_s^2 = t - s + B_s^2.$$

b) $\mathbb{E}(|M_t^v|) < \infty$ by Corollary (1.12). Then,

$$\mathbb{E}(M_t^v \,|\, \mathcal{F}_s) = \mathrm{e}^{-\frac{t}{2}\sum_{j=1}^d v_j^2} \, \mathrm{e}^{(v, B_s)} \, \mathbb{E}(\mathrm{e}^{(v, B_{t-s})}) \overset{(1.12)}{=} \mathrm{e}^{-\frac{t}{2}\sum_{j=1}^d v_j^2} \, \mathrm{e}^{(v, B_s)} \, \mathrm{e}^{\frac{t-s}{2}\sum_{j=1}^d v_j^2} = M_s^v.$$

c) Exercise. You should expand $M_t^\alpha = \mathrm{e}^{\alpha B_t - \frac{t}{2}\alpha^2} = \sum_{n=0}^\infty \alpha^n \frac{B_t^n}{n!} \sum_{n=0}^\infty \alpha^{2k} \big(-\frac{t}{2}\big)^k \frac{1}{k!}$, and sort by powers of α. □

All martingales that we built from Brownian motion in the previous point had the form $M_t = f(t, B_t)$ for some function f. In order to construct different ones, we need the following result.

(2.15) Lemma

The transition density

$$p_t(x) = \frac{1}{(2\pi t)^{d/2}} \, \mathrm{e}^{-\frac{|x|^2}{2t}}$$

of BM^d solves the **heat equation**

$$\partial_t p_t(x) = \frac{1}{2} \Delta p_t(x) := \frac{1}{2} \sum_{i=1}^d \partial_{x_i}^2 p_t(x) \qquad \forall t > 0, x \in \mathbb{R}^d.$$

Proof: $\hat{p}_t(k) := \int \mathrm{e}^{\mathrm{i}(k,x)} \, p_t(x) \, \mathrm{d}x \overset{(1.11)}{=} \mathrm{e}^{-\frac{1}{2}|k|^2 t}$. Clearly, $\partial_t \hat{p}_t(k) = -\frac{1}{2}|k|^2 \hat{p}_t(k)$, and so by Lebesgue's differentiation theorem,

$$\int \mathrm{e}^{\mathrm{i}(k,x)} \, \partial_t p_t(x) \, \mathrm{d}x = \partial_t \hat{p}_t(k) = -\frac{1}{2} \int \mathrm{e}^{\mathrm{i}(k,x)} \, |k|^2 p_t(x) \, \mathrm{d}x = +\frac{1}{2} \int \sum_{j=1}^d (\partial_{x_j}^2 \, \mathrm{e}^{\mathrm{i}(k,x)}) p_t(x) \, \mathrm{d}x = (*)$$

Now integration by parts (twice) in each coordinate gives

$$(*) = \frac{1}{2} \int \mathrm{e}^{\mathrm{i}(k,x)} \, (\Delta p_t)(x) \, \mathrm{d}x,$$

for all k. Now an inverse Fourier transform yields $\partial_t p_t(x) = \frac{1}{2} \Delta p_t(x)$ for λ-almost all x, and since p_t and its derivatives are continuous, the claim follows. □

(2.16) Theorem

Let (B_t) be a BM^d, $(\mathcal{F}_t)$ admissible. Let $f \in C([0,\infty) \times \mathbb{R}^d, \mathbb{R}) \cap C^{1,2}((0,\infty) \times \mathbb{R}^d, \mathbb{R})$ with the property that there exists $C < \infty$, and a locally bounded function $c : [0,\infty) \to \mathbb{R}_0^+$ with

$$\max\{\partial_t f(t,x), \partial_{x_i}^j f(t,x), f(t,x) : 1 \leqslant i \leqslant d, j = 1,2\} \leqslant c(t) \, \mathrm{e}^{C|x|} \quad \forall x, t.$$

For $t \geqslant 0$ define $Lf(t,x) := \partial_t f(t,x) + \frac{1}{2}\Delta_x f(t,x)$, and

$$M_t^f := f(t, B_t) - f(0, B_0) - \int_0^t (Lf)(r, B_r)\,\mathrm{d}r.$$

Then $(M_t^f)_{t \geqslant 0}$ is an $(\mathcal{F}_t)$-martingale. It is called the **fundamental martingale** associated to the function f.

For the proof, we first need the following lemma, which we already know for martingales with discrete time parameter:

(2.17) Doobs maximal inequality

Let (M_t) be a submartingale with continuous paths. Then

$$\forall t \geqslant 0, \forall p > 1: \quad \mathbb{E}\big(\sup_{s \leqslant t} |M_s|^p\big) \leqslant \Big(\frac{p}{p-1}\Big)^p \mathbb{E}\big(|M_t|^p\big).$$

Proof: in case of discrete time this is (4.39) from the probability theory lecture. The transfer to continuous time is left as an exercise. □

Proof of (2.16)**:** We have

$$\begin{aligned}\mathbb{E}(M_t \,|\, \mathcal{F}_s)(\bar\omega) =& f(s, B_s(0)) - f(0,0) - \int_0^s (Lf)(r, B_r(\bar\omega))\,\mathrm{d}r + \\ &+ \mathbb{E}\Big(f(t, B_t) - f(s, B_s) - \int_s^t (Lf)(r, B_r)\,\mathrm{d}r \,\Big|\, \mathcal{F}_s\Big)(\bar\omega) = (*).\end{aligned}$$

Proposition (3.23) from the probability theory lecture tells us that when $X \perp\!\!\!\perp Y$ for two random variables, then

$$\mathbb{E}(h(X,Y) \,|\, \sigma(Y))(\bar\omega) = \mathbb{E}(h(X, Y(\bar\omega))),$$

if the integrability condition is fulfilled. We will use this fact below for the case $X = (B_{s,r})_{r \geqslant 0}$ and $Y = (B_r)_{r \leqslant s}$, and for the function

$$h(X,Y) = f(t, \underbrace{B_{s,t}}_{=\pi_t(X)} + \underbrace{B_s}_{\pi_s(Y)}) - f(s, B_s) = \int_s^t (Lf)(r, \underbrace{B_{r,s}}_{=\pi_r(X)} + \underbrace{B_s}_{=\pi_s(Y)})\,\mathrm{d}r.$$

We then get

$$\begin{aligned}(*) &= M_s^f(\bar\omega) + \mathbb{E}\Big(f\big(t, B_{s,t} + B_s(\bar\omega)\big) - f\big(s, B_s(\bar\omega)\big) - \int_s^t (Lf)(r, B_{r,s} + B_s(\bar\omega))\,\mathrm{d}r\Big) \\ &\overset{(2.2)}{=} M_s^f(\bar\omega) + \mathbb{E}\Big(\tilde f_{\bar\omega}(t-s, B_{t-s}) - \tilde f_{\bar\omega}(0,0) - \int_0^{t-s} (L\tilde f_{\bar\omega})(r, B_{r-s})\,\mathrm{d}r\Big),\end{aligned}$$

with $\tilde f_{\bar\omega}(u,x) := f(u+s, x+B_s(\bar\omega))$. Here we used that L commutes with shifts of variables. For all $\bar\omega$, the function $\tilde f_{\bar\omega}$ fulfils the conditions of Theorem (2.16) if the function f does. Therefore the proof will be finished if we can show $\mathbb{E}(M_t^f) = 0$ for all t and all suitable f. To do this, let

$\varepsilon > 0$ consider the following computation, where in the first equality we used Fubinis theorem, and this is where the conditions on f enter in the proof:

$$\mathbb{E}\Big(\int_\varepsilon^t (Lf)(s,B_s)\,\mathrm{d}s\Big) = \int_\varepsilon^t \mathbb{E}(Lf(s,B_s))\,\mathrm{d}s = \int_\varepsilon^t \Big(\int_{\mathbb{R}^d} p_s(x)(Lf)(s,x)\,\mathrm{d}x\Big)\mathrm{d}s =$$
$$= \int_{\mathbb{R}^d}\mathrm{d}x\int_\varepsilon^t \mathrm{d}t\, p_s(x)(\partial_s f)(s,x) + \int_\varepsilon^t \mathrm{d}s\int_{\mathbb{R}^d}\mathrm{d}x\, p_s(x)(\tfrac{1}{2}\Delta_x f)(s,x) \overset{IBP}{=}$$
$$= \int_{\mathbb{R}^d}\mathrm{d}x\Big(p_t(x)f(t,x) - p_\varepsilon(x)f(\varepsilon,x) - \int_\varepsilon^t \mathrm{d}s\,\partial_s p_s(x)f(s,x)\Big) + \int_\varepsilon^t \mathrm{d}s\Big(0 + \int_{\mathbb{R}^d}\mathrm{d}x\,(\tfrac{1}{2}\Delta_x p_s)(x)f(s,x)\Big) =$$
$$= \mathbb{E}(f(t,B_t)) - \mathbb{E}(f(\varepsilon,B_\varepsilon)) - \int_\varepsilon^t \mathrm{d}s\int_{\mathbb{R}^d}\mathrm{d}x\,\underbrace{\big(\partial_s p_s(x) - \Delta_x p_s(x)\big)}_{=0 \text{ since } s\geqslant\varepsilon>0} f(s,x).$$

We have shown that

$$\mathbb{E}(M_t^f) = \mathbb{E}\big(f(\varepsilon,B_\varepsilon) - f(0,0)\big) - \mathbb{E}\Big(\int_0^\varepsilon (Lf)(s,B_s)\,\mathrm{d}s\Big) \qquad \forall \varepsilon > 0$$

and need to control this expression as $\varepsilon \to 0$. On the one hand, $Lf(s,B_s) \leqslant c(s)\,\mathrm{e}^{C|B_s|}$ for some c, C by our assumptions, and therefore

$$\Big|\mathbb{E}\Big(\int_0^\varepsilon (Lf)(s,B_s)\,\mathrm{d}s\Big)\Big| \leqslant \int_0^\varepsilon c(s)\mathbb{E}\big(\,\mathrm{e}^{C|B_s|}\,\big)\,\mathrm{d}s \leqslant \int_0^\varepsilon c(s)\mathbb{E}(\,\mathrm{e}^{C|B_\varepsilon|}\,)\,\mathrm{d}s \overset{\varepsilon\to 0}{\longrightarrow} 0.$$

In the second inequality we used that $(\,\mathrm{e}^{C|B_s|}\,)_{s\geqslant 0}$ is a submartingale since $\mathrm{e}^{|\cdot|}$ is convex. This fact will also be useful when we control the other part of the expression above: we get

$$\mathbb{E}(\sup_{s\leqslant 1}|f(s,B_s)|) \leqslant \sup_{s\leqslant 1} c(s)\mathbb{E}\Big(\sup_{s\leqslant 1}\Big(\,\mathrm{e}^{\frac{1}{2}C|B_s|}\,\Big)^2\Big) \overset{(2.17)}{\leqslant} \sup_{s\leqslant 1} c(s)4\mathbb{E}\Big(\,\mathrm{e}^{C|B_1|}\,\Big) < \infty.$$

This means that the function $\omega \mapsto \sup_{s\leqslant 1}|f(s,B_s(\omega))|$ is an integrable majorant for the family of functions $(\omega \mapsto g(\varepsilon,\omega) := f(\varepsilon,B_\varepsilon(\omega)))_{0\leqslant\varepsilon\leqslant 1}$. Since for all ω we have $\lim_{\varepsilon\to 0} f(\varepsilon,B_\varepsilon(\omega)) - f(0,0) = 0$ by the continuity of f and of $s \mapsto B_s(\omega)$, dominated convergence now shows that indeed $\mathbb{E}(M_t^f) = 0$. □

(2.18) Example

a) $f(x) = |x|^2$, then $Lf(x) = \frac{1}{2}\sum_{i=1}^d 2 = d$, and so $M_t^f = |B_t|^2 - \int_0^t d\ \mathrm{d}s = |B_t|^2 - dt$ is a martingale, see (2.14 a).

b) $f(x) = x^3$, $d = 1$, then $Lf(x) = \frac{1}{2}\cdot 3\cdot 2x$, so $M_t^f = B_t^3 - 3\int_0^t B_s\,\mathrm{d}s$ is a martingale. This one we haven't seen before!

c) Let $f : \mathbb{R}^d \to \mathbb{R}$ be **harmonic**, i.e. let $\Delta f(x) = 0$ for all x. Then $Lf = 0$, and thus $(f(B_t) - f(B_0))_{t\geqslant 0}$ is a martingale. As a concrete example, we can take $d = 2$, $f(x,y) = \mathrm{e}^x \sin(y)$, then $\Delta f(x,y) = 0$ and the integrability conditions of Theorem (2.16) are fulfilled. Thus the strange stochastic process $(\,\mathrm{e}^{B_t^1}\sin(B_t^2))_{t\geqslant 0}$ is a martingale if $(B_t^1, B_t^2)_{t\geqslant 0}$ is a two-dimensional Brownian motion. Note that neither $(\,\mathrm{e}^{B_t^1}\,)_{t\geqslant 0}$ nor $(\sin(B_t^2))_{t\geqslant 0}$ are martingales by themselves - only their product is one!

d) Let $\boldsymbol{B}_t = B_t^1 + \mathrm{i}B_t^2$ be **complex Brownian motion**, i.e. B^1 and B^2 are independent one-dimensional Brownian motions. Then $(\,\mathrm{e}^{\boldsymbol{B}_t}\,)_{t \geqslant 0}$ is a martingale: indeed, $\mathrm{Re}\,(\,\mathrm{e}^{\boldsymbol{B}_t}\,) = \mathrm{e}^{B_t^1}\cos(B_t^2)$ and $\mathrm{Im}\,(\,\mathrm{e}^{\boldsymbol{B}_t}\,) = \mathrm{e}^{B_t^1}\sin(B_t^2)$, and both are martingales by (resp. in analogy to) c). We stress again that for *real-valued* Brownian motion, $(\,\mathrm{e}^{B_t}\,)$ is not a martingale, but $\mathrm{e}^{B_t - t/2}$ is one thanks to (2.14 b).

(2.19) Defintion

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(\mathcal{F}_t)$ a filtration. A **random time** is a $\mathbb{R}_0^+ \cup \{\infty\}$-valued random variable. A **stopping time** for the filtration $(\mathcal{F}_t)$ (or: an $(\mathcal{F}_t)$-stopping time) is a random time τ that fulfils

$$\{\tau \leqslant t\} \in \mathcal{F}_t \qquad \text{for all } t \in \mathbb{R}_0^+.$$

(2.20) Example

Let (X_t) be an $(\mathcal{F}_t)$-adapted process with continuous paths, and let $F \subset \mathbb{R}^d$ be a closed set. Then the **first entry time of X into F** given by

$$\tau_F := \inf\{s \geqslant 0 : X_s \in F\}$$

is a stopping time.

Proof: Let $d(x, F) := \inf\{|x - y| : y \in F\}$. You should convince yourself that the map $x \mapsto d(x, F)$ is continuous. We have the following chain of equivalences, where two of them need further justification that will be given below:

$$\begin{aligned}
\tau_F(\omega) \leqslant t &\overset{\text{Def}}{\iff} \inf\{s \geqslant 0 : X_s(\omega) \in F\} \leqslant t \\
&\overset{(1)}{\iff} \exists x \in F, s \leqslant t : X_s(\omega) = x \\
&\overset{(2)}{\iff} \inf\{d(X_s(\omega), F) : s \in [0, t]\} = 0 \\
&\overset{s \mapsto d(X_s(\omega), F) \text{ cts.}}{\iff} \inf\{d(X_s(\omega), F) : s \in [0, t] \cap \mathbb{Q}\} = 0.
\end{aligned}$$

Since $\{\omega \in \Omega : \inf\{d(X_s(\omega), F) : s \in [0, t] \cap \mathbb{Q}\} = 0\} \in \mathcal{F}_t$, the claim follows once we have justified equivalences (1) and (2).

For (1), the direction $\Leftarrow$ is trivial. For the direction $\Rightarrow$, we use that F is closed and $s \mapsto X_s$ is continuous, therefore the set $X^{-1}(F) = \{s \geqslant 0 : X_s \in F\}$ is a closed subset of $[0, \infty)$. So the infimum in the first line is in fact a minimum, and it is attained for some $s \leqslant t$.

For (2), the direction $\Rightarrow$ is trivial, and for $\Leftarrow$ we use that by the continuity of $s \mapsto d(X_s(\omega), F)$, the image of $[0, t]$ under this map is compact. So the infimum in the third line above is attained somewhere, leading back to the second line. □

The proof of the above example does not work for open sets; however, we really want to the first entry into an open set to be a stopping time too. This leads to the necessity to enlarge the filtration $\mathcal{F}_t^B$ for Brownian motion, as already announced above.

(2.21) Definition

Let X be a stochastic process. The filtration (!)

$$\mathcal{F}^X_{t+} := \bigcap_{u>t} \mathcal{F}^X_u = \bigcap_{n \geqslant 1} \mathcal{F}^X_{t+1/n}$$

is called the **right continuous completion of** $\mathcal{F}^X$. Any filtration $(\mathcal{F}_t)$ with the property that $\mathcal{F}_t = \bigcap_{n \geqslant 1} \mathcal{F}_{t+1/n}$ is called **right-continuous**.

Of course, the right continuous completion of $(\mathcal{F}^X_t)$ is right continuous. With this definition we now have what we want:

(2.22) Lemma

Let X be a stochastic process with continuous paths, $U \subset \mathbb{R}^d$ an open subset. Then

$$\tau_U := \inf\{s \geqslant 0 : X_s \in U\}$$

is an $(\mathcal{F}_{t+})$-stopping time.

Proof: We have

$$\{\tau_U \leqslant t\} = \bigcap_{n\in\mathbb{N}} \{\tau_U < t + \tfrac{1}{n}\} = \bigcap_{n\in\mathbb{N}} \bigcup_{s<t+1/n} \{X_s \in U\} = \bigcap_{n\in\mathbb{N}} \bigcup_{s\in[0,t+1/n)\cap\mathbb{Q}} \{X_s \in U\} \in \mathcal{F}_{t+}.$$

In the last step we used the continuity of the paths of X. □

We now see that for Brownian motion (and in general for continuous martingales) the completion of the minimal filtration does not destroy the martingale property:

(2.23) Lemma

Let (X_t) be a $(\mathcal{F}^X_t)$-martingale such that $t \mapsto X_t(\omega)$ is right-continuous almost surely, and such that $\mathbb{E}(|X_t|^p) < \infty$ for some $p > 1$ and all t. Then (X_t) is also an $(\mathcal{F}^X_{t+})$-martingale. In particular, $(\mathcal{F}^B_{t+})$ is admissible for Brownian motion B.

Proof: We have

$$\mathbb{E}(X_t \,|\, \mathcal{F}^X_{s+}) = \mathbb{E}(X_{s,t} \,|\, \mathcal{F}^X_{s+}) + \mathbb{E}(X_s \,|\, \mathcal{F}^X_{s+})$$

as before. Since $\mathcal{F}^X_{s+} \supset \mathcal{F}^X_s$, the second term is equal to X_s almost surely. We need to show that the first term vanishes; since it is $\mathcal{F}^X_{s+}$-measurable, it is enough to show that for all $A \in \mathcal{F}^X_{s+}$, we have

$$\mathbb{E}\Big(\mathbb{E}(X_{s,t} \,|\, \mathcal{F}^X_{s+}) \mathbb{1}_A\Big) = \mathbb{E}(X_{s,t} \mathbb{1}_A) \overset{!}{=} 0.$$

To show the last equality, we first note that for $A \in \mathcal{F}^X_{s+}$, we have $A \in \mathcal{F}^X_{s+\varepsilon}$ for all $\varepsilon > 0$ by the definition of $\mathcal{F}^X_{s+}$. Since X is an $(\mathcal{F}^X_t)$-martingale,

$$\mathbb{E}(\mathbb{1}_A X_{s+\varepsilon,t+\varepsilon}) = \mathbb{E}(\mathbb{1}_A \mathbb{E}(X_{s+\varepsilon,t+\varepsilon} \,|\, \mathcal{F}^X_{s+\varepsilon})) = 0$$

for all $\varepsilon > 0$. We have $\sup_{0<\varepsilon<1} |X_{s+\varepsilon,t+\varepsilon}| \leqslant 2(\sup_{r \leqslant t+1} |X_r|^p + 1)$ for all $p > 1$, and the latter expression is integrable by Doobs maximal inequality (2.17) and our integrability assumption. So we have found an intergable function dominating $\omega \mapsto \sup_{0<\varepsilon<1} |X_{s+\varepsilon,t+\varepsilon}(\omega)|$, and since

$\lim_{\varepsilon\to 0} X_{s+\varepsilon,t+\varepsilon}(\omega) = X_{s,t}(\omega)$ for almost all ω by right-continuity of paths, the claim now follows by dominated convergence. □

Many statements below will be stated for martingales with right-continuous paths and admissible filtrations. Of course the most important example is Brownian motion, but the added generality is useful in many places, and the proofs are not made more difficult by the greater generality. First however, we need a few more properties of stopping times.

(2.24) Proposition

Let $(\mathcal{F}_t)$ be a filtration, and let τ, σ and $(\tau_n)_{n\in\mathbb{N}}$ be $(\mathcal{F}_t)$-stopping times. Then

a) $\{\tau < t\} \in \mathcal{F}_t$ for all t.

b) If $(\mathcal{F}_t)$ is right-continuous and ρ is a random time with $\{\rho < t\} \in \mathcal{F}_t$ for all t, then ρ is a stopping time.

c) $\tau + \sigma$, $\tau \wedge \sigma$, $\tau \vee \sigma$, and $\sup_n \tau_n$ are $(\mathcal{F}_t)$-stopping times.

d) If $(\mathcal{F}_t)$ is right-continuous, then $\inf_n \tau_n$, $\liminf_n \tau_n$ and $\limsup_n \tau_n$ are $(\mathcal{F}_t)$-stopping times.

Proof: exercise.

(2.25) Lemma: Approximation of stopping times

Let τ be an $(\mathcal{F}_t)$-stopping time. If we define

$$\tau_n(\omega) = \begin{cases} (m+1)2^{-n} & \text{if } m2^{-n} \leqslant \tau(\omega) < (m+1)2^{-n}, \\ \infty & \text{if } \tau(\omega) = \infty, \end{cases}$$

then τ_n is a stopping time for all n, each τ_n only takes countably many values, and

$$\tau_n(\omega) \searrow_{n\to\infty} \tau(\omega) \quad \forall \omega \in \Omega.$$

Proof: exercise.

(2.26) Definition

Let $(\mathcal{F}_t)$ be a filtration, τ an $(\mathcal{F}_t)$-stopping time. The σ-algebra (!)

$$\mathcal{F}_\tau := \{A \in \mathcal{F} : A \cap \{\tau \leqslant t\} \in \mathcal{F}_t \;\; \forall t \geqslant 0\}$$

is called the **$\boldsymbol{\sigma}$-algebra of the $\boldsymbol{\tau}$-past**.

(2.27) Example

Let $\Omega = C_0(\mathbb{R},\mathbb{R}^d)$, $\mathcal{F}_t = \sigma(\pi_s : s \leqslant t)$, $\mathcal{F} = \bigcup_{t\,\geqslant\,0}\mathcal{F}_t$. Then for $A \in \mathcal{F}$, we have

$$A \in \mathcal{F}_t \text{ iff } \begin{cases} \forall\omega \in \Omega,\ \text{knowledge of the values } (\pi_s(\omega))_{s\,\leqslant\,t} \\ \text{is enough to determine whether } \omega \in A \text{ or not.} \end{cases}$$

$$A \in \mathcal{F}_\tau \text{ iff } \begin{cases} \forall\omega \in \Omega \text{ with } \tau(\omega) \leqslant t,\ \text{knowledge of the values } (\pi_s(\omega))_{s\,\leqslant\,t} \\ \text{is enough to determine whether } \omega \in A \text{ or not, but } \ldots \\ \forall\omega \in \Omega \text{ with } \tau(\omega) > t,\ \text{knowledge of all the values } (\pi_s(\omega))_{s\,\geqslant\,0} \\ \text{might be necessary to determine whether } \omega \in A \text{ or not.} \end{cases}$$

For example, let F, G, H be closed subsets of $\mathbb{R}^d$ and τ_F, τ_G and τ_H the respective hitting times. You should check for the sets

$$A := \{\omega \in \Omega : t \mapsto \pi_t(\omega) \text{ first hits } F, \text{ then } G, \text{ then } H\},$$
$$B := \{\omega \in \Omega : t \mapsto \pi_t(\omega) \text{ hits } F \text{ before possibly entering } G \text{ or } H\}$$

whether they are in $\mathcal{F}_\sigma$ for $\sigma = \tau_F, \tau_G$ and τ_H. Try to first decide this by the semi-heuristic explanation above, and then by checking the definition.

(2.28) Proposition

Let $(\mathcal{F}_t)$ be a filtration, τ and τ_n, $n \in \mathbb{N}$ be stopping times.

a) $\tau \in \mathrm{m}\mathcal{F}_\tau$.

b) If $(\mathcal{F}_t)$ is right-continuous, then $\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau < t\} \in \mathcal{F}_t\ \forall t\}$.

c) If $\tau_1 \leqslant \tau_2$, then $\mathcal{F}_{\tau_1} \subset \mathcal{F}_{\tau_2}$.

d) If $(\mathcal{F}_t)$ is right-continuous and $\tau_n \searrow_{n\to\infty} \tau$, then $\mathcal{F}_\tau = \bigcap_{n\in\mathbb{N}}\mathcal{F}_{\tau_n}$.

e) If (X_t) is $(\mathcal{F}_t)$-adapted, $(\mathcal{F}_t)$ is right-continuous, and $t \mapsto X_t(\omega)$ is right-continuous for all $\omega \in \Omega$, then

$$\Big(\omega \mapsto X_{\tau(\omega)}(\omega)\mathbb{1}_{\{\tau(\omega)<\infty\}}\Big) \in \mathrm{m}\mathcal{F}_\tau.$$

Proof: a,b,c) exercise.

d) By c), $\mathcal{F}_\tau \subset \bigcap_{n\in\mathbb{N}}\mathcal{F}_{\tau_n}$. Conversely, if $A \in \mathcal{F}_{\tau_n}$ for all n, then

$$A \cap \{\tau < t\} = \bigcup_{n\in\mathbb{N}} \underbrace{A \cap \{\tau_n < t\}}_{\in\mathcal{F}_t\ \forall t} \in \mathcal{F}_t \qquad \text{for all } t.$$

e) Assume first that $\tau(\Omega)$ is countable, i.e. $\tau(\Omega) = \{t_i : i \in \mathbb{N}\}$ for some numbers t_i. Then

$$\{X_\tau \in B\} \cap \{\tau \leqslant t\} = \bigcup_{k:t_k\,\leqslant\,t} \{\tau = t_k, Z_{t_k} \in B\} \in \mathcal{F}_t$$

for all $B \in \mathcal{B}(\mathbb{R}^d)$, because τ is a stopping time. For general τ, we approximate as in (2.25) by a sequence of stopping times (τ_n) with $\tau_n(\omega) \searrow \tau(\omega)$ for all ω. Since X is right-continuous, we have $\lim_{n\to\infty} X_{\tau_n(\omega)}(\omega) = X_{\tau(\omega)}(\omega)$ for all ω. Thus for $n \geqslant m$, we have

$$X_{\tau_n} \overset{\tau_n(\Omega)\text{countable}}{\in} \mathrm{m}\mathcal{F}_{\tau_n} \overset{\tau_n\,\leqslant\,\tau_m,\text{c)}}{\subset} \mathcal{F}_{\tau_m}.$$

Since $\limsup_{n\to\infty} X_{\tau_n} = X_\tau$, this implies that $X_\tau \in \mathcal{F}_{\tau_m}$ for all m, and thus

$$X_\tau \in \mathrm{m} \bigcap_{m\in\mathbb{N}} \mathcal{F}_{\tau_m} \overset{\text{d)}}{=} \mathcal{F}_\tau.$$

□

(2.29) Theorem: Doobs optional stopping theorem

Let (X_t) be an $(\mathcal{F}_t)$-submartingale with continuous paths, and let σ, τ be $(\mathcal{F}_t)$-stopping times.

a) For all $k \geqslant 0$, $(X_{\tau\wedge k})_{k\in\mathbb{N}}$ is an $(\mathcal{F}_k)$-submartingale, and also an $(\mathcal{F}_{\tau\wedge k})$-submartingale.

b) If there exists $T < \infty$ with $\sigma(\omega) \leqslant \tau(\omega) < T$ for all ω, then

$$\mathbb{E}(X_\tau \,|\, \mathcal{F}_\sigma)(\bar\omega) \geqslant X_{\sigma(\bar\omega)}(\bar\omega)$$

for $\mathbb{P}$-almost all $\bar\omega$.

Proof: for discrete martingales, this was done in the probability theory lecture. The extension to continuous martingales (by approximation!) is left as an exercise. □

(2.30) Remark

The restriction $\tau(\omega) < T$ for all ω is rather strong. But, for example, the condition $\mathbb{P}(\tau < \infty) = 1$ would not be enough. A counterexample occurs when $X = B$ is one-dimensional Brownian motion, and $\tau = \tau_{\{1\}}$ is the hitting time of the value 1, and $\sigma = 0$. Then (e.g. by (2.7)), $\mathbb{P}(\tau < \infty) = 1$, but $B_{\tau(\omega)}(\omega) = 1$ for all $\omega \in \{\tau < \infty\}$, and thus $1 = \mathbb{E}(B_\tau) \neq \mathbb{E}(B_0) = 0$.

It is possible to prove the statement of Theorem (2.29) under various weaker conditions though, see e.g. Theorem 1.93 in [Liggett: continuous time Markov processes]. For our purposes, a version of (2.29 b) that works for L^2-martingales is particularly interesting. We start by introducing a very important quantity in the theory of L^2-martingales.

(2.31) Definition

Let (X_t) be a continuous, real-valued martingale with $\mathbb{E}(X_t^2) < \infty$ for all t. A stochastic process (A_t) is called **quadratic variation process (qvp)** of X if

(i): $A_0 = 0$ almost surely,

(ii): $t \mapsto A_t(\omega)$ is increasing for almost all ω,

(iii): $(X_t^2 - A_t)_{t\,\geqslant\,0}$ is a martingale.

Remark:

a) Proposition (2.14 a) shows that $A_t(\omega) := t$ is a qvp for Brownian motion.

b) For a discrete time martingale $(X_n)_{n\in\mathbb{N}}$, in Theorem (4.49) of the probability theory lecture we showed that

$$A_n := \sum_{k=1}^{n} \mathbb{E}(X_{k-1,k}^2 \,|\, \mathcal{F}_{k-1})$$

fulfils the conditions of Definition (2.31) and is thus a quadratic variation. We also showed that A_n is the unique *predictable* quadratic variation, i.e. the only quadratic variation such that A_n

is $\mathcal{F}_{n-1}$-measurable for all n. Since in continuous time, there is no „previous time step“, the condition of previsibility does not make sense in this setting. However, condition (2.31 (iii)) implies that a qvp (A_t) must be $(\mathcal{F}_t)$-adapted.

In general, the existence of a qvp is ensured by the following theorem:

(2.32) Theorem

Let $X \in \mathcal{M}_T^2$.

a) There exists a unique qvp A of X, with continuous paths.

b) If X is uniformly bounded, i.e. $|X_t(\omega)| \leqslant K$ for some $K \in \mathbb{R}$ and all $\omega \in \Omega$, $t \leqslant T$, then A is given by the limit

$$A_t := \lim_{n\to\infty} \sum_{j=1}^{\lfloor 2^n t \rfloor} (X_{k2^{-n}} - X_{(k-1)2^{-n}})^2,$$

which exists in L^2, and thus almost surely along a subsequence.

Proof:

(i): We start by showing uniqueness. Let A and $\tilde{A}$ be two continuous qvp for the same martingale X, then (D_t) with $D_t := A_t - \tilde{A}_t$ is a martingale. Since $t \mapsto D_t(\omega)$ is the difference of two increasing functions, it has finite total variation; a continuous martingale with finite total variation must be identical to zero. We will do both statements in an exercise. Thus $D = 0$ and $A = \tilde{A}$.

(ii): To show existence, we start with the situation in b) where X is uniformly bounded. Let $D_n(t) = \{k2^{-n} : k \in \mathbb{Z}\} \cap [0, t]$ denote the dyadic rationals, and let $P_n(t)$ be the partition with separating points $D_n(t) \cup \{t\}$. For each t, we put

$$A_t^{(n)}(\omega) := \sum_{[u,v) \in P_n(t)} X_{u,v}^2(\omega).$$

Then

$$(*) \qquad X_t^2 - A_t^{(n)} = \Big(\sum_{[u,v)\in P_n(t)} X_{u,v} \Big)^2 - \sum_{[u,v)\in P_n(t)} X_{u,v}^2(\omega) = 2 \sum_{\substack{[u,v),[u',v')\in P_n(t) \\ u<u'}} X_{u,v} X_{u',v'}.$$

Now let $s < t$, and $[u, v), [u', v') \in P_n(t)$ with $u < u'$. If $v' \leqslant s$, then $\mathbb{E}(X_{u,v} X_{u',v'} \,|\mathcal{F}_s) = X_{u,v} X_{u',v'}$ since X is adapted. If $s < u'$, then

$$\mathbb{E}(X_{u,v} X_{u',v'} \,|\mathcal{F}_s) = \mathbb{E}(X_{u,v} \underbrace{\mathbb{E}(X_{u',v'} \,|\, \mathcal{F}_{u'})}_{=0 \text{ since } X \text{ martingale}} \,|\, \mathcal{F}_s) = 0,$$

and if $u' \leqslant s < v'$, then

$$\mathbb{E}(X_{u,v} X_{u',v'} \,|\mathcal{F}_s) = \mathbb{E}(X_{u,v} \underbrace{\mathbb{E}(X_{u',v'} \,|\, \mathcal{F}_s)}_{=\mathbb{E}(X_{u',s}+X_{s,v'} \,|\, \mathcal{F}_s) = X_{u',s}} \,|\, \mathcal{F}_s) = \mathbb{E}(X_{u,v} X_{u',s} \,|\mathcal{F}_s).$$

Using these equalities in the representation $(*)$, we find that $X_t^2 - A_t^{(n)}$ is a martingale for each n. Clearly, it is also continuous, but it is not a qvp because it is not monotone increasing:

making t larger changes the last term $X_{u,t}^2$ in the sum, and this term can decrease. On the other hand, on the subset $D_n(T)$ of $[0,T]$, the process $X_t^2 - A_t^{(n)}$ is increasing, and so it is plausible that the non-monotonicity goes away in the limit $n \to \infty$. But first, we have to control that limit.

(iii): Let $n, m \in \mathbb{N}$ with $n > m$. In the same way that let to $(*)$, we see that

$$A_t^{(m)} - A_t^{(n)} = 2 \sum_{[u,v)\in P_m(t)} \sum_{\substack{[w,z),[w'z')\in P_n(t)\cap[u,v) \\ w<w'}} X_{w,z}X_{w',z'} =: \sum_{[u,v)\in P_m(t)} J_{[u,v)}(X).$$

We now want to square this expression, take expectation, and see that this becomes small as $n, m \to \infty$ to establish an L^2-Cauchy sequence. Squaring gives many terms, but most of them are zero. We go through all the terms that can appear systematically, then the calculation is rather manageable. By the same conditioning trick as above, we find that for $[u_i, v_i) \in P_n(t)$, $i = 1, \dots k$, and $u_i \leqslant u_{i+1}$ for all i we have $\mathbb{E}(\prod_{i=1}^k X_{u_i,v_i}) = 0$ except when the rightmost interval appears at least twice, i.e except when $u_k = u_{k-1}$. This means that $\mathbb{E}(J_{[u,v)}J_{[u',v')}) = 0$ when $u \neq u'$, and

$$\mathbb{E}(J_{[u,v)}^2) = \sum_{\substack{[w,z),[w',z'),[w'',z'') \\ \in P_n(t)\cap[u,v): w,w'<w''}} \mathbb{E}(X_{w,z}X_{w',z'}X_{w'',z''}^2) = \mathbb{E}\Big(\sum_{[w'',z'')\in P_n(t)\cap[u,v)} X_{u,w''}^2 X_{w'',z''}^2 \Big).$$

Let $\rho_\delta(X(\omega)) = \sup\{|X_s(\omega) - X_r(\omega)| : |s-r| \leqslant \delta, 0 \leqslant s,t \leqslant T$ be the modulus of continuity of a given path $t \mapsto X_t(\omega)$. Then

$$\begin{aligned}
\mathbb{E}((A_t^{(m)} - A_t^{(n)})^2) &= 4\mathbb{E}\Big(\sum_{[u,v)\in P_m(t)} \sum_{[w,z)\in P_n(t)\cap[u,v)} X_{u,w}^2 X_{w,z}^2 \Big) \leqslant \\
&\leqslant 4\mathbb{E}\Big(\big(\rho_{2^{-m}}(X)\big)^2 \sum_{[u,v)\in P_m(t)} \sum_{[w,z)\in P_n(t)\cap[u,v)} X_{w,z}^2 \Big) = \\
&= 4\mathbb{E}\Big(\big(\rho_{2^{-m}}(X)\big)^2 \sum_{[w,z)\in P_n(t)} X_{w,z}^2 \Big) \leqslant 4\mathbb{E}\Big(\big(\rho_{2^{-m}}(X)\big)^4 \Big)^{1/2} \mathbb{E}\Big(\big(\sum_{[w,z)\in P_n(t)} X_{w,z}^2 \big)^2 \Big)^{1/2}.
\end{aligned}$$

In the last step, we used Cauchy-Schwarz. Now the random variable $\rho_\delta(X(\omega))$ converges to zero by path continuity for all ω as $m \to \infty$, and it is bounded by $2K$ by the uniform boundedness[5)]. Therefore the first factor above converges to zero as $m \to \infty$, and we will have our Cauchy sequence once we show that the second factor is bounded uniformly in $n > m$. We will actually bound it independently of m, n.

In the following calculations, all intervals will be from $P_n(t)$ and we leave this out of the notation. The conditioning trick gives

$$\mathbb{E}\Big(\sum_{[u,v),[w,z):u<w} X_{u,v}^2 X_{w,z}^2 \Big) = \mathbb{E}\Big(\sum_{[u,v)} X_{u,v}^2 \big(\sum_{[w,z):u<w} X_{w,z} \big)^2 \Big) = \mathbb{E}\Big(\sum_{[u,v)} X_{u,v}^2 X_{v,t}^2 \Big),$$

[5)]Only here we really need this - up to now, fourth moments would still be enough.

and we have

$$\mathbb{E}\Big(\big(\sum_{[u,v)} X_{u,v}^2\big)^2\Big) = \mathbb{E}\Big(\sum_{[u,v)} X_{u,v}^4 + 2 \sum_{[u,v),[w,z):u<w} X_{u,v}^2 X_{w,z}^2\Big) =$$

$$= \mathbb{E}\Big(\sum_{[u,v)} X_{u,v}^2 \underbrace{X_{u,v}^2}_{\leqslant 4K^2} + 2\sum_{[u,v)} X_{u,v}^2 \underbrace{X_{v,t}^2}_{\leqslant 4K^2}\Big) \leqslant 12K^2 \mathbb{E}\Big(\sum_{[u,v)} X_{u,v}^2\Big) =$$

$$= 12K^2 \mathbb{E}\Big(\big(\sum_{[u,v)} X_{u,v}\big)^2\Big) \leqslant 48K^4.$$

In the final equality, we used the conditioning trick one last time. We have thus shown that

$$\mathbb{E}((A_t^{(m)} - A_t^{(n)})^2) \leqslant 28K^2 \mathbb{E}\Big(\big(\rho_{2^{-m}}(X)\big)^4\Big)^{1/2}$$

for all $n \geqslant m$. Since both $M^2 - A^{(n)}$ and $M^2 - A^{(m)}$ are martingales, also $A^{(m)} - A^{(n)}$ is a martingale, and so Doobs inequality even gives

$$\mathbb{E}\Big(\|A^{(m)} - A^{(n)}\|_{L^\infty(0,t)}\Big) = \mathbb{E}\Big(\sup\{|A_s^{(m)} - A_s^{(n)}|^2 : 0 \leqslant s \leqslant t\}\Big) \leqslant 4\mathbb{E}\big((A^{(m)} - A^{(n)})^2\big),$$

and so even $(A^{(n)}_{.})_{n\in\mathbb{N}}$ is a L^2-Cauchy sequence in the Banach space $L^2(\mathbb{P}, L^\infty([0,t))$. This means that an almost sure pointwise limit (which exists) converges uniformly in $[0,t)$ for each ω, and since all $A^{(n)}$ are continuous, this defines a continuous function. This function is monotone on all dyadic rationals, and so it is monotone on $[0,t)$.

(iV): We have just shown b). To show a), we need to remove the condition that X is uniformly bounded. Here one uses the technique of *loclization*, which we will see later in more detail. The basic idea is to first replace (X_t) by $(X_{t\wedge\tau_M})$, wehre $\tau_M = M \wedge \inf\{s \geqslant 0 : |X_s| \geqslant M\}$. This is a bounded martingale (why?) and so the previous reasoning applies, giving (unique) quadratic variations $A_t^{(M)}$ for each M. One then checks that $A_t^{(M)}(\omega) = A_t^{(M')}(\omega)$ for all $M' \geqslant M$ and all ω such that $\tau_M(\omega) > t$, and defines $A_t(\omega) := A_t^{(M)}(\omega)$ for such ω, and checks that this is the desired qvp. The details can be found in the Proof of Theorem 5.3 in [Liggett], which is also the model for the proof given here. □

(2.33) Theorem

Let (X_t) be a continuous, square integrable martingale, (A_t) its qvp, and σ, τ stopping times. Assume that $\sigma \leqslant \tau$, $\mathbb{P}(\tau < \infty) = 1$, and $\mathbb{E}(A_\tau) < \infty$. Then (with the convention $X_\infty := 0$) we have

a) $\mathbb{E}(X_\tau \,|\, \mathcal{F}_\sigma) = X_\sigma$ almost surely, and

b) $\mathbb{E}(X_\tau^2 - A_\tau \,|\, \mathcal{F}_\sigma) = X_\sigma^2 - A_\sigma$ almost surely.

In particular, $\mathbb{E}(X_\tau) = \mathbb{E}(X_0)$ and $\mathbb{E}(X_\tau^2) = \mathbb{E}(X_0^2) + \mathbb{E}(A_\tau)$.

Proof: Let $n \geqslant k$. Since $\mathcal{F}_{\tau\wedge n} \supset \mathcal{F}_{\tau\wedge k} \supset \mathcal{F}_{\sigma\wedge k}$, for any continuous martingale M we have

$$\mathbb{E}(M_{\tau\wedge n} \,|\, \mathcal{F}_{\sigma\wedge k}) = \mathbb{E}\Big(\mathbb{E}(M_{\tau\wedge n} \,|\, \mathcal{F}_{\tau\wedge k}) \Big| \mathcal{F}_{\sigma\wedge k}\Big) \overset{(2.29a)}{=} \mathbb{E}(M_{\tau\wedge k} \,|\, \mathcal{F}_{\sigma\wedge k}) \overset{(2.29b)}{=} M_{\sigma\wedge k} \qquad (*)$$

Since $\mathbb{P}(\sigma < \infty) = 1$ and the paths of M are continuous, the right hand side above always converges to $M_{\sigma\wedge k}\mathbb{1}_{\{\sigma<\infty\}}$ almost surely. What remains is to show

$$(**) \qquad \lim_{k\to\infty}\lim_{n\to\infty}\mathbb{E}(M_{\tau\wedge n}\,|\,\mathcal{F}_{\sigma\wedge k}) = \mathbb{E}(M_\tau\,|\,\mathcal{F}_\sigma) \qquad \text{almost surely,}$$

for the choices $M = X$ (giving (a)) and $M = X^2 - A$ (giving (b)).

For this, we set $\sigma = \tau$ in $(*)$ to see that $(M_{\tau\wedge n})_{n\in\mathbb{N}}$ is a discrete time martingale. In particular,

$$\mathbb{E}\Big(\big(X_{\tau\wedge n} - X_{\tau\wedge k}\big)^2\Big) = \mathbb{E}(X^2_{\tau\wedge n} - X^2_{\tau\wedge k}) = \mathbb{E}(A_{\tau\wedge n} - A_{\tau\wedge k}) = \mathbb{E}\big((A_{\tau\wedge n} - A_{\tau\wedge k})\mathbb{1}_{\{\tau\geqslant k\}}\big).$$

The first equality above holds for all square integrable martingales, the second is obtained by applying $(*)$ to $\tau = \sigma$ and $M_t = X_t^2 - A_t$, taking expectations and rearranging, and the third holds because on $\{\tau < k\}$ both terms in the difference are equal to A_k.

Since $t \mapsto A_t$ is increasing, we have $\sup_{n\geqslant k}|A_{\tau\wedge n} - A_{\tau\wedge k}| \leqslant A_\tau \in L^1$, and since $\mathbb{1}_{\{\tau\geqslant k\}} \to 0$ almost surely as $k\to\infty$, and thus

$$\lim_{k\to\infty}\sup_{n\geqslant k}\mathbb{E}\big((A_{\tau\wedge n} - A_{\tau\wedge k})\mathbb{1}_{\{\tau\geqslant k\}}\big) = 0$$

by dominated convergence. This implies that the sequence $(X_{\tau\wedge n})_{n\in\mathbb{N}}$ is an L^2-Cauchy sequence (and thus an L^1-Cauchy sequence), and since $A_{\tau\wedge n} - A_{\tau\wedge k} = |A_{\tau\wedge n} - A_{\tau\wedge k}|$, the sequence $(A_{\tau\wedge n})_{n\in\mathbb{N}}$ is an L^1-Cauchy sequence. Finally, since for any L^2-Cauchy sequence (f_n) the sequence (f_n^2) is an L^1-Cauchy sequence, also $(X^2_{\tau\wedge n})_{n\in\mathbb{N}}$ is an L^1-Cauchy sequence.

Since we already know that $M_{\tau\wedge n}\to M_\tau$ almost surely, we now have $\lim_{n\to\infty}M_{\tau\wedge n} = M_\tau$ in L^1 for both choices of M, and in particular $M_\tau \in L^1$. We know from Theorem (4.66) of the probability theory (PT) lecture that a discrete time martingale (like $(M_{\tau\wedge n})_{n\in\mathbb{N}}$) converging in L^1 to an integrable limit is uniformly integrable. By Theorem (4.68 a) of [PT], this implies

$$\mathbb{E}(M_\tau|\mathcal{F}_{\tau\wedge n}) = M_{\tau\wedge n} \qquad \text{almost surely for all } n\in\mathbb{N},$$

and thus for $k \leqslant n$,

$$\mathbb{E}\big(M_{\tau\wedge n}\big|\,\mathcal{F}_{\sigma\wedge k}\big) = \mathbb{E}\big(\mathbb{E}(M_\tau\,|\,\mathcal{F}_{\tau\wedge n})\,\big|\,\mathcal{F}_{\sigma\wedge k}\big) = \mathbb{E}(M_\tau\,|\,\mathcal{F}_{\sigma\wedge k}),$$

which allows us to take the $n\to\infty$ limit in $(**)$. By Theorem (4.68 b) of [PT], we have

$$\lim_{k\to\infty}\mathbb{E}(Y\,|\,\mathcal{G}_k) = \mathbb{E}(Y\,|\,\mathcal{G}_\infty) \qquad \text{in } L^1 \text{ and almost surely}$$

for any integrable random variable Y and any filtration $(\mathcal{G}_k)$, where $\mathcal{G}_\infty = \sigma(\mathcal{G}_k : k\in\mathbb{N})$. We put $Y = M_\tau$ and $\mathcal{G}_k = \mathcal{F}_{\tau\wedge k}$ to justify the $k\to\infty$ limit in $(*)$ (you should check that indeed $\mathcal{F}_\sigma = \sigma(\mathcal{F}_{\sigma\wedge k} : k\in\mathbb{N})$), finishing the proof of a) and b). The additional statement is obtained by taking $\sigma = 0$ and by taking expectations. □

(2.34) Corollary: Wald's identities

Let B be a BM[1], τ a stopping time. If $\mathbb{E}(\tau) < \infty$, then

$$\mathbb{E}(B_\tau = 0) \qquad \text{and} \qquad \mathbb{E}(B_\tau^2) = \mathbb{E}(\tau).$$

Proof: B_t is a continuous martingale with qvp $A_t = t$. The claim follows from (2.33) and the fact that $A_{\tau(\omega)}(\omega) = \tau(\omega)$ so that $\mathbb{E}(A_\tau) = \mathbb{E}(\tau)$. □

(2.35) Proposition

Let (X_t) be a continuous, real-valued martingale with $\mathbb{P}(X_0 = x) = 1$ for some $x \in \mathbb{R}$. For $a < x < b$, let $\tau := \inf\{t > 0 : X_t \notin (a, b)\}$, and assume $\mathbb{P}(\tau < \infty) = 1$. Then we have

$$\mathbb{P}(X_\tau = a) = \mathbb{P}(X \text{ hits } a \text{ before it hits } b) = \frac{b - x}{b - a},$$

and

$$\mathbb{P}(X_\tau = b) = \frac{x - a}{b - a}.$$

Proof: Let (A_t) be the qvp of X. Since $\tau \wedge n \leqslant n$, Theorem (2.29 b) gives $\mathbb{E}(X^2_{\tau\wedge n} - A_{\tau\wedge n}) = \mathbb{E}(X_0^2) = x^2$, thus $\mathbb{E}(X^2_{\tau\wedge n}) = \mathbb{E}(A_{\tau\wedge n}) + x^2$. Monotone convergence now gives

$$\mathbb{E}(A_\tau) = \lim_{n\to\infty} \mathbb{E}(A_{\tau\wedge n}) = \lim_{n\to\infty} \mathbb{E}(X^2_{\tau\wedge n}) + x^2 \leqslant \max\{a^2, b^2\} + x^2.$$

So, Theorem (2.33) applies, and we get

$$x = \mathbb{E}(X_0) = \mathbb{E}(X_\tau) = a\mathbb{P}(X_\tau = a) + b\mathbb{P}(X_\tau = b).$$

Together with the equality $\mathbb{P}(X_\tau = a) + \mathbb{P}(X_\tau = b) = 1$, we obtain the claim. □

Combined with Example (2.18 c), the previous proposition gives very interesting results about the long time behaviour of Brownian motion paths. To state them, we make the following

(2.36) Definition

Let B be a BM^d, $x \in \mathbb{R}^d$. The stochastic process $(B_t + x)_{t\geqslant 0}$ is called the d-dimensional Brownian motion started in x. Its path measure is denoted by $\mathbb{P}^x$.

(2.37) Theorem

Let $x \in \mathbb{R}^d$, B a BM^d started in x. For $r \in \mathbb{R}^+$, let $K(0, r) := \{y \in \mathbb{R}^d : |y| \leqslant r\}$. For $r < |x| < R$, we have

$$\mathbb{P}^x(\tau_{K(0,r)} < \tau_{K(0,R)^c}) = \begin{cases} \frac{R-|x|}{R-r} & \text{if } d = 1, \\ \frac{\ln R - \ln|x|}{\ln R - \ln r} & \text{if } d = 2, \\ \frac{R^{2-d} - |x|^{2-d}}{R^{2-d} - r^{2-d}} & \text{if } d \geqslant 3. \end{cases}$$

Proof: The case $d = 1$ was done in (2.35). For $d = 2$, note that the function $f : \mathbb{R}^2 \setminus \{0\} \to \mathbb{R}, x \mapsto \ln|x|$ is harmonic, as can be seen by computing

$$\partial_{x_i} f(x) = \frac{1}{|x|}\partial_{x_i}|x| = \frac{x_i}{|x|^2} \quad \text{and} \quad \partial^2_{x_i} f(x) = \frac{1}{|x|^2} - 2\frac{x_1}{|x|^3}\partial_{x_i}|x| = \frac{1}{|x|^2} - 2\frac{x_i^2}{|x|^4},$$

and adding for $i = 1, 2$. Now let $\tilde{f}$ be a bounded C^2-function with $\tilde{f}(x) = f(x)$ when $|x| > r$. One example would be the choice $f(x) = h(|x|)$ with

$$h(u) = \begin{cases} \ln u & \text{if } u > r, \\ au^3 + bu^2 + c & \text{if } u \leqslant r, \end{cases}$$

where a, b, c hare chosen such that h is C^2 at r. One can calculate that $a = -\frac{2}{3}r^{-3}$, $b = \frac{3}{2}r^{-2}$ and $c = \ln r - \frac{5}{6}$, but this is not so important. What matters is that now $\tilde{f}$ fulfils the assumptions of Theorem 2.16, so $(M_t^{\tilde{f}})$ with $M_t^{\tilde{f}} = \tilde{f}(B_t) - \int_0^t L\tilde{f}(B_s)\,\mathrm{d}s$ is a martingale. Let

$$\tau = \inf\{t \geqslant 0 : |B_t| \leqslant r \text{ or } |B_t| \geqslant R\}.$$

be the hitting time of the complement of the annulus with radii r and R. By the optional stopping theorem, also the process $M_{t\wedge\tau}^{\tilde{f}}$ is a martingale. On the other hand,

$$M_{t\wedge\tau}^{\tilde{f}} = \tilde{f}(B_{t\wedge\tau}) - \int_0^{t\wedge\tau} (L\tilde{f})(B_s)\,\mathrm{d}s = f(B_{t\wedge\tau}),$$

where the last inequality holds because $|B_{t\wedge\tau}(\omega)| \geqslant r$ for all ω, and thus in particular

$$(L\tilde{f})(B_s) = (L\tilde{f})(B_{s\wedge\tau}) = (Lf)(B_{s\wedge\tau}) = 0$$

for all $s \leqslant t \wedge \tau(\omega)$ and all ω. It follows that $(f(B_{t\wedge\tau}))$ is a (bounded) martingale starting in $x \in (r, R)$, and now (2.35) gives

$$\mathbb{P}^x(\tau_{K(0,r)} < \tau_{K(0,R)}) = \mathbb{P}(f(B_\tau) = \ln r) = \frac{\ln R - \ln|x|}{\ln R - \ln r},$$

as claimed. For $d \geqslant 3$, check that $f : \mathbb{R}^d \setminus \{0\} \to \mathbb{R}$, $x \mapsto |x|^{2-d}$ is harmonic, and the proceed as above. □

(2.38) Corollary

For BM^d starting in x with $|x| > r > 0$, we have

a) $\mathbb{P}^x(\tau_{\{0\}} < \infty) = \begin{cases} 1 & \text{if } d = 1, \\ 0 & \text{if } d \geqslant 2. \end{cases}$.

b) $\mathbb{P}^x(\tau_{K(0,r)} < \infty) = \begin{cases} 1 & \text{if } d \leqslant 2, \\ \left(\frac{|x|}{r}\right)^{2-d} & \text{if } d \geqslant 3. \end{cases}$

Proof: a) By continuity from above,

$$\mathbb{P}^x(\tau_{\{0\}} \leqslant \tau_{K(0,R)^c}) = \lim_{r\to 0} \mathbb{P}^x(\tau_{K(0,r)} \leqslant \tau_{K(0,R)^c}) = 0$$

for $d \geqslant 2$. By continuity from below, we can now take $R \to \infty$ and obtain the result, since $\tau_{K(0,R)^c} \to \infty$ almost surely as $R \to \infty$, e.g. by Doobs maximal inequality. The same reasoning works for $d = 1$ and is left as an exercise. For b), just use Theorem (2.37) and let $R \to \infty$ as above. □

(2.39) Remark

The interpretation of (2.38) is:
a) says that BM^d never „finds“ pre-determined points in $d \geqslant 2$; this is maybe not too surprising, it is a little bit like the fact that $\mathbb{P}(X = q) = 0$ if $X \sim \mathcal{U}([0, 1])$ for all $x \in [0, 1]$. Note however that for $d = 1$, things are different!
b) says that BM^d finds arbitrarily small balls from arbitrarily far away with probability 1 for

$d = 2$ (this *is* surprising), but has a (good) chance to never find them when $d \geqslant 3$.

It is tempting to claim that BM^2 finds all balls not only once, but infinitely often. The reason is that after the Brownian motion has visited a ball and left it again, we can „restart" it by using the invariances (2.2) and (2.3). In the same spirit, it looks like we should have $\mathbb{P}(\lim_{t\to\infty} |B_t| = \infty)$ for $d = 3$ (but not for $d \leqslant 2$). The reason is that any compact set is only visited finitely many times by part b) above: each time a Brownian motion leaves a compact set, we restart it, and then it has a positive probability of never finding it again. So eventually, it will succeed in never finding it again.

Both considerations are correct in principle, but suffer from the fact that the „restarting" of Brownian motion is a bit vague - we should specify where and when we restart, and investigate what happens then. For the correct way to handle such things, we need the Markov property of Brownian motion, which we treat next.

C) Markov Properties of Brownian Motion

(2.40) Definition

Let E be a metric space. For $s \geqslant 0$, the map

$$\theta_s : C(\mathbb{R}_0^+, E) \to C(\mathbb{R}_0^+, E), \quad (X_t)_{t\,\geqslant\,0} \mapsto (X_{t+s})_{t\,\geqslant\,0}$$

is called the *shift to the left* by s.

(2.41) Definition

Let

- $(\Omega, \mathcal{F})$ be a measurable space with a filtration $(\mathcal{F}_t)$,
- E be a metric space and $\mathcal{E}$ its Borel-σ-algebra,
- $(X_t)_{t\,\geqslant\,0}$ be a family of functions from Ω to E, such that each X_t is $\mathcal{F}_t$-$\mathcal{E}$-measurable,
- $(\mathbb{P}^x)_{x\in E}$ be a family of probability measures on $(\Omega, \mathcal{F})$ such that $\mathbb{P}^x(X_0 = x) = 1$ for all $x \in \mathbb{E}$.

Note that for each $\mathbb{P}^x$, the measurable maps (X_t) form a stochastic process on $(\Omega, \mathcal{F}, \mathbb{P}^x)$ with values in E, and for all x these stochastic processes may be different even though the maps X_t are always the same. The collection $((X_t)_{t\,\geqslant\,0}, (\mathbb{P}^x)_{x\in E})$ of all such pairs is called

a) **a weak Markov process** if for all $f \in C_b(E, \mathbb{R})$, all $x \in E$ and all $s \geqslant 0$, we have

$$\mathbb{E}^x(f(X_{t+s}) \,|\, \mathcal{F}_s)(\bar\omega) = \mathbb{E}^{X_s(\bar\omega)}(f(X_t)) \qquad \text{for } \mathbb{P}^x\text{-almost all } \bar\omega \in \Omega.$$

b) **a Markov process** if for all measurable, bounded $F : (E^{\mathbb{R}_0^+}, \mathcal{E}^{\otimes \mathbb{R}_0^+}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, all $x \in E$ and all $s \geqslant 0$, we have

$$\mathbb{E}^x\Big(F \circ \theta_s\big((X_t)_{t\,\geqslant\,0}\big) \,\Big|\, \mathcal{F}_s\Big)(\bar\omega) \equiv \mathbb{E}^x\Big(F\big((X_{t+s})_{t\,\geqslant\,0}\big) \,\Big|\, \mathcal{F}_s\Big)(\bar\omega) = \mathbb{E}^{X_s(\bar\omega)}(F((X_t)_{t\,\geqslant\,0})),$$

for $\mathbb{P}^x$-almost all $\bar\omega \in \Omega$.

c) **a strong Markov process** if for all $\mathcal{B}(\mathbb{R}) \otimes \mathcal{E}^{\otimes \mathbb{R}_0^+} - \mathcal{B}(\mathbb{R})$-measurable, bounded functions

$F : \mathbb{R}_0^+ \times \mathbb{E}^{\mathbb{R}_0^+} \to \mathbb{R}$, all $x \in E$ and all $(\mathcal{F}_t)$-stopping times τ, we have

$$\mathbb{E}^x\Big(F\big(\tau, (\theta_\tau X)_t\big) \,\Big|\, \mathcal{F}_\tau\Big)(\bar\omega) = \mathbb{E}^{X_{\tau(\bar\omega)}(\bar\omega)}\Big(F\big(\tau(\bar\omega), (X_t)_{t \geqslant 0}\big)\Big),$$

for $\mathbb{P}^x$-almost all $\bar\omega \in \{\tau < \infty\}$.

(2.42) Remark

Observe how the definition gives an intricate connection between the (usually uncountably many) probability measures $\mathbb{P}^x$ in terms of how the stochastic processes $((X_t), \mathbb{P}^x)$ behave under conditioning. The best way to understand this definition is to take $(\Omega, \mathcal{F}) = (E^{\mathbb{R}+0^+}, \mathcal{E}^{\otimes \mathbb{R}_0^+})$, $X_t(\omega) = \pi_t(\omega)$ and $\mathcal{F}_t = \sigma(X_s : s \leqslant t)$. Then the different probability measures $\mathbb{P}^x$ describe the random behaviour of the process when started at x. Note however that the maps $\omega \mapsto X_t(\omega)$ are always the same, it is the $\mathbb{P}^x$ that change!

a) (2.41 a) means that if we know what (X_t) did up to time s and want to know the probability of if (e.g.) being in a set A at time $t+s$, then all we need to know is the point $X_s(\bar\omega)$ to which it arrived at time s; the information about things that the process did before time s has no influence on the answer. Moreover, the answer can be computed by asking what the probability of a process started (at time 0) in the point $X_s(\bar\omega)$ is to be in the set A, i.e. we switch probability measures depending on where we find the process (and decrease the run-time from $t+s$ to t). Compare also with Proposition (2.3).

b) (2.41 b) is an improved version of a): instead of just answering questions about the (conditional) behaviour of X at *one* time point in the future by switching probability measures and shifting time, we can do the same for questions about the whole *path* after time s. Examples for the function F might be integrals, e.g. $F(X) = \int_0^7 \sin(X_r)\,\mathrm{d}r$. Then $F \circ \theta_s(X) = \int_0^7 \sin(X_{r+3})\,\mathrm{d}r = \int_3^{10} \sin(X_r)\,\mathrm{d}r$. Another example would be $F(X) = \max_{r \leqslant 5} |X(r)|$.

c) (2.41 c) means that we can even replace s by a stopping time in the context of b), i.e. the time when we stop may depend on (usually) what the process has done in the past. An example of what we can achieve with this is the following calculation, for $E = \mathbb{R}$, $r < R$ and under the assumption that X is a strong Markov process with continuous paths, and that with $\tau_R = \inf\{s \geqslant 0 : X_s \geqslant R\}$ we have $\mathbb{P}^0(\tau_R < \infty) = 1$:

$$\begin{aligned} &\mathbb{P}^0(X_t \text{ returns to } K(0,r) \text{ after having been larger than } R) = \mathbb{E}^0(\mathbb{1}_{\{X_{t+\tau_R} : t \geqslant 0\} \cap K(0,r) \neq \emptyset\}}) = \\ &= \mathbb{E}^0\Big(\mathbb{E}^0(\mathbb{1}_{\{X_{t+\tau_R} : t \geqslant 0\} \cap K(0,r) \neq \emptyset\}} \,|\, \mathcal{F}_{\tau_R})\Big) \overset{(2.41c)}{=} \mathbb{E}^0\Big(\mathbb{E}^{X_{\tau_R(.)}(.)}(\mathbb{1}_{\{X_t : t \geqslant 0\} \cap K(0,r) \neq \emptyset\}})\Big) = \\ &= \mathbb{E}^R(\mathbb{1}_{\{X_t : t \geqslant 0\} \cap K(0,r) \neq \emptyset\}}) = \mathbb{P}^R(X \text{ ever hits } K(0,r)). \end{aligned}$$

The equality in the last line is because by path continuity, we have $X_{\tau_R(\omega)}(\omega) = R$ for all $\omega \in \{\tau_R < \infty\}$. This is precisely what we were aiming for in the last part of Remark (2.39).

d) In many textbooks, the strong Markov property is written in the „classical probability theory“ way, i.e. without writing the arguments $\bar\omega$ in the conditional expectation. Here is one of the places where this classical way is really inadequate and creates a lot of confusion: there is then simply no way to write that in the first slot of the function F on the right hand side, the τ must be given the (fixed) $\bar\omega$, while the $(X_t)_{t \geqslant 0}$ is integrated over. In integral notation, the right hand side actually reads

$$\int F(\tau(\bar{\omega}), (X_t(\omega))_{t \geqslant 0}) \, \mathbb{P}^{X_{\tau(\bar{\omega})}(\bar{\omega})}(\omega),$$

while the integrand (inside the conditional expectation) on the left hand side ist given by $F(\tau(\omega), (X_{t+\tau(\omega)}(\omega))_{t \geqslant 0})$. Notice the precise places where $\bar{\omega}$ and ω are used, respectively!

In many important cases, all three notions of Markov processes coincide. We prepare this statement by an important variant of the π-λ-Theorem from probability theory.

(2.43) Monotone Class Theorem

Let $\mathcal{F}$ be a σ-algebra on a set Ω, $\mathcal{A} \subset \mathcal{F}$ be a π-system (i.e.: $A, B \in \mathcal{A}$ implies $A \cap B \in \mathcal{A}$), and let $\mathcal{H}$ be a collection of $\mathcal{F}$-measurable functions with the following properties:

(i): $\Omega \in \mathcal{A}$, and $1_A \in \mathcal{H}$ for all $A \in \mathcal{A}$.

(ii): $\mathcal{H}$ is a vector space, i.e. $\alpha f + g \in \mathcal{H}$ whenever $f, g \in \mathcal{H}$ and $\alpha \in \mathbb{R}$.

(iii): $\mathcal{H}$ is closed under monotone limits, i.e. if $f_n \in \mathcal{H}$ for all n, $f_n(\omega) \leqslant f_{n+1}(\omega)$ for all $\omega \in \Omega$ and all $n \in \mathbb{N}$, and $\lim_{n\to\infty} f_n = f$ for some bounded function f, then $f \in \mathcal{H}$ follows.

Under these conditions, $\mathcal{H}$ contains all bounded, $\sigma(\mathcal{A})$-measurable functions.

Proof: exercise using the π-λ-Theorem. □

(2.44) Theorem

Consider the situation as in Definition (2.41), and assume in addition that $(\mathcal{F}_t)$ is the right-continuous filtration generated by the maps X_t, i.e. $\mathcal{F}_t = \bigcap_{n\in\mathbb{N}} \sigma(X_s : s \leqslant t + 1/n)$, and that $\mathcal{F} = \sigma(\mathcal{F}_t : t \geqslant 0)$. Let $((X_t), (\mathbb{P}^x))$ be a weak Markov process.

a) Assume in addition that for all $0 < t_1 < t_2 \cdots < t_n$ and all $f_1, \dots f_n \in C_b(E, \mathbb{R})$, the map

$$x \mapsto \mathbb{E}^x(f_1(X_{t_1}) \cdots f_n(X_{t_n})) \quad \text{is continuous.} \qquad \text{(i)}$$

Then $((X_t), (\mathbb{P}^x))$ is a Markov process.

b) Assume that (i) holds, and in addition assume that

$$\forall \omega \in \Omega : t \mapsto X_t(\omega) \quad \text{is right-continuous.} \qquad \text{(ii)}$$

Then $((X_t), (\mathbb{P}^x))$ is a strong Markov process.

Proof:

a) We first show by induction that (i) implies that for all $n \in \mathbb{N}$, all $f_1, \dots, f_n \in C_b(E, \mathbb{R})$ and all $0 < t_1 < \dots t_n$, we have

$$\mathbb{E}^x(f_1(X_{t_1+s}) \cdots f_n(X_{t_n+s}) \,|\, \mathcal{F}_s)(\bar{\omega}) = \mathbb{E}^{X_s(\bar{\omega})}(f_1(X_{t_1}) \cdots f_n(X_{t_n})) \qquad (*).$$

For $n = 1$, this is just the weak Markov property. Assuming it holds up to $n-1$, we have

$$\mathbb{E}^x\Big(\prod_{i=1}^n f_i(X_{t_i+s})\,\Big|\,\mathcal{F}_s\Big)(\bar\omega) \overset{\text{Tower}}{=} \mathbb{E}^x\Big(\mathbb{E}^x\big(\prod_{i=1}^n f_i(X_{t_i+s})\,|\,\mathcal{F}_{t_1+s}\big)\,\Big|\,\mathcal{F}_s\Big)(\bar\omega) =$$

$$= \mathbb{E}^x\Big(f_1(X_{t_1+s})\mathbb{E}^x\big(\prod_{i=2}^n f_i(X_{t_i+s})\,|\,\mathcal{F}_{t_1+s}\big)\,\Big|\,\mathcal{F}_s\Big)(\bar\omega) = \qquad \text{(induction hypothesis)}$$

$$= \mathbb{E}^x\Big(f_1(X_{t_1+s})\mathbb{E}^{X_{t_1+s}}\big(\prod_{i=2}^n f_i(X_{t_i-t_1})\big)\,\Big|\,\mathcal{F}_s\Big)(\bar\omega) = \qquad \text{(weak Markov, continuity of } x\mapsto\mathbb{E}^x(\cdots))$$

$$= \mathbb{E}^{X_s(\bar\omega)}\Big(f_1(X_{t_1})\mathbb{E}^{X_{t_1}}\big(\prod_{i=2}^n f_i(X_{t_i-t_1})\big)\Big)(\bar\omega) = (**)$$

By the induction hypothesis, for almost all $\omega\in\Omega$ and for the (fixed) $\bar\omega$ from above, we have

$$\mathbb{E}^{X_s(\bar\omega)}\Big(\prod_{i=2}^n f_i(X_{t_i})\,\Big|\,\mathcal{F}_{t_1}\Big)(\omega) = \mathbb{E}^{X_{t_1}(\omega)}\Big(\prod_{i=2}^n f_i(X_{t_i-t_1})\Big),$$

We read this from right to left, plug it into $(**)$ and use the tower property (with respect to the measure $\mathbb{E}^{X_s(\bar\omega)}$) in order to prove $(*)$.

The next step is to extend $(*)$ to the case when the f_i are indicators of open subsets of E. This is done by the monotone convergence theorem on both sides, by approximating indicators of open sets U by continuous functions from below: $f_n(x) = \min\{1, nd(x,U^c)\}$, where d is the metric. We then have shown the validity of the defining equation

$$\mathbb{E}^x\Big(F\circ\theta_s(X)\,\Big|\,\mathcal{F}_s\Big)(\bar\omega) = \mathbb{E}^{X_s(\bar\omega)}(F(X)) \qquad (***)$$

for a Markov process in the case where $F(X) = \prod_{i=1}^n \mathbb{1}_{A_i}(X_{t_i})$ for open sets $A_i\subset E$, $t_1,\dots t_n\geqslant 0$, and $n\in\mathbb{N}$. Thus letting

$$\mathcal{H} = \{F:\Omega\to\mathbb{R} \text{ bounded, measurable, such that } (***) \text{ holds}\},$$

we first see that $\mathcal{H}$ is closed under monotone limits (this follows from monotone convergence for conditional expectations and ordinary monotone convergence), is a vector space by the linearity of (conditional) expectation, and contains the indicator functions for the set system

$$\mathcal{A} = \{X_{t_1}^{-1}(A_1)\cap\cdots\cap X_{t_n}^{-1}(A_n) : t_i\geqslant 0, A_i\subset E \text{ open } \forall i\leqslant n, n\in\mathbb{N}\}.$$

Since $\sigma(\mathcal{A}) = \mathcal{F}$ by assumption, the monotone class theorem now shows that $(***)$ holds for all bounded, measurable $\mathcal{F}$, which is the Markov property.

b) Consider first functions of the type

$$F(t,X) = f_1(t,X_{t_1})\cdots f_n(t,X_{t_n}) \qquad \text{for } f_i\in C_b(\mathbb{R}_0^+\times E,\mathbb{R}), 0<t_1<\ldots<t_n, n\in\mathbb{N}.$$

For such F, the map $Z:\bar\omega\mapsto\mathbb{E}^{X_{\tau(\bar\omega)}(\bar\omega)}(F(\tau(\bar\omega),X))$ is $\mathcal{F}_\tau$-measurable by ((2.28) a,e) and by the continuity of the map $x\mapsto\mathbb{E}^x\big(\prod_{i=1}^n f_i(\tau(\bar\omega),X_{t_i})\big)$. So Z already has the right measruability for being the conditional expectation $\mathbb{E}^x(F(\tau,\theta_\tau X)\,|\,\mathcal{F}_\tau)$. To prove that it really is the conditional expectation we still need to check the defining equality for conditional expectation, which in general reads $\mathbb{E}(\mathbb{E}(Y\,|\,\mathcal{G})\mathbb{1}_G) = \mathbb{E}(Y\mathbb{1}_G)$ for all $G \in \mathcal{G}$, and in our particular case is best written as

$$\forall A \in \mathcal{F}_\tau, A \subset \{\tau < \infty\} : \int \mathbb{E}^{X_{\tau(\bar\omega)}(\bar\omega)}\big(F(\tau(\bar\omega), X)\big)\mathbb{1}_A(\bar\omega)\mathbb{P}^x(\mathrm{d}\bar\omega) = \mathbb{E}^x(F(\tau, \theta_\tau X)\mathbb{1}_A) \qquad (*^4)$$

Let us first assume that τ only takes countably many values, i.e. $\tau(\Omega) = \{s_i : i \in \mathbb{N}\} \cup \{\infty\}$ with $s_i \in \mathbb{R}_0^+$. Then for $A \in \mathcal{F}_\tau$ with $A \subset \{\tau < \infty\}$,

$$\begin{aligned}\mathbb{E}^x(F(\tau, \theta_\tau X)\mathbb{1}_A) &= \sum_{i\in\mathbb{N}} \mathbb{E}^x\big(F(s_i, \theta_{s_i} X)\underbrace{\mathbb{1}_{A\cap\{\tau=s_i\}}}_{\in \mathrm{m}\mathcal{F}_{s_i}}\big) \overset{a)}{=} \sum_{i\in\mathbb{N}} \mathbb{E}^x\Big(\mathbb{E}^{X_{s_i}}(F(s_i, X))\mathbb{1}_{A\cap\{\tau=s_i\}}\Big) = \\ &= \sum_{i\in\mathbb{N}} \int \mathbb{E}^{X_{\tau(\bar\omega)}(\bar\omega)}\big(F(\tau(\bar\omega), X)\big)\mathbb{1}_{A\cap\{\tau=s_i\}}(\bar\omega)\,\mathbb{P}^x(\mathrm{d}\bar\omega) = \\ &= \int \mathbb{E}^{X_{\tau(\bar\omega)}(\bar\omega)}\big(F(\tau(\bar\omega), X)\big)\mathbb{1}_A(\bar\omega)\,\mathbb{P}^x(\mathrm{d}\bar\omega),\end{aligned}$$

so $(*^4)$ holds for discrete τ. For general τ, use Lemma (2.25) to approximate τ from above by discrete $(\tau_k)_{k\in\mathbb{N}}$. By the right-continuity of paths of X and the continuity of the f_i, we have

$$\begin{aligned}\lim_{k\to\infty} F(\tau_k(\bar\omega), \theta_{\tau_k(\omega)} X(\omega)) &= \lim_{k\to\infty} \prod_{i=1}^n f_i(\tau_k(\bar\omega), X_{t_i+\tau_k(\omega)}) = \\ &= \prod_{i=1}^n f_i(\tau(\bar\omega), X_{t_i+\tau(\omega)}) = F(\tau(\bar\omega), \theta_{\tau(\omega)} X(\omega)).\end{aligned}$$

If we take $\bar\omega = \omega$ and use dominated convergence, we find

$$\lim_{k\to\infty} \mathbb{E}^x(F(\tau_k, \theta_{\tau_k} X)\mathbb{1}_A) = \mathbb{E}^x(F(\tau, \theta_\tau X)\mathbb{1}_A).$$

By keeping $\bar\omega$ and ω different, using of the triangle inequality (to separate the k-dependence of the upper index below from the k-dependence of the function that is integrated), recalling assumption (i) and using dominated convergence we get

$$\begin{aligned}\lim_{k\to\infty} \mathbb{E}^{X_{\tau_k(\bar\omega)}(\bar\omega)}((\big(F(\tau_k(\bar\omega), X)) &= \lim_{k\to\infty} \mathbb{E}^{X_{\tau_k(\bar\omega)}(\bar\omega)}\Big(\prod_{i=1}^n f_i(\tau_k(\bar\omega), X_{t_i+\tau_k}))\Big) = \\ &= \mathbb{E}^{X_{\tau(\bar\omega)}(\bar\omega)}\Big(\prod_{i=1}^n f_i(\tau(\bar\omega), X_{t_i+\tau}))\Big) = \mathbb{E}^{X_{\tau(\bar\omega)}(\bar\omega)}\big(F(\tau(\bar\omega), X)\big)\end{aligned}$$

for all $\bar\omega$. We multiply this by $\mathbb{1}_A(\bar\omega)$, integrate using to $\mathbb{P}^x$ and employ dominated convergence once again to see that $(*^4)$ holds for general stopping times, too.

The final step is to allow general bounded, $\mathcal{B}(\mathbb{R}_0^+)\otimes\mathcal{F}$-measurable functions $F : \mathbb{R}_0^+ \times E^{\mathbb{R}_0^+} \to \mathbb{R}$. Here we use again the monotone class theorem: fist of all, just like in a) we use monotone convergence to get $(*^4)$ for the case where $F(t, X) = \prod_{i=1}^n \mathbb{1}_{A_i}(t, X_{t_i})$ with $A_i \subset \mathbb{R}_0^+ \times E$ open in the product topology, and $t_1, \dots, t_n \geqslant 0$. The set system

$$\mathcal{A} = \{(t, \omega) \in \mathbb{R}_0^+ \times \Omega : (t, X_{t_1}(\omega)) \in A_1, \dots, (t, X_{t_n}(\omega)) \in A_n, t_i \geqslant 0, A_i \subset \mathbb{R}_0^+ \times E \text{ open } \forall i \leqslant n, n \in \mathbb{N}\}$$

generates $\mathcal{B}(\mathbb{R}_0^+) \otimes \mathcal{F}$, and the set of functions

$$\mathcal{H} = \{F : \mathbb{R}_0^+ \times \Omega \to \mathbb{R} \text{ bounded, measurable such that } (*^4) \text{ holds for all } A \in \mathcal{F}_\tau, A \subset \{\tau < \infty\}\}$$

is closed under monotone limits (again monotone convergence!), a vector space, and contains the indicators for all sets in $\mathcal{A}$ as we have proved above. This shows the claim. □

(2.45) Corollary

Define $\mathbb{P}^x$, $x \in \mathbb{R}^d$, as in (2.36), and let (B_t) be a BM^d with its right-continuous filtration. Then $((B_t), (\mathbb{P}^x))$ is a strong Markov process.

Proof: By (2.2) and (2.3), the maps $(W_t) = (B_{s,t})_{t \geqslant s}$ and $(B_r)_{r \leqslant s}$ are independent Brownian motions under $\mathbb{P}^0$. Moreover, (W_t) has the same distribution under all $\mathbb{P}^x$, i.e. is a Brownian motion starting in zero under all of these measures. Therefore

$$\begin{aligned}\mathbb{E}^x\big(F((B_{t-s})_{t \geqslant 0}) \,\big|\, \mathcal{F}_s\big)(\bar\omega) &= \mathbb{E}^x\big(F((W_t)_{t \geqslant s} + B_s) \,\big|\, \mathcal{F}_s\big)(\bar\omega) \\ &\overset{(*)}{=} \mathbb{E}^x\big(F((W_t)_{t \geqslant s} + B_s(\bar\omega))\big) = \mathbb{E}^0\big(F((W_t)_{t \geqslant s} + B_s(\bar\omega))\big) \\ &= \mathbb{E}^0\big(F((B_t)_{t \geqslant 0} + B_s(\bar\omega))\big) = \mathbb{E}^{B_s(\bar\omega)}\big(F((B_t)_{t \geqslant 0})\big),\end{aligned}$$

which is the regular Markov property. The equality $(*)$ is a special case of the fact that $\mathbb{E}(f(X,Y) \,|\, \mathcal{F})(\bar\omega) = \mathbb{E}(f(X, Y(\bar\omega)))$ in the case when $X \perp\!\!\!\perp \mathcal{F}$ and $Y \in \mathrm{m}\mathcal{F}$. The strong Markov property now follows from path continuity and Theorem (2.44). □

The strong Markov property is very powerful. In the remainder of this chapter, we give several examples where it (or the regular Markov property) is used to extract very interesting information about Brownian motion. We start by completing what we began in Remark (2.39):

(2.46) Theorem: transience and recurrence of Brownian Motion

a) BM^1 is **point recurrent**, i.e. $\mathbb{P}^x(B_t = 0 \text{ infinitely often}) = 1$ for all $x \in \mathbb{R}$.

b) BM^2 is **neighbourhood recurrent**. i.e.

$$\forall r > 0, x \in \mathbb{R}^2 : \mathbb{P}^x(B_t \text{ returns to } K(0,r) \text{ infinitely often}) = 1.$$

Here, „returns to" means that there are two time $t_1 < s_2 < t_2$ so that $B_{t_1}, B_{t_2} \in K(0,r)$, but $B_{s_2} \notin K(0,r)$.

c) BM^d is **transient** for $d \geqslant 3$, i.e.

$$\forall x \in \mathbb{R}^d : \mathbb{P}^x(\lim_{t \to \infty} |B_t| = \infty) = 1.$$

Proof:

a) holds by (2.8) and path continuity.

b) Let $R > r$ and define $\sigma_r = \inf\{s > 0 : |B_s| \leqslant r\}$, $\tau_{R,0} := \inf\{s > 0 : B_s \geqslant R\}$, and

$$\tau_{R,n} := \inf\{t > 0 : \exists s_1 < t_1 < s_2 < t_2 < \dots < s_n < t_n < t : |B_{s_i}| < r, |B_{t_i}| > R \;\forall i \leqslant n\}.$$

By (2.38 b), $\mathbb{P}^y(\sigma_r < \infty) = 1$ for all $y \in \mathbb{R}^2$, and by (2.8) (e.g. for the first component), $\mathbb{P}^y(\tau_{R,0} < \infty) = 1$ for all $y \in \mathbb{R}$. The claim will be shown once we prove the same for $\tau_{R,k}$ for all k. Assume we have it for some k, i.e. we know that $\mathbb{P}^y(\tau_{R,k} < \infty) = 1$ for all x. Then by the strong Markov property (twice), we get

$$\begin{aligned}
\mathbb{P}^x(\tau_{R,k+1} < \infty) &= \mathbb{E}^x\Big(\mathbb{E}^x\big(\mathbb{1}_{\{\exists 0<s<t: |X_{\tau_{R,k}+s}|<r, |X_{\tau_{R,k}+t}|>R\}} \,\big|\, \mathcal{F}_{\tau_{R,k}}\big)\Big) = \\
&= \mathbb{E}^x\Big(\mathbb{E}^{X_{\tau_{R,k}}}\big(\underbrace{\mathbb{1}_{\{\sigma_r<\infty\}}}_{=1 \text{ almost surely}} \mathbb{1}_{\{\sup_{t\geqslant 0}|B_{\sigma_r+t}|>R\}}\big)\Big) = \\
&= \mathbb{E}^x\Big(\mathbb{E}^{X_{\tau_{R,k}}}\big(\mathbb{E}^{X_{\tau_{R,k}}}(\mathbb{1}_{\{\sup_{t\geqslant 0}|B_{\sigma_r+t}|>R\}} \,|\, \mathcal{F}_{\sigma_r})\big)\Big) = \\
&= \mathbb{E}^x\Big(\mathbb{E}^{X_{\tau_{R,k}}}\big(\underbrace{\mathbb{P}^{X_{\sigma_r}}(\sup_{t\geqslant 0}|B_t| > R)}_{=1}\big)\Big) = 1,
\end{aligned}$$

which completes the induction.

c) exercise, similar to b). ☐

The most famous application of the strong Markov property is

(2.47) Theorem: the reflection principle

Let B be a BM[1], and $M_t(\omega) := \max\{B_s(\omega) : s \leqslant t\}$. For all $0 < b < a$, $t > 0$, we have

$$\mathbb{P}^0(M_t > a, B_t < b) = \mathbb{P}^0(B_t > 2a - b).$$

Proof: Let us first give an informal proof. Pick $\omega \in \Omega$ with $M_t(\omega) > a$ and $B_t(\omega) < b$, i.e. an ω that contributes to the first probability. Now we take the part of the path $s \mapsto B_s(\omega)$ that comes after the time $\tau_a(\omega)$ when the path first hits of a (which must have happened since $M_t(\omega) > a$), and „reflect“ it. This means we now consider the path

$$\tilde{B}_s(\omega) = \begin{cases} B_s(\omega) = a + (B_s(\omega) - a) & \text{if } s \leqslant \tau_a(\omega) \\ 2a - B_s(\omega) = a - (B_s(\omega) - a) & \text{if } \tau_a(\omega) \leqslant s \leqslant t. \end{cases}$$

Then $\tilde{B}_t(\omega) > 2a - b$ because $B_t(\omega) < b$. Therefore, the set of paths making up the second probability emerges from those that make up the first probability by reflection, and there as a bijective map between the two sets. By the strong Markov property, each reflected path has the same („infinitesimal“) probability as the original path: there is the (infinitesimal) probability p_1 of getting to a in time τ_a in the first place, at which time the path is „restarted“ at a, and the difference between $\tilde{B}$ and B is just that in one case the negative of the reflected path is taken, but this has the same probability p_2 by orthogonal invariance (2.1). Since the part after τ_a is independent of how we got there by the strong Markov property, the infinitesimal probability of both $B(\omega)$ and $\tilde{B}(\omega)$ is just $p_1 p_2$. Of course, the problem with this proof is the use of the suggestive but ill-defined notion of „infinitesimal probability“, which is why we will now give the formal proof.

For this, let

$$\tau_a(\omega) := \begin{cases} \inf\{s \geqslant 0 : B_s \geqslant a\} & \text{if } M_t \geqslant a, \\ \infty & \text{if } M_t < a. \end{cases}$$

This is a stopping time. We also define

$$F(s,\omega) := \mathbb{1}_{\{s<t\}}\big(\mathbb{1}_{\{B_{t-s}(\omega)>2a-b\}} - \mathbb{1}_{\{B_{t-s}(\omega)<b\}}\big),$$

and observe that for fixed $s < t$, we have

$$\mathbb{E}^a(F(s,\cdot)) = \mathbb{P}^a(B_{t-s} > 2a-b) - \mathbb{P}^a(B_{t-s} < b) = \mathbb{P}^0(B_{t-s} > a-b) - \mathbb{P}^0(B_{t-s} < b-a) = 0,$$

by symmetry (i.e. by (2.1)), while for $s \geqslant t$ we have $F(s,.) = 0$ by definition. Therefore

$$\mathbb{E}^{X_{\tau_a(\bar\omega)}(\bar\omega)}(F(\tau_a(\bar\omega),.))\mathbb{1}_{\{\tau_a(\bar\omega)<\infty\}} = 0 \qquad \forall \bar\omega \in \Omega.$$

Together with the fact that $\{\tau_a < \infty\} = \{M_t \geqslant a\}$, we get

$$\begin{aligned}
0 &= \int \mathbb{E}^{X_{\tau_a(\bar\omega)}(\bar\omega)}(F(\tau_a(\bar\omega),.))\mathbb{1}_{\{\tau_a(\bar\omega)<\infty\}}\,\mathbb{P}^0(\mathrm{d}\bar\omega) \overset{(2.41c)}{=} \\
&= \int \mathbb{E}^0\big(F(\tau_a,\theta_{\tau_a}B)\,\big|\,\mathcal{F}_{\tau_a}\big)(\bar\omega)\mathbb{1}_{\{\tau_a(\bar\omega)<\infty\}}\,\mathbb{P}_0(\mathrm{d}\bar\omega) \overset{\text{tower}}{=} \\
&= \mathbb{E}^0\big(F(\tau_a,\theta_{\tau_a}B)\mathbb{1}_{\{\tau_a<\infty\}}\big) = \mathbb{E}^0\big(\mathbb{1}_{\{M_t\,\geqslant\,a\}}(\mathbb{1}_{\{B_{t-\tau_a+\tau_a}>2a-b\}} - \mathbb{1}_{\{B_{t-\tau_a+\tau_a}><b\}})\big) = \\
&= \mathbb{P}^0(M_t > a, B_t > 2a-b) - \mathbb{P}^0(M_t > a, B_t < b) = \mathbb{P}^0(B_t > 2a-b) - \mathbb{P}^0(M_t > a, B_t < b).
\end{aligned}$$

□

The important point about Theorem 2.47 is that for many stochastic processes, we have some idea what the distribution of the process X_t at some time t is (for BM it is Gaussian), but the maximum M_t up to time t is often a complicated object, where we know little about the distribution. For Brownian motion, however, the theorem shows that M_t is also not too difficult. For example:

(2.48) Corollary

Let B be one-dimensional Brownian motion, $a > 0$ and $t > 0$, and define $\tau_a = \inf\{s \geqslant 0 : B_s \geqslant a\}$. Then

a) $\mathbb{P}^0(\tau_a < t) = \mathbb{P}^0(M_t > a) = \mathbb{P}(|B_t| > a)$, i.e. $M_t \sim |B_t|$ for all t.

b) $\mathbb{P}^0(M_t - B_t > a) = \mathbb{P}^0(|B_t| > a)$ for all t.

Proof: a) The first equality holds by definition of τ_a and M_t. For the second, take $b \nearrow a$ in Theorem (2.47), and use that

$$\mathbb{P}^0(M_t > a) = \mathbb{P}^0(B_t > a) + \mathbb{P}^0(M_t > a, B_t < a) \overset{(2.47)}{=} 2\mathbb{P}^0(B_t > a) \overset{(2.1)}{=} \mathbb{P}^0(|B_t| > a).$$

b) We have

$$M_t - B_t = \sup_{s\,\leqslant\,t}(B_s - B_t) = \sup_{s\,\leqslant\,t}(B_{t-s} - B_t) \overset{(2.5)}{\sim} \sup_{s\,\leqslant\,t} B_s = M_t \overset{a)}{\sim} |B_t|.$$

□

Remark: The equality b) in (2.48) actually holds for the full distribution of the processes: we have $(M_t - B_t)_{t\,\geqslant\,0} \sim (|B_t|_{t\,\geqslant\,0})$. For the proof, see Remark 1.75 and after in the book of Liggett.

(2.49) The arcsine law

Let B be a one-dimensional Brownian motion, and define $\xi_t := \sup\{s < t : B_s = 0\}$ (note that this is *not* a stopping time!). Then we have

$$\mathbb{P}^0(\xi_t < s) = \frac{2}{\pi} \arcsin \sqrt{s/t} \qquad \forall 0 \leqslant s \leqslant t.$$

Proof: For $s < t$, let $h(s) = \mathbb{P}^0(\xi_t < s)$. Then we have

$$\begin{aligned} h(s) &= \mathbb{P}^0(B_0 \neq 0 \ \forall u \in [s,t]) = \mathbb{E}^0\big(\mathbb{P}^0(B_0 \neq 0 \ \forall u \in [s,t] \,|\, \mathcal{F}_s)\big) \overset{(2.41b)}{=} \\ &= \mathbb{E}^0\big(\mathbb{P}^{B_s(.)}(B_u \neq 0 \ \forall u \in [0, t-s])\big) = \\ &= \mathbb{E}^0\big(\mathbb{1}_{\{B_s \geqslant 0\}} \mathbb{P}^{B_s(.)}(B_u > 0 \ \forall u \leqslant t-s)\big) + \mathbb{E}^0\big(\mathbb{1}_{\{B_s < 0\}} \mathbb{P}^{B_s(.)}(B_u < 0 \ \forall u \leqslant t-s)\big) \overset{(2.1)}{=} \\ &= 2\mathbb{E}^0(\mathbb{1}_{\{B_s \geqslant 0\}} \mathbb{P}^{B_s(.)}(B_u > 0 \ \forall u \leqslant t-s)\big) = 2\int_0^\infty \mathrm{d}x \, \frac{1}{\sqrt{2\pi s}} \, \mathrm{e}^{-\frac{x^2}{2s}} \, \mathbb{P}^x(B_u > 0 \ \forall u \leqslant t-s). \end{aligned}$$

Now

$$\begin{aligned} \mathbb{P}^x(B_u > 0 \ \forall u \leqslant t-s) &= \mathbb{P}^0(-M_{t-s} > -x) = \mathbb{P}^0(M_{t-s} < x) \overset{(2.48a)}{=} \\ &= \mathbb{P}^0(|B_{t-s}| < x) = 2\mathbb{P}^0(0 < B_{t-s} < x), \end{aligned}$$

and thus

$$h(s) = 2\int_0^\infty \mathrm{d}x \, \frac{1}{\sqrt{2\pi s}} \, \mathrm{e}^{-\frac{x^2}{2s}} \int_0^x \mathrm{d}y \, \frac{2}{\sqrt{2\pi(t-s)}} \, \mathrm{e}^{-\frac{y^2}{2(t-s)}} = \frac{2}{\pi} \int \mathrm{d}\tilde{x} \, \mathrm{e}^{-\frac{\tilde{x}^2}{2}} \int_0^{\tilde{x}\sqrt{\frac{s}{t-s}}} \mathrm{d}\tilde{y} \, \mathrm{e}^{-\frac{\tilde{y}^2}{2}} \,,$$

where we used the substitution $\tilde{x} = x/\sqrt{s}$ and $\tilde{y} = y/\sqrt{t-s}$ in the integrals. It is easiest to compute the value of this final integral by first differentiating with respect to s: since

$$\partial_s \sqrt{\frac{s}{t-s}} = \frac{1}{2}\sqrt{\frac{t-s}{s}} \frac{t}{(t-s)^2} = \frac{1}{2\sqrt{s(t-s)}} \frac{t}{t-s},$$

we find

$$\partial_s h(s) = \frac{2}{\pi} \int_0^\infty \mathrm{d}\tilde{x} \, \mathrm{e}^{-\frac{\tilde{x}^2}{2}} \, \mathrm{e}^{-\left(\tilde{x}\sqrt{\frac{s}{t-s}}\right)^2/2} \, \partial_s\Big(\tilde{x}\sqrt{\frac{s}{t-s}}\Big) = \frac{1}{\pi\sqrt{s(t-s)}} \underbrace{\frac{t}{t-s} \int_0^\infty \mathrm{d}\tilde{x} \, \tilde{x} \, \mathrm{e}^{-\frac{\tilde{x}^2}{2} \frac{t}{t-s}}}_{=1} .$$

Since $h(0) = 0$, integrating this gives $h(s) = \frac{2}{\pi} \arcsin(\sqrt{s/t})$. □

(2.50) Proposition: zeroes of Brownian motion

Let B be a one-dimensional Brownian motion, and let $Z(\omega) := \{t \geqslant 0 : B_t(\omega) = 0\}$ be the (random) subset of zeroes of Brownian motion.

a) $Z(\omega)$ is Lebesgue-measurable for all ω, and we have $\mathbb{E}^x(\lambda(Z)) = 0$ for all $x \in \mathbb{R}$, where λ is the Lebesgue measure.

b) Almost surely, Z is a *perfect subset* of $\mathbb{R}_0^+$, which means that

(i): Z is closed, and

(ii): Z has no isolated points, i.e. for every $z \in Z$ and every open $U \subset \mathbb{R}_0^+$ with $z \in U$, we have $U \cap (Z \setminus \{z\}) \neq \emptyset$.

Proof: a) $Z(\omega) = (B_\cdot(\omega))^{-1}(\{0\})$ is closed (and thus Lebesgue-measurable) as the inverse image of a closed set under a continuous function. The claimed equality is simply an application of Fubini: $\mathbb{E}^x(\lambda(Z)) = \int_0^\infty \mathbb{P}^x(B_t = 0)\,\mathrm{d}t = \int_0^\infty 0\,\mathrm{d}t = 0$.

b) We have already seen that $Z(\omega)$ is closed. To show that Z is perfect, let us define

$$\tau_{0,a}(\omega) := \inf\{s \geqslant a : B_s = 0\},$$

and

$$\Omega_1 := \{\omega \in \Omega : \forall a \in \mathbb{Q}, \exists (s_n) \subset \mathbb{R}_0^+ \text{ such that } s_n > \tau_a(\omega), B_{s_n}(\omega) = 0\ \forall n, \lim_{n\to\infty} s_n = \tau_a(\omega)\}.$$

In words, for all $\omega \in \Omega_1$, the next Brownian zero after each rational number can be approximated from the right by a sequence of other Brownian zeroes, and is therefore not isolated in Z. In particular, this means that if $\omega \in \Omega_1$ and $s \in Z(\omega)$, then

- either $s = \tau_{0,a}(\omega)$ for some $a \in \mathbb{Q}$, in which case it is approximated from the right by other Brownian zeroes, by the definition of Ω_1,
- or it is not of this form, in which case for all $a \in \mathbb{Q} \cap [0, s)$, there must be another Brownian zero in the interval (a, s). Then it easily follows that s is approximated from the left by Brownian zeroes.

In both cases, s is not isolated. This means that we can finish the proof by showing that $\mathbb{P}^x(\Omega_1) = 1$. Let

$$A = \{\omega \in \Omega : \exists (s_n) \subset \mathbb{R}_0^+ \text{ such that } s_n > 0, B_{s_n}(\omega) = 0\ \forall n, \lim_{n\to\infty} a_n = 0\}.$$

By (2.7), we have $\mathbb{P}^0(A) = 1$, and the strong Markov property gives

$$\mathbb{E}^x(\mathbb{1}_A \circ \theta_{\tau_{0,a}} \,|\, \mathcal{F}_{\tau_{0,a}})(\bar\omega) = \mathbb{E}^{\tau_{0,a}(\bar\omega)}(\mathbb{1}_A) = \mathbb{P}^0(A) = 1$$

for $\mathbb{P}^x$-almost all $\bar\omega$, and taking $\mathbb{E}^x$-expectation shows $\mathbb{E}^x(\mathbb{1}_A \circ \theta_{\tau_{0,a}}) = 1$ for all $a \in \mathbb{R}$, and thus $\mathbb{1}_A \circ \theta_{\tau_{0,a}} = 1$ almost surely. Since $\Omega_1 = \bigcap_{a\in\mathbb{Q}}\{\omega \in \Omega : \mathbb{1}_A \circ \theta_{\tau_{0,a}} = 1\}$, we have $\mathbb{P}^x(\Omega_1) = 1$. $\square$

Remarks: a) Perfect sets are always uncountable (exercise!). On the other hand, the zeroes of Brownian motion have Lebesgue measure zero almost surely, and therefore are not a „boring“ perfect set like a closed interval. Indeed, they look very much like a (random) Cantor set, with lots of holes but no isolated points.

b) It is also known that complements of perfect sets are countable unions of open intervals. Thus when Brownian motion leaves the value zero, it stays away from zero during an open time interval - but on the other hand, many of these open time intervals are so small that when we follow a Brownian path and find we have just hit zero, we know that we will hit zero again infinitely many times in any arbitrarily small future interval.

c) On the other hand, if we just came from somewhere else and have hit zero for the first time, then of course we have not hit zero at all in a sufficiently small time interval in the past. Thus when we „sit“ on a typical zero of Brownian motion, the view into the future is very different from the view into the past. Again on the other hand, we have the time reversal symmetry

(2.5) which tells us that we can invert the flow of time - how does this combine with what we just said?

d) The resolution is that while it is very *easy* to find the *first* zero of Brownian motion that has been started from $x \neq 0$, it is very *hard* to nail down the *last* zero before it goes onto an excursion. Of course, fore each Brownian path, such „last zeroes“ exist, but we cannot *predict* them by looking at our past, while we *can* tell when we are sitting on a „first zero“. In other words, while the *distribution* of Brownian motion is invariant under time reversal, the Markov property strongly needs the time direction because it separates past from future. Put differently, Brownian paths indeed „look the same“ when time-reversed, but only if we look globally on them as a whole - if we try to follow them as time goes on, we will find some features of them (first entry points) while others (last entry points) will be hidden from us.

e) Brownian paths are extremely intriguing mathematical objects, but in this lecture we will not have further time to investigate their geometry. I highly recommend looking into the Book „Brownian motion“ by Peter Mörters and Yuval Peres, which is available for free from the author's home pages. It contains many very beautiful further facts about geometry, fractal dimension, and many other aspects of Brownian paths.

D) Brownian Motion and partial differential equations

(2.51) The Brownian semigroup

Let B be a d-dimensional Brownian motion, and $f : \mathbb{R}^d \to \mathbb{R}$ be a bounded function. For all $t \geqslant 0$ let

$$P_t f(x) := \mathbb{E}^x(f(B_t)).$$

Then $(P_t)_{t \geqslant 0}$ is a semigroup of operators mapping $L^\infty(\mathbb{R}^d)$ into $C_b^\infty(\mathbb{R}^d, \mathbb{R})$ for all $t > 0$, i.e.

$$x \mapsto P_t f(x) \text{ is smooth for all } t > 0, \text{ and } [P_{t+s} f](x) = [P_t(P_s f)](x) \text{ for all } x.$$

Proof: Since

$$P_t f(x) = \frac{1}{(2\pi t)^{d/2}} \int \mathrm{e}^{-\frac{|x-y|^2}{2t}} f(y) \, \mathrm{d}y,$$

smoothness of $P_t f$ is clear e.g. by differentiating under the integral and invoking dominated convergence. The semigroup property follows from the Markov property of Brownian motion:

$$P_t f(x) = \mathbb{E}^x(f(B_{s+t})) = \mathbb{E}^x\big(\mathbb{E}^x(f(B_{t+s}) \,|\, \mathcal{F}_t)\big) = \mathbb{E}^x\big(\mathbb{E}^{B_t}(f(B_s))\big) = \mathbb{E}^x([P_s f](B_t)) = [P_t(P_s f)](x).$$

□

(2.52) Theorem

Let $f : \mathbb{R}^d \to \mathbb{R}$ be bounded and continuous on an open subset $A \subset \mathbb{R}^d$ with $\lambda(A^c) = 0$. Then the Brownian semigroup $(P_t f)_{t \geqslant 0}$ solves the **heat equation** with initial condition f, i.e. for all $t > 0$, we have

$$\partial_t P_t f(x) = \frac{1}{2} \Delta_x [P_t f](x) := \frac{1}{2} \sum_{j=1}^{d} \partial_{x_i}^2 [P_t f](x), \qquad \text{and} \qquad \lim_{t \to 0} P_t f(x) = f(x) \text{ for almost all } x.$$

Proof: For all $x \in A$, a change of variables in the integral gives

$$\mathbb{E}^x(f(B_t)) = \frac{1}{(2\pi t)^{d/2}} \int \mathrm{e}^{-\frac{|x-y|^2}{2t}} f(y)\,\mathrm{d}y = \frac{1}{(2\pi)^{d/2}} \int \mathrm{e}^{-\frac{|z|^2}{2}} f(x - \sqrt{t}z)\,\mathrm{d}z \xrightarrow{t\to 0} f(x)$$

by dominated convergence. For the differential equation, we recall that by Lemma (2.16), the transition density $p_t(x)$ of Brownian motion satisfies the heat equation, and so for $f \in C_b^2$ and $0 < \varepsilon < t$ we have

$$\begin{aligned} P_t f(x) - f(x) &= \int p_t(x-y) f(y)\,\mathrm{d}y - f(x) = \\ &= \int \mathrm{d}y \int_\varepsilon^t \mathrm{d}s \partial_s p_s(x-y) f(y) + \int \mathrm{d}y g_\varepsilon(x-y) f(y) - f(x) = \\ &= \int_\varepsilon^t \mathrm{d}s \int \mathrm{d}y \frac{1}{2}[\Delta p_s](x-y) f(y) + [P_\varepsilon f](x) - f(x). \end{aligned}$$

We have just proved that the last two terms cancel as $\varepsilon \to 0$. For the first term, note that $[\Delta p](x-y) = \Delta_y p(x-y)$ in the sense that in the first expression, the Laplace operator acts on the function p and only then the argument $x-y$ is inserted, and in the second expression it acts on the function $y \mapsto p_t(x-y)$. Then we can perform two integrations by part in the first expression and find

$$P_t f(x) - f(x) = \lim_{\varepsilon\to 0} \int_\varepsilon^t \mathrm{d}s \int \mathrm{d}y p_s(x-y)[\tfrac{1}{2}\Delta f](y) = \int_0^t \mathrm{d}s \int \mathrm{d}y p_s(x-y)[\tfrac{1}{2}\Delta f](y)$$

So,

$$\partial_t P_t f(x)|_{t=0} = \lim_{t\to 0} \frac{1}{t}(P_t f(x) - f(x)) = \lim_{t\to 0} \frac{1}{t} \int_0^t \mathrm{d}s \Big(\int \mathrm{d}y p_s(x-y)[\tfrac{1}{2}\Delta f](y) \Big).$$

Since $f \in C^2$, we have

$$\sup \Big\{ \Big| \int p_s(x-y)[\tfrac{1}{2}\Delta f](y)\,\mathrm{d}y - \frac{1}{2}\Delta f(x) \Big| : s \leqslant t \Big\} \to 0 \qquad \text{as } t \to 0,$$

and thus $\partial_t P_t f(x)|_{t=0} = \frac{1}{2}\Delta f(x)$. For $t > 0$, apply this result to $P_{t+h} f - P_t f = P_h[P_t f] - P_t f$ with C^2 (even smooth) function $P_t f$ instead of f. □

Remark: Theorem (2.52) is also called **Kolmogorov's backward equation** for Brownian motion. The term „backward“ comes from the fact that the „last“ point in time t of where $f(B_t)$ is evaluated becomes the starting point for the heat equation, and the starting point x of Brownian motion has to be inserted into the solution of the heat equation at its final time t. The backward equation is much more general than this, and holds for many continuous Markov processes. When we study solutions of stochastic differential equations, we will meet a few more processes for which it holds, only that then of course a different partial differential equation needs to be solved.

(2.53) The mean value property

Let $\partial K_r(x) := \{y \in \mathbb{R}^d : |y-x| = r\}$ be the sphere of radius r around x. Let $\sigma_{x,r}$ denote rotation invariant probability measure concentrated on $\partial K_r(x)$, i.e. the unique probability measure on

$\mathbb{R}^d$ with $\sigma_{x,r}(\partial K_r(x)) = \sigma_{x,r}(\mathbb{R}^d) = 1$, and $\sigma_{x,r}(UA) = \sigma_{x,r}(A)$ for all $A \in \mathcal{B}(\mathbb{R}^d)$, where U is a rotation around the point x.

Let $D \subset \mathbb{R}^d$ be open. A function $h : D \to \mathbb{R}$ is said to have the **mean value property** if for all $x \in D$ and all $r > 0$ so that $S_r(x) \subset D$, we have

$$h(x) = \int_{\partial K_r(x)} h(y)\sigma_{x,r}(\mathrm{d}y).$$

Theorem: A function $u \in C^2(D)$ has the mean value property on D if and only it is harmonic on D, i.e. if $\Delta u(x) = 0$ for all $x \in D$.

Proof: Let $u \in C^2$. A change of variables and dominated convergence give

$$\phi_x(r) := \int_{\partial K_r(x)} u(y)\,\sigma_{x,r}(\mathrm{d}y) = \int_{S_1(0)} u(x+ry)\,\sigma_{0,1}(\mathrm{d}y) \stackrel{r\to 0}{\longrightarrow} u(x).$$

We investigate $\phi_x'(r)$. We have

$$\begin{aligned}\phi_x'(r) &= \int_{\partial K_1(0)} \partial_r u(x+ry)\,\sigma_{0,1}(\mathrm{d}y) = \int_{\partial K_1(0)} [\nabla u](x+ry)\cdot y\,\sigma_{0,1}(\mathrm{d}y) = \\ &= \int_{\partial K_1(0)} \frac{1}{r}\nabla_y u(x+ry)\cdot y\,\sigma_{0,1}(\mathrm{d}y) = \frac{1}{r}\int_{K_1(0)} \Delta_y u(x+ry)\mathrm{d}y,\end{aligned}$$

where in the last step we used the Gauss divergence formula, since the vector y is just the normal vector of unit length at the point y of the unit sphere. $K_r(x)$ denotes the ball of radius r around x. Since $\Delta_y u(x+ry) = r^2[\Delta u](x+ry)$, we change variables back and get

$$\phi_x'(r) = r^{-d+1}\int_{K_r(x)} [\Delta u](y)\,\mathrm{d}y$$

for all $x \in D$ and all $r > 0$. The function u has the mean value property if and only if $\phi_x'(r) = 0$ for all x and r. On the one hand, for harmonic u, this is true. On the other hand, if $\Delta u(x) \neq 0$ for some $x \in D$, then by continuity and because D is open, we find $r > 0$ so that $K_r(x) \subset D$ and $\Delta u(x) \neq 0$ on $K_r(x)$, showing that u does not have the mean value property in this case. □

(2.54) The harmonic measure

Let $D \subset \mathbb{R}^d$, $D \neq \mathbb{R}^d$ be open, $x \in D$. Let ∂D denote the boundary of D, B a d-dimensional Brownian motion, and let $\tau_{\partial D} := \inf\{t \geqslant 0 : B_t \in \partial D\}$ be the hitting time of the boundary of D. The probability measure $\nu_{D,x}$ on $\mathcal{B}(\mathbb{R}^d) \cap \partial D$ defined through

$$\nu_{D,x}(A) = \mathbb{P}^x(B_{\tau_{\partial D}} \in A \,|\, \tau_{\partial D} < \infty)$$

is called the **harmonic measure** on ∂D of Brownian motion started in $x \in D$. For bounded D, the condition $\tau_{\partial D} < \infty$ can be dropped (why?). Rotational invariance of Brownian motion shows that when $D = K_r(x)$, then $\nu_{K_r(x),x} = \sigma_{x,r}$.

(2.55) Proposition

Let $D \subset \mathbb{R}^d$ be open, $h : \mathbb{R} \to \mathbb{R}$ be continuous with $\mathbb{E}^x(|h(B_t)|) < \infty$ for all t and all x. Let $\tau_{\partial D}$ be the hitting time of ∂D of Brownian motion as above (and $\tau_{\partial D} := \infty$ if $D = \mathbb{R}^d$).

(i): If h is C^2 on $\mathbb{R}^d$ and h is harmonic on D, then $(h(B_{t\wedge\tau_{\partial D}}))_{t\geqslant 0}$ is a martingale.
(ii): If $(h(B_{t\wedge\tau_{\partial D}}))_{t\geqslant 0}$ is a $(\mathcal{F}_{t\wedge\tau_{\partial D}})_{t\geqslant 0}$-martingale under all $\mathbb{P}^x$, then h is harmonic on D.

Proof: (i): By Theorem (2.17), the process (M_t^h) with

$$M_t^h = h(B_t) - \int_0^t \frac{1}{2}\Delta h(B_s)\,\mathrm{d}s$$

is a martingale. The same holds for $(M^h_{t\wedge\tau_{\delta D}})$ by optional stopping. Since $B_s \in D$ for all $s \leqslant \tau_{\partial D}$, the integral is zero when h is harmonic in D.
(ii): Let $x \in D$, $r > 0$ with $\partial K_x(r) \subset D$. By optional stopping, also the process $(h(B_{t\wedge\tau_{\partial K_x(r)}}))_{t\geqslant 0}$ is a martingale. This implies $h(x) = \mathbb{E}^x(h(B_{t\wedge\tau_{\partial K_x(r)}}))$ for all t, and since $\tau_{\partial K_x(r)} < \infty$ almost surely, we can take $t \to \infty$ to obtain

$$h(x) = \mathbb{E}^x(h(B_{\tau_{S_x(r)}})) = \int_{\partial K_r(x)} h(y)\,\mathbb{P}(B_{\tau_{\partial K_x(r)}} \in \mathrm{d}y) = \int_{\partial K_r(x)} h(y)\nu_{K_r(x)}(\mathrm{d}y).$$

This shows that h has the mean value property and is therefore harmonic. □

The reason for including the case $D \neq \mathbb{R}^d$ (and thus inducing additional notation and complexity) is that for $D = \mathbb{R}^d$, one can show that all bounded harmonic functions are constant (the so-called maximum principle). While Proposition (2.55) can then still be interesting for non-bounded (but locally bounded) harmonic functions, it is good to have it for bounded domains too, because there we can have many harmonic functions that are bounded. In fact, we can have a (unique) harmonic function for *every* boundary condition on domains D whose boundary is not too irregular. We will show this after a technical Lemma that is „obvious“ but somewhat unpleasant to prove.

(2.56) Lemma

Let B be a d-dimensional Brownian motion, $\kappa > 0$, and let $A = \{x \in \mathbb{R}^d : x_1 \geqslant \kappa \sum_{i=2}^d x_i^2\}$. Let τ_A be the hitting time of A, and $\tau_r := \tau_{\partial K_r(0)}$ be the hitting time of the boundary of the ball with radius r around 0. Then

$$\lim_{y\to 0} \mathbb{P}^y(\tau_A < \tau_{\sqrt{|y|}}) = 1.$$

Proof: We note two general facts:
1) For $z \in \mathbb{R}^d$ and $\delta > 0$, $z_1 > \delta$ and $|z| \leqslant \sqrt{\delta/\kappa}$ implies $z \in A$.
2) For $t > 0$, $c > 0$, and a one-dimensional Brownian motion B, we have $\mathbb{P}^0(|B_t| \leqslant c) = \mathbb{P}^0(|B_1| \leqslant c/\sqrt{t})$ by Brownian scaling, i.e. by (2.4). Then (2.48 b) implies

$$\mathbb{P}^0(\sup_{s\leqslant t} |B_s| \geqslant c) \leqslant \mathbb{P}^0(\sup_{s\leqslant t} B_s \geqslant c) + \mathbb{P}^0(\inf_{s\leqslant t} B_s \leqslant -c) = 2\mathbb{P}^0(|B_t| \geqslant c) = 2\mathbb{P}^0(|B_1| \geqslant c/\sqrt{t}).$$

Let now $t > 0$ be arbitrary for the moment. Then

$$\mathbb{P}^y(\tau_A < \tau_{\sqrt{|y|}}) \geqslant \mathbb{P}^y(\tau_A < t < \tau_{\sqrt{|y|}}) \overset{1),\delta=|y|}{\geqslant} \mathbb{P}^y(\sup_{s\leqslant t} X_s^{(1)} > |y|, \sup_{s\leqslant t} |X_s^{(i)}| \leqslant \sqrt{\tfrac{|y|}{\kappa d^2}} \ \forall i \geqslant 1)$$

$$\geqslant \mathbb{P}^{y_1}\big(\sup_{s\leqslant t} X_s^{(1)} > |y|, \sup_{s\leqslant t} |X_s^{(1)}| \leqslant \sqrt{\tfrac{|y|}{\kappa d^2}}\big) \prod_{i=2}^{n} \mathbb{P}^{y_i}\big(\sup_{s\leqslant t} |X_s^{(i)}| \leqslant d\sqrt{\tfrac{|y|}{\kappa d^2}}\big) = (*)$$

Now let us choose $t = M^2|y|^2$ for $M > 0$. Then for all i, fact 2) gives

$$\mathbb{P}^{y_i}\big(\sup_{s\leqslant t} |X_s^{(i)}| \leqslant \sqrt{\tfrac{|y|}{\kappa d^2}}\big) \geqslant 1 - \mathbb{P}^0\big(\sup_{s\leqslant t} |X_s^{(i)}| > \sqrt{\tfrac{|y|}{\kappa d^2}} - |y|\big) \geqslant 1 - 2\mathbb{P}^0\Big(|B_1| > \frac{1}{Md\sqrt{\kappa|y|}} - \frac{1}{M}\Big)$$

So the product on the right hand side of $(*)$ converges to 1 as $y \to 0$ for all fixed M. For the first factor of $(*)$, we have

$$\mathbb{P}^{y_1}\big(\sup_{s\leqslant t} X_s^{(1)} > |y|, \sup_{s\leqslant t} |X_s^{(1)}| \leqslant \sqrt{\tfrac{|y|}{\kappa d^2}}\big) \geqslant \mathbb{P}^{y_1}\big(\sup_{s\leqslant t} X_s^{(1)} > |y|\big) - \mathbb{P}^{y_1}\big(\sup_{s\leqslant t} |X_s^{(1)}| > \sqrt{\tfrac{|y|}{\kappa d^2}}\big)$$

We have

$$\mathbb{P}^{y_1}\big(\sup_{s\leqslant t} X_s^{(1)} > |y|\big) = \mathbb{P}^0\big(\sup_{s\leqslant t} X_s^{(1)} > |y| - y_1\big) = \mathbb{P}^0\big(|X_t^{(1)}| > |y| - y_1\big) \geqslant$$

$$\geqslant \mathbb{P}^0(|X_t|^{(1)} > 2|y|) = \mathbb{P}^0(|X_1^{(1)}| > \frac{2}{M}) \geqslant 1 - \frac{4}{M},$$

and $\lim_{y\to 0} \mathbb{P}^{y_1}\big(\sup_{s\leqslant t} |X_s^{(1)}| > \sqrt{\tfrac{|y|}{\kappa d^2}}\big) = 0$, as above. This shows $\lim_{y\to 0} \mathbb{P}^y(\tau_A < \tau_{\sqrt{|y|}}) \geqslant 1 - 4/M$ for all $M > 0$, proving the claim. □

(2.57) The solution of the Dirichlet problem

The Dirichlet problem: Let $D \subset \mathbb{R}^d$ be open, $f : \partial D \to \mathbb{R}$ any locally bounded function. A function $h : D \to \mathbb{R}$ is said to **solve the Dirichlet problem** with boundary condition f on D if

(i): h is harmonic on D.

(ii): $\lim_{x\to y} h(x) = f(y)$ for all point y where f is continuous.

A solution formula for the Dirichlet problem: Let $D \subset \mathbb{R}^d$ be bounded and open, and assume that ∂D can be locally described by a C^2-manifold. Let $f : \partial D \to \mathbb{R}$ be bounded. Then the function

$$h : D \to \mathbb{R}, \qquad x \mapsto \mathbb{E}^x\big(f(B_{\tau_{\partial D}})\big)$$

solves the Dirichlet problem with boundary condition f on D.

Proof: We first prove that h is harominc on D. We write $\tau = \tau_{\partial D}$ for brevity. By (2.55), we need to confirm that $(h(B_{\tau\wedge t}))_{t\geqslant 0}$ is a $(\mathcal{F}_{\tau\wedge t})_{t\geqslant 0}$-martingale. We use the strong Markov property on the function $F(B) = f(B_\tau)$. The strange thing about that function is that $F \circ \theta_s(B) = F(B)$ for all s, namely

$$F \circ \theta_s(B) = f(B_{s+\inf\{u:B_{s+u}\notin D\}}) = f(B_{s+\inf\{u-s:B_u\notin D\}}) = f(B_{\inf\{u:B_u\notin D\}}) = F(B).$$

Thus for $s < t$, we have

$$\begin{aligned}\mathbb{E}^x\big(h(B_{\tau\wedge t})\,|\,\mathcal{F}_{\tau\wedge s}\big) &= \mathbb{E}^x\Big(\mathbb{E}^{B_{\tau\wedge t}}(f(B_\tau))\,\Big|\,\mathcal{F}_{\tau\wedge s}\Big) = \mathbb{E}^x\Big(\mathbb{E}^{B_{\tau\wedge t}}(F(B))\,\Big|\,\mathcal{F}_{\tau\wedge s}\Big) = \\ &= \mathbb{E}^x\Big(\mathbb{E}^x\big(\theta_\tau F(B)\,|\,\mathcal{F}_{\tau\wedge t}\big)\,\Big|\,\mathcal{F}_{\tau\wedge s}\Big) = \mathbb{E}^x\Big(\mathbb{E}^x\big(F(B)\,|\,\mathcal{F}_{\tau\wedge t}\big)\,\Big|\,\mathcal{F}_{\tau\wedge s}\Big) \\ &= \mathbb{E}^x\big(F(B)\,|\,\mathcal{F}_{\tau\wedge s}\big) = \mathbb{E}^x\big(f(B_\tau)\,|\,\mathcal{F}_{\tau\wedge s}\big) = h(B_{\tau\wedge s}).\end{aligned}$$

Therefore h is harmonic in D.

The second task is to prove that $\lim_{x\to y} h(x) = f(y)$ when $x \to y$ and f is continuous at y. For this, let $(y_n) \subset D$ with $y_n \to y$ as $n \to \infty$. For each $n \in \mathbb{N}$, let z_n be the point in ∂D that is closest to y_n. We change to a coordinate system so that in the new coordinates, z_n is the origin and y_n lies on the negative x_1-axis. For large enough n (and thus small enough $z_n - y_n$), we can then find $\kappa_n > 0$ so that, in the new coordinate system,

$$\{x \in \mathbb{R}^d : x_1 > \kappa_n \sum_{i=2}^n x_i^2\} \cap K_{\sqrt{|y_n|}}(y_n) \subset D^c.$$

By the assumed curvature of the surface ∂D and the fact that $z_n \to y$, the sequence (κ_n) can be chosen to be bounded by some $\kappa > 1$. Now we apply Lemma (2.56) to show that $\lim_{n\to\infty} \mathbb{P}^{y_n}(\tau_{D^c} \leqslant \tau_{K_{\sqrt{|z_n-y_n|}}(z_n)}) = 0$. Since $|z_n - y_n| \leqslant |y - y_n|$ and thus $|z_n - y| \leqslant 2|y - y_n|$, we have $K_{\sqrt{|z_n-y_n|}}(z_n) \subset K_{3\sqrt{|z_n-y_n|}}(y)$ if $|y - y_n| \leqslant 1$. This shows that $\lim_{n\to\infty} \mathbb{P}^{y_n}(B_\tau \in K_{3\sqrt{|y_n-y|}}(y) = 1$. If f is continuous at y, we conclude

$$\begin{aligned}|h(y_n) - f(y)| \leqslant\ & \sup\{|f(z) - f(y)| : z \in \partial D, |z - y| \leqslant 3\sqrt{|y_n - y|})\} \\ &+ \|f\|_\infty \mathbb{P}^{y_n}(B_\tau \notin K_{3\sqrt{|y_n-y|}}(y)) \stackrel{n\to\infty}{\longrightarrow} 0.\end{aligned}$$

This finishes the proof. □

Remark: The condition on the boundary of D is not optimal. With additional effort, one can show that the *exterior cone condition* (or Poincaré cone condition) is sufficient: at each point z of ∂D, it must be possible to find a cone with tip z such that the intersection of that cone with a sufficiently small ball around z is fully inside D^c. See the book of Mörters and Peres (chapter 3) for details.

3. Stochastic integrals

(3.1) Motivation

In the introduction, we considered (informally) stochastic differential equations (SDE) of the type

$$\partial_t X_t = b(t, X) + \sigma(t, X_t)\xi_t,$$

where ξ_t is white noise; white noise is (again, for us only formally) the derivative of Brownian motion, and so we can integrate both sides above and arrive to the integral formulation of the

SDE

$$X_{0.t} = \int_0^t b(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\underbrace{\dot{B}_s\,\mathrm{d}s}_{=:\mathrm{d}B_s}$$

Here, $\dot{B}_s$ denotes the derivative of B at time s. We still don't know what the last integral should be, but we can try to interpret is as a Riemann-Stieltjes integral, i.e. we define

$$\int_0^t f(s)\mathrm{d}g(s) := \lim_{n\to\infty} \sum_{k=1}^{\lfloor nt \rfloor} f\big(\tfrac{k}{n}\big)\Big(g\big(\tfrac{k+1}{n}\big) - g\big(\tfrac{k}{n}\big)\Big). \qquad (*)$$

Notice that when $g \in C_b^1$ and $f \in C_b$, then as $n \to \infty$, $k \to \infty$ and $k/n \to s$, we have

$$\lim_{n\to\infty} n\Big(g\big(\tfrac{k+1}{n}\big) - g\big(\tfrac{k}{n}\big)\Big) = \dot{g}(s),$$

and so by multiplying and dividing the n-th element in the sequence on the right hand side of $(*)$ by n, using the n for the limit just described, and the other one for the Riemann sum, and after being a bit careful with justifying swapping the order of limits, we arrive at the equality

$$\int_0^t f(s)\,\mathrm{d}g(s) = \int_0^t f(s)\dot{g}(s)\,\mathrm{d}s.$$

In cases where g is not differentiable but where the right hand side of $(*)$ still converges, we can interpret the limit as a generalization of the integral $\int_0^t f(s)\dot{g}(s)\,\mathrm{d}s$, which is what we want. There is one caveat: just like for Riemann integrals, we would like the convergence in $(*)$ to hold not only for the regular partition of $[0,t]$ into intervals of length $1/n$, but for any sequence of partitions where the length of the shortest interval converges to zero. Our next few items will investigate when this is possible.

(3.2) Definition

Let $I = [a, b)$ with $0 \leqslant a < b \leqslant \infty$, $\mathcal{T} \subset I$ finite. The set of intervals

$$P(\mathcal{T}) := \Big\{\big\{x \in I : s \leqslant x < \inf\{t \in \mathcal{T} \cup \{b\} : t > s\}\big\} : s \in \mathcal{T}\Big\}$$

is the **partition generated by** $\mathcal{T}$. It $\mathcal{T} \subset \mathcal{T}'$, we say that $P(\mathcal{T}')$ **refines** $P(\mathcal{T})$, and write $P(\mathcal{T}) \subset P(\mathcal{T}')$. The number

$$|P(\mathcal{T})| := \max\{|t - s| : [s, t) \in P(\mathcal{T})\}$$

is the **fineness** of $P(\mathcal{T})$. For a function F from the set of partitions of $[a.b)$ into a metric space, we say that $\lim_{|P|\to 0} F(P)$ exists if for all sequences (P_n) of partitions with $P_n \subset P_{n+1}$ for all n and $\lim_{n\to\infty} |P_n| = 0$, the limit $\lim_{n\to\infty} F(P_n)$ exists and is independent from the chosen sequence.

(3.3) Example

a) Let $f : [0, T] \to \mathbb{R}$ be Riemann-integrable. Then

$$\lim_{|P|\to 0} \sum_{[s,t)\in P} f(s)(t - s) = \int_0^T f(s)\,\mathrm{d}s.$$

b) If $g \in C^2$ and f is continuous, then for $s < t$ the expansion $g(t) = g(s) + g'(s)(t-s) + g''(\xi)(t-s)^2$ with $s \leqslant \xi(s) \leqslant t$ gives

$$\sum_{[s,t)\in P} f(s)(g(t)-g(s)) = \sum_{[s,t)\in P} f(s)g'(s)(t-s) + \sum_{[s,t)\in P} f(s)g''(\xi) \underbrace{(t-s)^2}_{\leqslant |P|(t-s)} .$$

The last term above is bounded by $\|g''\|_\infty |P| \sum_{[s,t)\in P} f(s)(t-s)$ which by a) converges to zero as $|P| \to 0$. Thus

$$\lim_{|P|\to 0} \sum_{[s,t)\in P} f(s)(g(t)-g(s)) = \int_0^T f(s)g'(s)\,\mathrm{d}s =: \int_0^T f\,\mathrm{d}g.$$

This is the Riemann-Stieltjes integral.

The following relaxes the conditions of Example (3.3 b) significantly. Recall that C^α is the space of α-Hölder continuous functions.

(3.4) Theorem: Young integral, Young (1936)

Let $f \in C^\alpha([a,b],\mathbb{R})$, $X \in C^\beta([a,b],\mathbb{R})$ and assume $\gamma := \alpha + \beta > 1$. Then[6)]

$$\int_a^b f_r\,\mathrm{d}X_r := \lim_{|P|\to 0} \sum_{[u,v)\in P\cap[a,b)} f_u X_{u,v}$$

exists and is called the **Young integral** with **integrand** f and **integrator** integrator X.

Proof: For $a \leqslant u < v < b$, let

$$\Xi_{u,v} := f_u X_{u,v}$$

denote the „approximation of order zero to the integral $\int_u^v f_r\,\mathrm{d}X_r$“. The quantity $\Xi_{u,v}$ clearly does *not* enjoy the additivity property that a proper integral should have, i.e. in general we have $\Xi_{u,r} + \Xi_{r,v} \neq \Xi_{u,v}$ when $u < r < v$. Let

$$(\delta\Xi)_{u,r,v} := \Xi_{u,v} - (\Xi_{u,r} + \Xi_{r,v}) = f_u \underbrace{(X_{u,v} - X_{u,r})}_{=X_{r,v}} - f_r X_{r,v} = -f_{u,r} X_{r,v}$$

be the „extent of non-additivity“ that one gets for the points $u < r < v$. We have the estimate

$$\frac{|(\delta\Xi)_{u,r,v}|}{|v-u|^\gamma} \leqslant \frac{|f_{u,r}|}{|v-u|^\alpha}\frac{|X_{r,v}|}{|v-u|^\beta} \leqslant \frac{|f_{u,r}|}{|r-u|^\alpha}\frac{|X_{r,v}|}{|v-r|^\beta} \leqslant \|f\|_\alpha \|X\|_\beta.$$

Here and below, we write $\|f\|_\alpha$ instead of $\|f\|_{[s,t],\alpha}$ etc. for brevity. Our assumption $\gamma > 1$ now guarantees that

$$\sup_{r\in[u,v)} |(\delta\Xi)_{u,r,v}| \leqslant \|f\|_\alpha \|X\|_\beta |v-u|^\gamma \qquad \text{(A)}$$

vanishes faster than linearly when $|v-u| \to 0$. This is already the key estimate for the proof.

[6)] recall the increment notation $X_{u,v} = X_v - X_u$

Now consider $a \leqslant s = u_0 < u_1 < \cdots < u_m = t < b$, and the partition $P = P(\{u_0, \ldots, u_m\})$ of $[s, t)$. We have

$$t - s = \sum_{j=1}^{m} (u_j - u_{j-1}) \geqslant \frac{1}{2}\Big(\sum_{j=1}^{m-1} (u_j - u_{j-1}) + \sum_{j=2}^{m} (u_j - u_{j-1}) \Big) =$$
$$= \frac{1}{2} \sum_{j=1}^{m-1} (u_{j+1} - u_{j-1}) \geqslant \frac{1}{2}(m-1) \min_{1 \leqslant j \leqslant m-1} (u_{j+1} - u_{j-1}),$$

which means that there must be (at least) one $i \leqslant m-1$ with $u_{i+1} - u_{i-1} \leqslant \frac{2|t-s|}{m-1}$. This means that

$$\Big| \sum_{[v,w)\in P(\{u_j : j \neq i\})} \Xi_{v,w} - \sum_{[v,w)\in P} \Xi_{v,w} \Big| = |(\delta \Xi)_{u_{i-1}, u_i, u_{i+1}}| \leqslant \|f\|_\alpha \|X\|_\beta \Big(\frac{2|t-s|}{m-1} \Big)^\gamma$$

We now repeat this argument with the partition $P(\{u_0, \ldots, u_{i-1}, u_{i+1}, \ldots, u_m\})$, and continue recursively until we reach the trivial partition $P_0 = [s, t]$. A telescopic sum together with the triangle inequality then yields

$$\big| \Xi_{s,t} - \sum_{[v,w)\in P} \Xi_{v,w} \big| \leqslant \|f\|_\alpha \|X\|_\beta \sum_{j=1}^{m-1} \Big(\frac{2|t-s|}{m-j} \Big)^\gamma \leqslant \|f\|_\alpha \|X\|_\beta (2|t-s|)^\gamma \zeta(\gamma), \qquad (*)$$

and $\zeta(\gamma) = \sum_{j=1}^\infty j^{-\gamma} < \infty$ thanks to $\gamma > 1$.

Now let P' and P two partitions of $[a, b)$ such that P' refines P. We apply the estimate $(*)$ to each element $[u, v)$ of the partition P and obtain

$$\Big| \sum_{[u,v)\in P} \Xi_{u,v} - \sum_{[u',v')\in P'} \Xi_{u',v'} \Big| = \Big| \sum_{[u,v)\in P} \big(\Xi_{u,v} - \sum_{[w,z)\in P' \cap [u,v)} \Xi_{w,z} \big) \Big|$$
$$\leqslant \sum_{[u,v)\in P} \|f\|_\alpha \|X\|_\beta (2|v-u|)^\gamma \zeta(\gamma)$$
$$\leqslant 2^\gamma \|f\|_\alpha \|X\|_\beta \zeta(\gamma) \max_{[u,v)\in P} |v-u|^{\gamma-1} \underbrace{\sum_{[u,v)\in P} (v-u)}_{=b-a} = 2^\gamma \|f\|_\alpha \|X\|_\beta \zeta(\gamma)(b-a)|P|^{\gamma-1}.$$

So if we have two partitions P_n and P_m (not necessarily refining each other), the difference between their approximating sum is (by the triangle inequality) bounded by the sum of the difference of each partition to their common refinement. Each of these distances is bounded by a constant (not depending on the partitions) times the mesh (fine-ness) of the partition. This shows that for a sequence (P_n) with $|P_n| \to 0$, the sequence of approximating integrals is Cauchy. Thus the limit exists and is independent of the sequence of approximating partitions. The details of this argument are left as an exercise. □

Remark: The same proof shows the following extension of the Young integral: Let V, W, Z be Banach spaces, $f \in C^\alpha([a, b], V)$, $X \in C^\beta([a, b], W)$ and $H : V \times W \to Z$ continuous and bilinear. Then $\alpha + \beta > 1$ again guarantees the existence of the limit

$$\int_a^b H(f_r, \mathrm{d}X_r) := \lim_{|P|\to 0} \sum_{[u,v)\in P} H(f_u, X_{u,v}).$$

Indeed, bilinearity of H guarantees the validity of the important formula for $(\delta\Xi)_{u,r,v}$ (with $\Xi_{u,v} := H(f_u, X_{u,v})$ in this case), while continuity guarantees the validity of the estimates that come later. This extension will become important in the next chapter, when we apply it to $V = \mathbb{R}^n$, $W = \mathbb{R}^m$, and $H(v, w) = v \otimes w$ is the matrix $n \times m$-matrix with elements $(v \otimes w)_{i,j} = v_i w_j$.

(3.5) Remark

We can not take $f = X = B$ (with B a one-dimensional Brownian motion) in the previous Theorem, because $t \mapsto B_t$ is α-Hölder-continuous for all $\alpha < 1/2$ but not for $\alpha \geqslant 1/2$. So $\int_0^t B_s \,\mathrm{d}B_s$ is not a Young integral. It is not very difficult to see that if g is differentiable and $h \in C^\alpha$, then also the composition $t \mapsto g(h(t))$ is in C^α for the same α, but that on the other hand if $h \notin C^\alpha$, then also $g \circ h \notin C^\alpha$, at least at all places where $g' \neq 0$. Thus the integral $\int_0^t g(B_s) \,\mathrm{d}B_s$ is not a Young integral either in this case. We will see, however, that this is the type of integral that we need to understand if we want to solve stochastic differential equations. We therefore need a further extension of the Young integral. Unfortunately, the key estimate (A) in the proof of Theorem (3.4) is simply not true (with $\gamma > 1$) if $f = X = B$, and the sequence of approximating integrals does then *not* converge; more precisely, $\int_0^t B_s(\omega) \,\mathrm{d}B_s(\omega)$ can not be approximated as a Young integral for almost all ω.

a) The first is the classical Itô integral, and the idea is that while e.g. the expression

$$\lim_{|P|\to 0} \sum_{[u,v)\in P} g(B_u(\omega)) B_{u,v}(\omega)$$

from the statement of Theorem (3.4) might not exist for almost all ω, it could still exist in $L^2(\mathrm{d}\mathbb{P})$-sense, where $\mathbb{P}$ is the path measure of Brownian motion. The Itô theory establishes that this is indeed the case, and defines $\int_0^t g(B_s) \,\mathrm{d}B_s$ (and more general integrals) as the L^2-limit of its approximations.

Maybe this is already a good place to reflect about how it is possible that a sequence converges in L^2, but that there is no way to make sense of its pointwise limit. Note that the $L^2(\mathrm{d}\mathbb{P})$-convergence does imply almost sure convergence along a subsequence, so for each fixed sequence of partitions (P_n) with $|P_n| \to 0$, the limit $\lim_{k\to\infty} \sum_{[u,v)\in P_{n_k}} g(B_u(\omega)) B_{u,v}(\omega)$ exists almost surely for some subsequence (n_k). We could be tempted to take that as the definition of the pointwise limit. But for different sequences of partitions we will have to take different subsequences, and get different subsequential limits; all of these limits agree with the L^2-limit L^2-limit for almost all ω, but there are uncountably many sequences of partitions, and therefore uncountably many exceptional sets Ω_0 of measure zero where this subsequential limit does not agree with the L^2-limit. Therefore, there is no sensible way to go much beyond the L^2-limit in Itô theory.

b) The relatively new theory of rough paths provides such a way. The main idea is that the expression $\Xi_{u,v} = f_u X_{u,v}$, which is just the zero-order approximation of the integral $\int_u^v f_r \, \mathrm{d}X_r$, is not good enough, and one has to add to it a correction term that
(i): vanishes faster than $\Xi_{u,v}$ itself as $|u-v| \to 0$, i.e. it is indeed a correction term,
(ii): makes equation (A) (with the corrected $\Xi_{u,v}$) in the proof of (3.4) true for some $\gamma > 1$.

This choice is far from unique, but once we make it, we can actually prove convergence along all sequences of partitions, and independence of the limit from the chosen sequence. We will do this in the last chapter of the lecture and for now focus on the Itô integral.

(3.6) Definition

Let $\mathcal{T} = \{s_0, s_1 \dots, s_n\}$ be a finite subset of $\mathbb{R}_0^+$ with $0 = s_0 < s_1 < \cdots < s_n$. Let $(\Omega, \mathcal{F})$ be a measurable space and $(\mathcal{F}_t)$ a filtration of $\mathcal{F}$. A family $(Y_t)_{t \in \mathbb{R}_0^+}$ of measurable maps is called **$\mathcal{T}$-elementary** if for each $i \leqslant n$ there exists a bounded, $\mathcal{F}_{s_i}$-measurable random variable ϕ_{s_i} such that

$$Y_t(\omega) = \sum_{[u,v) \in P(\mathcal{T})} \phi_u(\omega) \mathbb{1}_{[u,v)}(t) = \begin{cases} \phi_0(\omega) & \text{if } 0 \leqslant t < s_1 \\ \phi_{s_1}(\omega) & \text{if } s_1 \leqslant t < s_2 \\ \dots & \\ \phi_{s_{n-1}}(\omega) & \text{if } s_{n-1} \leqslant t < s_n \\ \phi_{s_n}(\omega) & \text{if } s_n \leqslant t < \infty. \end{cases}$$

We write

$$\mathcal{E}_T := \{(Y_t)_{t \in \mathbb{R}_0^+} : (Y_t) \text{ is } \mathcal{T}\text{-elementary for some finite } \mathcal{T} \subset [0,T]\}.$$

Renark: Note that the integral is well-defined in the sense that its value does not depend on adding artificial points to the partition. If $s \notin \mathcal{T}$, we may add it without changing the value of Y_t: indeed, let $\tilde{\mathcal{T}} = \mathcal{T} \cup \{s\}$, and define $\tilde{\phi}_s(\omega) = \phi_{\max\{u \in \mathcal{T} : u < s\}}(\omega)$, and $\tilde{\phi}_u = \phi_u$ for all $u \in \mathcal{T}$. Then the elementary process $\tilde{Y}$ with $\tilde{Y}_t(\omega) = \sum_{[u,v) \in P(\tilde{\mathcal{T}})} \phi_u(\omega) \mathbb{1}[u,v)(t)$ is equal to Y, and also $[\tilde{Y} \bullet X]_t(\omega) = [Y \bullet X]_t(\omega)$ for all t and all ω.

(3.7) Proposition and Definition

Let (Y_t) be a $\mathcal{T}$-elementary stochastic process as in (3.6) for some $\mathcal{T}$, (X_t) an $(\mathcal{F}_t)$-martingale. Then with

$$[Y \bullet X]_t \equiv \int_0^t Y_s \, \mathrm{d}X_s := \sum_{[u,v) \in P(\mathcal{T})} \phi_u \int_0^t \mathbb{1}_{[u,v)}(s) \, \mathrm{d}X_s := \sum_{[u,v) \in P(\mathcal{T})} \phi_u (X_{v \wedge t} - X_{u \wedge t}),$$

the process $([Y \bullet X]_t)_{t \geqslant 0}$ is a martingale. It is called the **(elementary) stochastic integral** with integrand Y and integrator X.

Proof of the martingale property: Let $s > 0$. By the remark after (3.7), we may assume that $s \in \mathcal{T}$. We then get

$$\mathbb{E}([Y \bullet X]_t \,|\, \mathcal{F}_s) = \mathbb{E}\Big(\sum_{\substack{[u,v) \in \mathcal{P}: \\ [u,v) \subset [0,s)}} \phi_u X_{u,v} \,\Big|\, \mathcal{F}_s \Big) + \mathbb{E}\Big(\sum_{\substack{[u,v) \in \mathcal{P}: \\ [u,v) \subset [s,\infty)}} \mathbb{E}(\phi_u X_{u,v} \,|\, \mathcal{F}_u) \,\Big|\, \mathcal{F}_s \Big),$$

where in the last sum we used the tower property on each term. The sum in the first term above is acutally $\mathcal{F}_s$-measurable and thus gives $[Y \bullet X]_s$. In the second term, the $\mathcal{F}_u$-measurable quantity ϕ_u can be pulled in front of the inner conditional expectation, and $\mathbb{E}(X_{u,v} \,|\, \mathcal{F}_u) = 0$ since X is a martingale. Thus the second term vanishes, and the martingale property is proved. $\square$

To make more progress, we will need to impose more integrability than just the L^1 minimum requirement for a martingale.

(3.8) Definition

Let $0 < T < \infty$. We define

$$\mathcal{M}_T^2 := \big\{(M_t)_{0 \leqslant t \leqslant T} : M \text{ is a continuous martingale with } \mathbb{E}(M_t^2) < \infty \ \forall t \leqslant T\big\},$$

and

$$\|M\|_{\mathcal{M}_T^2} = \mathbb{E}(\sup_{s \in T} |M_s|^2)^{1/2}.$$

for $M \in \mathcal{M}_T^2$. Note that $\|M\|_{\mathcal{M}_T^2} < \infty$ follows from Doobs L^2-maximal inequality.

(3.9) Proposition

Let $X \in \mathcal{M}_T^2$, and let A be the qvp of X. Let $Y \in \mathcal{E}_T$. Then $M := Y \bullet X \in \mathcal{M}_T^2$, and its qvp $\tilde{A}$ is given by

$$\tilde{A}_t(\omega) = \int_0^t Y_t^2(\omega)\,\mathrm{d}A_t(\omega) := \sum_{[u,v) \in P(\mathcal{T})} \phi_u^2(\omega)(A_{v\wedge t}(\omega) - A_{u \wedge t}(\omega)).$$

Moreover, $\|Y \bullet X\|^2_{\mathcal{M}_T^2} \leqslant 4\mathbb{E}(\tilde{A}_T)$.

Proof: Let Y be given in the notation of Definition (3.6). $Y \bullet X$ is a martingale by (3.7), and Doobs L^2-inequality gives us

$$\mathbb{E}(\sup_{s \leqslant T} |M_s|^2) \leqslant 4\mathbb{E}(M_T^2) \leqslant 4 \max_{u \in \mathcal{T}} \|\phi_u\|_\infty^2 \mathbb{E}(X_T^2) < \infty.$$

Also, $t \mapsto \tilde{A}_t$ is continuous and increasing. Let us check that $(M_t^2 - \tilde{A}_t)$ is a martingale: let $s < t < T$. We may assume that $s, t \in \mathcal{T}$, because if they are not, we re-define Y by putting $\mathcal{T}' = \mathcal{T} \cup \{s,t\}$, $\phi_r(\omega) := \phi_{\max\{u \in \mathcal{T} : u \leqslant r\}}(\omega)$ for $r \in \{s,t\}$, and observing that this changes none of the values $Y_u(\omega)$, $M_u(\omega)$ or $\tilde{A}_u(\omega)$ for any ω and any u.

Now let $[u,v), [\tilde{u}, \tilde{v}) \in P(\mathcal{T})$ with $s \leqslant u$ and $\tilde{u} < u$ (or equivalently with $\tilde{v} \leqslant u$). Then

$$(*) \qquad \mathbb{E}(\phi_u \phi_{\tilde{u}} X_{u,v} X_{\tilde{u},\tilde{v}} \,|\, \mathcal{F}_s) = \mathbb{E}\Big(\phi_{\tilde{u}} X_{\tilde{u},\tilde{v}} \phi_u \underbrace{\mathbb{E}(X_{u,v} \,|\, \mathcal{F}_u)}_{=0} \Big| \mathcal{F}_s\Big) = 0,$$

and

$$(**) \quad \mathbb{E}(\phi_u^2 X_{u,v}^2 - \phi_u^2 A_{u,v} \,|\, \mathcal{F}_s) = \mathbb{E}\Big(\phi_u^2 \underbrace{\mathbb{E}(X_{u,v}^2 - A_{u,v} \,|\, \mathcal{F}_u)}_{=0} \Big| \mathcal{F}_s\Big) = 0.$$

Therefore

$$\begin{aligned}&\mathbb{E}(M_t^2-\tilde{A}_t\,|\,\mathcal{F}_s)=M_s^2-\tilde{A}_s+\mathbb{E}(M_{s,t}(M_t+M_s)-\tilde{A}_{s,t}\,|\,\mathcal{F}_s)=\\&=M_s^2-\tilde{A}_s+\mathbb{E}\Big(\sum_{\substack{[u,v):\\ s\leqslant u,v\leqslant t}}\phi_u X_{u,v}\big(2\sum_{[\tilde{u},\tilde{v}):\tilde{v}\leqslant s}\phi_{\tilde{u}}X_{\tilde{u},\tilde{v}}+\sum_{\substack{[\tilde{u},\tilde{v}):\\ s\leqslant \tilde{u},\tilde{v}\leqslant t}}\phi_{\tilde{u}}X_{\tilde{u},\tilde{v}}\big)-\sum_{\substack{[u,v):\\ s\leqslant u,v\leqslant t}}\phi_u^2A_{u,v}\,\Big|\,\mathcal{F}_s\Big)\\&\overset{(*)}{=}M_s^2-\tilde{A}_s+\mathbb{E}\Big(\sum_{\substack{[u,v):\\ s\leqslant u,v\leqslant t}}\phi_u^2X_{u,v}^2-\sum_{\substack{[u,v):\\ s\leqslant u,v\leqslant t}}\phi_u^2A_{u,v}\,\Big|\,\mathcal{F}_s\Big)\overset{(**)}{=}M_s^2-\tilde{A}_s^2,\end{aligned}$$

which is the martingale property. The claimed inequality just follows from Doobs L^2-inequality: $\|M\|^2_{\mathcal{M}_T^2}\leqslant 4\mathbb{E}(M_T^2)=4\mathbb{E}(\tilde{A}_T)$. □

(3.10) Remark and Definition

Since $t\mapsto A_t(\omega)$ is increasing and continuous for all ω, a finite measure μ_ω on $[0,T]$ is uniquely defined by the equations $\mu_{A,\omega}([s,t]):=A_t(\omega)-A_s(\omega)$ for all $0\leqslant s<t\leqslant T$. We write

$$\int_0^T Y_t(\omega)\,\mathrm{d}A_t(\omega):=\int_0^T Y_t(\omega)\,\mu_{A,\omega}(\mathrm{d}t)$$

whenever $\int_0^T|Y_t(\omega)|\,\mu_\omega(\mathrm{d}t)<\infty$ for a stochastic process Y.

(3.11) Definition

The measure μ_A on $\Omega\times[0,T]$ (with product σ-algebra) is uniquely defined by the equations

$$\mu_A(C\times[s,t])=\int\mathbb{1}_C(\omega)\mu_{A,\omega}([s,t])\mathbb{P}(\mathrm{d}\omega)\qquad\text{for all } C\in\mathcal{F},0\leqslant s<t\leqslant T.$$

We define $L^2(\mu_A)$ in the usual way. Note that for $Y=(Y_t)_{0\leqslant t\leqslant T}\in L^2(\mu_A)$, we have

$$\int_0^T Y_t^2(\omega)\,\mathrm{d}A_t(\omega)<\infty\qquad\mathbb{P}\text{-almost surely},$$

and that

$$\|Y\|_{L^2(\mu_A)}=\mathbb{E}\Big(\int_0^T Y_t^2(\omega)\,\mathrm{d}A_t(\omega)\Big)=\int\mathbb{P}(\mathrm{d}\omega)\int_0^T\mathrm{d}A_t(\omega)Y_t^2(\omega).$$

(3.12) Lemma

Let $X\in\mathcal{M}_T^2$ with qvp A. For $t\leqslant T$, the map

$$\mathcal{E}_t\to L^2(\mathrm{d}\mathbb{P}),\qquad Y\mapsto\int_0^t Y_s\,\mathrm{d}X_s$$

(see Definition (3.7)) is a linear isometry between the spaces $(\mathcal{E}_t,\|.\|_{L^2(\mu_A)})$ and $L^2(\mathrm{d}\mathbb{P})$. In particular,

$$\mathbb{E}\Big(\big(\int_0^t Y_s\,\mathrm{d}X_s\big)^2\Big)=\mathbb{E}\Big(\int_0^t Y_s^2\,\mathrm{d}A_s\Big)$$

for all $Y\in\mathcal{E}_T$ and all $t\leqslant T$.

Proof: We use the equality $\mathbb{E}(M_t^2) = \mathbb{E}(A_t)$, valid for any square integrable martingale M and its qvp A, in the case where $M_t = \int_0^t Y_s \, \mathrm{d}X_s$. By (3.9), the qvp is given by $\tilde{A}_t = \int_0^t Y_s^2 \, \mathrm{d}A_s$, which shows the claimed equality, and thus the isometry property. Linearity is clear. □

(3.13) Definition

Let $X \in \mathcal{M}_T^2$. $\mathcal{L}_T^2(M)$ denotes the closure of $\mathcal{E}_t$ with respect to the norm $\|.\|_{L^2(\mu_A)}$, i.e.

$$f \in \mathcal{L}_T^2(M) \qquad \overset{\text{Def}}{\Longleftrightarrow} \qquad \exists (f_n) \subset \mathcal{E}_T \text{ with } \|f_n - f\|_{L^2(\mu_A)} \overset{n\to\infty}{\longrightarrow} 0.$$

(3.14) The Itô-integral

Let $0 \leqslant t \leqslant T$, $X \in \mathcal{M}_T^2$, $Y \in \mathcal{L}_T^2(X)$. Then $M = (X_s)_{s \leqslant t} \in \mathcal{M}_t^2$ and $(Y_s)_{s \leqslant t} \in \mathcal{L}_t^2(M)$. We define

$$(*) \qquad \int_0^t Y_s \, \mathrm{d}X_s := \lim_{n\to\infty} \int_0^t Y_s^{(n)} \, \mathrm{d}X_s, \qquad (\text{limit in } L^2(\mathrm{d}\mathbb{P})),$$

where $(Y^{(n)})_{0 \leqslant s \leqslant t}$ is any sequence in $\mathcal{E}_t$ that fulfills $\lim_{n\to\infty} \|Y^{(n)} - Y\|_{L^2(\mu_A)} = 0$. The random variable $\int_0^t Y_s \, \mathrm{d}X_s$ is independent of the approximating sequence in the sense of $L^2(\mathrm{d}\mathbb{P})$, i.e. we have

$$\lim_{n\to\infty} \mathbb{E}\Big(\big(\int_0^t Y_s^{(n)} \, \mathrm{d}X_s - \int_0^t \tilde{Y}_s^{(n)} \, \mathrm{d}X_s \big)^2 \Big) = 0 \qquad \text{whenever} \qquad \lim_{n\to\infty} \|Y^{(n)} - \tilde{Y}^{(n)}\|_{L^2(\mu_A)} = 0.$$

$\int_0^t Y_s \, \mathrm{d}X_s$ is called the **Itô-integral** with integrand Y and integrator X.

Proof: $M \in \mathcal{M}_t^2$ is clear, and $(Y_s)_{s \leqslant t} \in \mathcal{L}^2(M)$ can be seen by cutting off the an approximating sequence of elementary functions from $\mathcal{E}_T$ at time t, it then remains elementary. A sequence $(Y_s^{(n)})_{s \leqslant t}$ with $\|Y^{(n)} - Y\|_{L^2(\mu_A)} \to 0$ is in particular a Cauchy sequence in $L^2(\mu_A)$. By the isometry proved in (3.12), then $(\int_0^t Y_s^{(n)} \, \mathrm{d}X_s)_{n\in\mathbb{N}}$ is a Cauchy sequence in $L^2(\mathrm{d}\mathbb{P})$. Since this space is complete, the limit exists and defines $\int_0^t Y_s \, \mathrm{d}X_s$. Independence of this limit from the approximating sequence is proved exactly in the same way, using the isometry. □

(3.15) Proposition

Let $X \in \mathcal{M}_t^2, Y \in \mathcal{L}_t^2(M)$. Then

a) The map $Y \mapsto \int Y_s \, \mathrm{d}X_s$ is a linear ismoetry from $L^2(\mu_A)$ to $L^2(\mathrm{d}\mathbb{P})$.

b) The **Itô-isometry** holds:

$$\mathbb{E}\Big(\big(\int_0^t Y_s \, \mathrm{d}X_s \big)^2 \Big) = \mathbb{E}\Big(\int_0^t Y_s^2 \, \mathrm{d}A_s \Big).$$

Proof: this follows directly from the corresponding properties of the approximating sequences. □

(3.16) Proposition

a) Let $(\mathcal{F}_t)$ be a filtration, $X \in \mathcal{M}_T^2$ with qvp A, and Y a stochastic process such that
(i): Y is $(\mathcal{F}_t)$-adapted.
(ii): $t \mapsto Y_t$ is almost surely continuous.
(iii): $Y \in L^2(\mu_A)$.
Then $Y \in \mathcal{L}_T^2$.
b) If in the situation of a), either Y is bounded, or if for

$$\rho(\delta) := \sup\Big\{\frac{1}{v-u}\mathbb{E}\Big(\int_u^v (Y_{u,s})^2 \,\mathrm{d}A_s\Big) : 0 \leqslant u < v \leqslant T, |u-v| \leqslant \delta\Big\}$$

we have $\lim_{\delta\to 0}\rho(\delta) = 0$, then $\int_0^t Y_s \,\mathrm{d}X_s$ can be approximated (in $L^2(\mathrm{d}\mathbb{P})$) by arbitrary partitions, with the limit independent of the partition, i.e.

$$(*) \qquad \lim_{|P|\to 0} \sum_{[u,v)\in P} Y_u X_{u,v} = \int_0^T Y_s \,\mathrm{d}X_s \qquad \text{in } L^2(\mathrm{d}\mathbb{P}).$$

Proof: We first consider the case where Y is bounded and fulfils conditions (i)-(iii), and prove the statements of a) and b) in this case. For a partition P (with set of separating points $\mathcal{T}$), we define for all $t \leqslant T$

$$Y_t^P(\omega) := \sum_{[u,v)\in P} Y_u(\omega)\mathbb{1}_{[u,v)}(t).$$

Then $Y^P \in \mathcal{E}_T$ by (i) and our boundedness assumption. By (ii), we have

$$Y_t^P(\omega) = Y_{\max\{r\in\mathcal{T} : r \leqslant t\}}(\omega) \overset{|P|\to 0}{\longrightarrow} Y_t(\omega)$$

for almost all ω. We have $\| \|Y\|_\infty \|^2_{L^2(\mu_A)} = \|Y\|_\infty^2 \mathbb{E}(A_T) < \infty$, and so we can use $\|Y\|_\infty$ as an integrable, dominating function for Y^P for all partitions P and find $Y^P \to Y$ in $L^2(\mu_A)$ as $|P| \to 0$. This shows that $Y \in \mathcal{L}_T^2$, and so $\int_0^t Y_s \,\mathrm{d}X_s$ exists and is independent of the approximating sequence of elements from $\mathcal{E}_T$. Since the sequence of partitions P above was arbitrary, this shows $(*)$ for the case of bounded functions.

To complete the proof of a) without assuming boundedness of Y (and instead using (iii)), consider $Y^{(n)} = (Y_t \mathbb{1}_{\{|Y_t| \leqslant n\}})_{t \leqslant T}$. By the first part of the proof, $Y^{(n)} \in \mathcal{L}_T^2$ for all n. On the other hand, $\|Y - Y^{(n)}\|_{L^2(\mu_A)} \to 0$ as $n \to \infty$ by dominated convergence. Since $\mathcal{L}_T^2$ is a closed subspace of $L^2(\mu_A)$, this implies $Y \in \mathcal{L}_T^2$.

Turning to equality $(*)$, note that the dominated convergence argument in the final step can (and in general will) destroy its validity for general unbounded Y. The reason is that for $(*)$ to hold, we would need to exchange the limit where $Y^{(n)}$ approximates Y with the limits of partitions, and this is in general not possible. However, the second conditions guarantees that for any sequence of partitions,

$$\mathbb{E}\Big(\int_0^T \big(Y_s^P - Y_s\big)^2 \,\mathrm{d}A_s\Big) = \sum_{[u,v)\in P} \mathbb{E}\Big((v-u)\frac{1}{v-u}\int_u^v (Y_{u,s})^2 \,\mathrm{d}A_s\Big) \leqslant \rho(|P|) \underbrace{\sum_{[u,v)\in P}(v-u)}_{=T} \overset{|P|\to 0}{\longrightarrow} 0,$$

which shows $(*)$ in this case. □

(3.17) Remark

a) In assumption (ii) above, left-continuity of the paths of Y is enough, as can be seen from the proof.

b) The condition $\lim_{\delta\to 0}\rho(\delta) = 0$ is a bit hard to understand intuitively, but is actually not very strong. Let us assume for simplicity that $A_t = t$, i.e. the case of Brownian motion. Then

$$\mathbb{E}\Big(\int_u^v (Y_{u,s})^2 \,\mathrm{d}A_s\Big) = \int_u^v \mathbb{E}((Y_{u,s})^2)\,\mathrm{d}s = (*),$$

and if we now assume any kind of Kolmogorov-Chentsov-type control over the second moment of the increments, we are done. More precisely, assume that $\mathbb{E}((Y_t - Y_s)^2) \leqslant |t-s|^\alpha$ for some $\alpha > 0$, then $(*) \leqslant (v-u)^{1+\alpha}$ as $|u - v| \to 0$, which shows that (ii) is valid. This is a much weaker condition than the one in Theorem 1.36, so in particular we can include some (Y_t) with discontinuous paths, as long as the second moment of increments remains a Hölder-continuous function, i.e. Y^2 is not discontinuous „on average“.

It is a bit strange that for uniformly bounded Y, we need no restriction at all on the regularity of increments, while for possibly unbounded Y, we need (ii), but that's just what we get at the moment. Also, when A_t itself is a nontrivial stochastic process, we have to take the behaviour of its paths into account, and the simple picture above becomes more complicated.

(3.18) Important example

Let B be a BM^1, then

$$\int_0^t B_s \,\mathrm{d}B_s = \frac{1}{2}B_t^2 - \frac{1}{2}t.$$

Proof: By (3.16), $B \in \mathcal{L}_T^2$, and condition (ii) holds since $\rho(\delta) = \frac{1}{2}\delta^2$ by an easy calculation. So the Itô integral can be approximated by partitions. For a partition P of $[0,t)$, we have

$$\sum_{[u,v)\in P} B_u B_{u,v} = \sum_{[u,v)\in P} \tfrac{1}{2}B_v^2 - \tfrac{1}{2}B_u^2 - \big(\tfrac{1}{2}B_u^2 + \tfrac{1}{2}B_v^2 - B_u B_v\big) = \tfrac{1}{2}B_t^2 - \tfrac{1}{2}\sum_{[u,v)\in P} B_{u,v}^2.$$

You can prove as an exercise that the second term in the final expression converges to $\frac{1}{2}t$ in $L^2(\mathrm{d}\mathbb{P})$ as $|P| \to 0$. □

(3.19) Remark

a) The same calculation shows $\int_0^t X_s \,\mathrm{d}X_s = \frac{1}{2}X_t^2 - \frac{1}{2}A_t$ for $X \in \mathcal{M}_t^2$ with qvp A.

b) For $f \in C^1$ with $f_0 = 0$, we can define $\int_0^t f_s \,\mathrm{d}f_s$ as a Young integral, and

$$\int_0^t f_s \,\mathrm{d}f_s = \int_0^t f_s f_s' \,\mathrm{d}s = \int_0^t \big(\tfrac{1}{2}f^2\big)' \,\mathrm{d}s = \frac{1}{2}f_t^2.$$

So for functions where the function $s \mapsto f_s^2$ can be differentiated using the chain rule, the extra term $\frac{1}{2}t$ or $\frac{1}{2}A_t$ is absent. So, the Itô-integral for Brownian motion (or non-trivial continuous martingales) does not fulfill the integral version of the chain rule. This is an important point that will come up again many times in this lecture.

c) There actually is a way to define an integral of Brownian motion against itself that does fulfill the chain rule: note that the limit in $(*)$ in (3.16) is actually the limit of a Riemann sum where the function Y is always evaluated at the *left endpoint* of the intervals in the Riemann approximation. Of course, we could also try to evaluate at the right end point, or use the trapezoid rule of numerics, i.e. take the average of left and right end point. This then leads to

$$\int_0^t B_s \circ \mathrm{d}B_s := \lim_{|P|\to 0} \sum_{[u,v)\in P} \frac{B_u + B_v}{2} B_{u,v} = \lim_{|P|\to 0} \frac{1}{2} \sum_{[u,v)\in P} (B_v^2 - B_u^2) = \frac{1}{2} B_t^2.$$

This looks very nice, so why don't we always do it like this? There are two reasons: first of all, it only works so nicely for the case where $Y_t = B_t$, but e.g. for $Y_t = \sin(B_t)$ it is not so clear what to do! The second reason is that when we do *not* use the left endpoint approximation, then the process $t \mapsto \int_0^t f(B_s)\,\mathrm{d}B_s$ is no longer a martingale, and this is a big advantage to lose. Nevertheless, the integral $\int_0^t B_s \circ \mathrm{d}B_s$ is very useful in many instances, and the general strategy of using the trapezoid rule leads to the so-called **Stratonovich integral**.

d) You should verify the following important equality as an exercise: if $B = (B^{(1)}, B^{(2)})$ is a two-dimensional Brownian motion, then

$$\int_0^t B_s^{(1)}\,\mathrm{d}B_s^{(2)} + \int_0^t B_s^{(2)}\,\mathrm{d}B_s^{(1)} = B_t^{(1)} B_t^{(2)}.$$

Notice that this means that *independent* Brownian motions behave like differentiable functions when integrated against each other: To see what this means, let f, g be differentiable, and integrate the equality $(fg)' = f'g + g'f$ from 0 to t, and use $\int f_s g_s'\,\mathrm{d}s = \int f_s\,\mathrm{d}g_s$. To prove the equality for Brownian motion, you should first try to show that $\sum_{[u,v)} B_u^{(1)} B_{u,v}^{(2)} - \sum_{[u,v)} B_v^{(1)} B_{u,v}^{(2)}$ converges to zero in L^2 as $|P| \to 0$.

e) The choice of trapezoid rule that we made in d) is not the only alternative to the left endpoint rule that leads to the Itô-integral. As an example of what else is possible, let $f \in C^2$ and $Y = f(X)$. Then for all $\lambda \in [0,1]$ we can define

$$\sum_{[u,v)\in P} f(X_u + \lambda X_{u,v}) X_{u,v} = \sum_{[u,v)\in P} f(X_u) X_{u,v} + \sum_{[u,v)\in P} f'(X_u) \lambda X_{u,v} X_{u,v} + \frac{1}{2} \sum_{[u,v)\in P} f''(X_{\xi_u}) \lambda^2 X_{u,v}^3,$$

with $u \leqslant \xi_u \leqslant v$. You should use this expansion to show as an exercise that

$$\lim_{|P|\to 0} \sum_{[u,v)\in P} f(X_u + \lambda X_{u,v}) X_{u,v} = \int_0^t f(X_s)\,\mathrm{d}X_s + \lambda \int_0^t f'(X_s)\,\mathrm{d}A_s,$$

in particular the third term in the expansion vanishes. So in general, this gives a *different* integral for each choice of λ. For some λ, the resulting integrals have names: $\lambda = 0$ is the Itô-integral, $\lambda = 1/2$ is the Stratonovich-integral (written as $\int_0^t f(X_s) \circ \mathrm{d}X_s$), and $\lambda = 1$ is the *backwards-Itô-integral*. The most important cases are $\lambda = 0$ and $\lambda = 1/2$. The relation

$$\int_0^t f(X_s) \circ \mathrm{d}X_s = \int_0^t f(X_s)\,\mathrm{d}X_s + \frac{1}{2} \int_0^t f'(X_s)\,\mathrm{d}A_s$$

is called the **Itô-Stratonovich-correction**.

In the next items, we will investigate the stochastic process $(\int_0^t Y_s \,\mathrm{d}X_s)_{t \geqslant 0}$ made from stochastic integrals.

(3.20) Proposition

$\mathcal{M}_T^2$ is a Banach space under the norm $M \mapsto \|M\|_{\mathcal{M}_T^2} := \mathbb{E}\big(\sup_{s \leqslant T} |M_s|^2\big)^{1/2}$.

Proof: L^2-spaces where the target space is itself a Banach space are Banach spaces. Consider the space $C([0,T])$ of continuous functions with the sup norm and the σ-algebra generated by the point evaluations, and define

$$L^2(\mathrm{d}\mathbb{P}, C([0,T])) := \Big\{ f : \Omega \to C([0,T]) : f \text{ measurable}, \int \|f(\omega)\|_\infty^2 \,\mathbb{P}(\mathrm{d}\omega) < \infty \Big\}$$

Since $C([0,T])$ with the sup norm is a Banach space, we only need to show that $\mathcal{M}_T^2$ is a closed subspace of $L^2(\mathrm{d}\mathbb{P}, C([0,T]))$. So let $(M^{(n)}) \subset \mathcal{M}_T^2$ and $M \in L^2(\mathrm{d}P, C([0,T]))$ with $\|M^{(n)} - M\|_{\mathcal{M}_T^2} \to 0$ as $n \to \infty$. We need to show that M is also a continuous martingale. First of all, we can choose a subsequence, also denoted by $(M^{(n)})$, so that $\|M^{(n)} - M\|_\infty \to 0$ $\mathbb{P}$-almost surely. This shows that M has continuous paths almost surely. To show that M is a martingale, let $s < t$, and note that for all $r \leqslant T$,

$$\lim_{n\to\infty} \mathbb{E}(|M_r^{(n)} - M_r|) \leqslant \lim_{n\to\infty} \mathbb{E}(\|M^{(n)} - M\|_\infty) \leqslant \lim_{n\to\infty} \mathbb{E}(\|M^{(n)} - M\|_\infty^2)^{1/2} = 0$$

by assumption. Thus for all $A \in \mathcal{F}_s$, we have

$$\mathbb{E}(\mathbb{1}_A M_t) = \lim_{n\to\infty} \mathbb{E}(\mathbb{1}_A M_t^{(n)}) \overset{(*)}{=} \lim_{n\to\infty} \mathbb{E}(\mathbb{1}_A M_s^{(n)}) = \mathbb{E}(\mathbb{1}_A M_s),$$

where the equality $(*)$ holds because each $M^{(n)}$ is a martingale. Since

$$M_s(\omega) = \limsup_{n\to\infty} M_s^{(n)}(\omega) \in \mathrm{m}\mathcal{F}_s,$$

we see that M is a martingale. □

(3.21) Theorem

Let $X \in \mathcal{M}_T^2$ with qvp A. The map

$$\mathcal{L}_T^2 \to \mathcal{M}_T^2, \qquad Y \mapsto \Big(\int_0^t Y_s \,\mathrm{d}X_s \Big)_{t \leqslant T}$$

is a bounded linear operator from $\mathcal{L}_T^2$ to $\mathcal{M}_T^2$, and the qvp of the martingale $(M_t)_{t \leqslant T} := (\int_0^t Y_s \,\mathrm{d}X_s)_{t \leqslant T}$ is given by

$$(\tilde{A}_t)_{t \leqslant T} = (\int_0^t Y_s^2 \,\mathrm{d}A_s)_{t \leqslant T}.$$

Proof: let $Y^{(n)}$ be a sequence of elementary stochastic processes approximating Y, then for all n, $M^{(n)} = \int_0^\cdot Y_s^{(n)} \,\mathrm{d}X_s$ is in $\mathcal{M}_T^2$ with qvp $\tilde{A}^{(n)} = \int_0^\cdot (Y_s^{(n)})^2 \,\mathrm{d}A_s$ by (3.9). Since $M^{(n)} - M^{(m)}$ is a martingale for all m, n, we have

$$\|M^{(n)} - M^{(m)}\|_{\mathcal{M}^2_T} \overset{\text{Doob}}{\leqslant} 2\|M^{(n)}_T - M^{(m)}_T\|_{L^2(\mathrm{d}\mathbb{P})} \overset{(3.12)}{=} \|Y^{(n)} - Y^{(m)}\|_{L^2(\mu_A)} \overset{n,m\to\infty}{\longrightarrow} 0.$$

So the sequence $(M^{(n)})$ is $\mathcal{M}^2_T$-Cauchy, and by (3.20) it thus converges to an element of $\mathcal{M}^2_T$. This shows that M is a continuous martingale. We have

$$\|M\|_{\mathcal{M}^2_T} \overset{\text{Doob}}{\leqslant} 2\|M_T\|_{L^2(\mathrm{d}\mathbb{P})} \overset{\text{Itô}}{=} 2\|Y\|_{L^2(\mu_A)},$$

so the map $Y \to M$ is bounded. Its linearity is trivial.

To show that $\tilde{A}$ is the qvp, first note that for all t,

$$\|\tilde{A}^{(n)}_t - \tilde{A}_t\|_{L^1} \leqslant \mathbb{E}\Big(\int_0^t \underbrace{\big|(Y^{(n)}_s)^2 - Y^2_s\big|}_{=|Y^{(n)}_s - Y_s||Y^{(n)}_s + Y_s|} \,\mathrm{d}A_s\Big) \leqslant \mathbb{E}\Big(\int_0^t |Y^{(n)}_s - Y_s|^2 \,\mathrm{d}A_s\Big)^{\frac12} \mathbb{E}\Big(\int_0^t |Y^{(n)}_s + Y_s|^2 \,\mathrm{d}A_s\Big)^{\frac12}$$

by the Cauchy-Schwarz inequality applied in $L^2(\mu_A)$. The first term on the right above goes to zero, the second is bounded, showing that $\tilde{A}^{(n)}_t \to \tilde{A}_t$ in L^1. A similar argument shows

$$\|(M^{(n)}_t)^2 - M^2_t\|_{L^1} \overset{n\to\infty}{\longrightarrow} 0.$$

Finally, for each n, all $s < t$, and all $C \in \mathcal{F}_s$,

$$\mathbb{E}\Big(\big((M^{(n)})^2_t - \tilde{A}^{(n)}_t\big)\mathbb{1}_C\Big) = \mathbb{E}\Big(\big((M^{(n)})^2_s - \tilde{A}^{(n)}_s\big)\mathbb{1}_C\Big),$$

because $(M^{(n)})^2 - \tilde{A}^{(n)}$ is a martingale, and the L^1-convergence that we just proved shows that this equality survives the limit $n \to \infty$. Therefore also $M^2 - \tilde{A}$ is a martingale, which shows that $\tilde{A}$ is the qvp of M. □

So far, we have only defined the stochastic integral for $Y \in \mathcal{L}^2_T$. On the other hand, it is quite reasonable to want integrals also for stochastic processes Y with locally bounded paths, but with $Y \notin \mathcal{L}^2_T$. One example is $Y_t = \exp(B^2_t)$, which is a bounded function on $[0, T]$ for every Brownian path, but where

$$\mathbb{E}(Y^2_t) = \frac{1}{\sqrt{2\pi t}} \int \mathrm{e}^{2x^2}\ \mathrm{e}^{-\frac{x^2}{2t}}\ \mathrm{d}x = \infty$$

for $t \geqslant 1/4$. So, the reason why we can not define $\int Y_t \,\mathrm{d}B_t$ in this case is because there are „too many“ paths where Y_t gets large, while for each individual path, the integral should be harmless! The problem is that our Itô-integral needs all paths to work together to achieve its L^2-convergence, and if too many of them are bad, they spoil the convergence also for the well-behaved ones. The solution is to hit the bad paths on the head, more precisely when $Y_t(\omega)$ exceeds a certain level D, we just freeze the path belonging to ω for all future points in time. We then define the integral in this case, and in the end take the freezing threshold D to infinity *in the integral itself.* This technique, which we will elaborate now, is known as **localization**. We start with a technical result on stopped stochastic integrals.

(3.22) Lemma

Let $X \in \mathcal{M}_T^2$, $Y \in \mathcal{L}_T^2$ and τ a stopping time. Then

$$\int_0^{t\wedge\tau} Y_s \mathrm{d}X_s = \int_0^t Y_s \mathrm{d}X_{s\wedge\tau} = \int_0^t Y_s \mathbb{1}_{[0,\tau)}(s)\, \mathrm{d}X_s$$

almost surely.

Remark: it is useful to think about what the difference between the three expressions above is. The first is the Martingale $(\int_0^t Y_s \,\mathrm{d}X_s)_{s \leqslant T}$ stopped with the stopping time τ. The second is the (un-stopped) stochastic integral if Y with the stopped martingale $X_{s\wedge\tau}$ as integrator - this works since stopped martingales are martingales. The final expression is the normal stochastic integral with integrator X, but with an integrand that is zero for all pairs (s, ω) where $s \geqslant \tau(\omega)$. The Lemma says that these three ways of stopping lead to the same result.

Proof of Lemma (3.22):

a) We start with the first equality. It is true for elementary $Y \in \mathcal{E}_T$, since then for the partition P belonging to Y, we have

$$\int_0^{t\wedge\tau(\omega)} Y_s \mathrm{d}X_s(\omega) = \sum_{[u,v)\in P} Y_u(\omega)\Big(X_{v\wedge(t\wedge\tau(\omega))}(\omega) - X_{u\wedge(t\wedge\tau(\omega))}(\omega)\Big) =$$

$$= \sum_{[u,v)\in P} Y_u(\omega)\Big(X_{(v\wedge\tau(\omega))\wedge t}(\omega) - X_{(u\wedge\tau(\omega))\wedge t}(\omega)\Big) = \int_0^t Y_s \,\mathrm{d}X_{s\wedge\tau(\omega)}(\omega),$$

for all ω. For $Y \in \mathcal{L}_T^2$, let $(Y^{(n)})$ be a sequence approximating Y in $L^2(\mu_A)$. Since the claimed equality holds for all n, it remains to show that both sides converge to the claimed expressions in $L^2(\mathrm{d}\mathbb{P})$. For the left expression, let $M_t = \int_0^t Y_s \,\mathrm{d}X_s$ and $M_t^{(n)} = \int_0^t Y_s^{(n)} \,\mathrm{d}X_s$, and observe that for each ω,

$$\sup\Big\{\Big|M_{t\wedge\tau(\omega)}(\omega) - M^{(n)}_{t\wedge\tau(\omega)}(\omega)\Big| : t \leqslant T\Big\} \leqslant \sup\Big\{\Big|M_t(\omega) - M_t^{(n)}(\omega)\Big| : t \leqslant T\Big\}$$

simply because the supremum on the right hand side is over a larger set of numbers when $\tau(\omega) < T$. This implies that

$$\|(M_{\tau\wedge t})_{t\leqslant T} - (M^{(n)}_{\tau\wedge t})_{t\leqslant T}\|_{\mathcal{M}_T^2} \leqslant \|M - M^{(n)}\|_{\mathcal{M}_T^2} \stackrel{n\to\infty}{\longrightarrow} 0,$$

showing convergence of the approximations for the leftmost expression in the claim. For the middle expression, we first find that $(A_{t\wedge\tau})_{t\leqslant T}$ is the qvp of the stopped martingale $(X_{t\wedge\tau})_{t\leqslant T}$. This is true because $t \mapsto A_{t\wedge\tau(\omega)}(\omega)$ is increasing and continuous for all ω, and $(X^2_{t\wedge\tau} - A_{t\wedge\tau})_{t\leqslant T}$ is a martingale due to the optional stopping theorem. Moreover, we have

$$\mathbb{E}\Big(\int_0^t \big(Y_s - Y_s^{(n)}\big)^2 \,\mathrm{d}A_{s\wedge\tau}\Big) \leqslant \mathbb{E}\Big(\int_0^t \big(Y_s - Y_s^{(n)}\big)^2 \,\mathrm{d}A_s\Big) \stackrel{n\to\infty}{\longrightarrow} 0,$$

which means that $A^{(n)}$ also converges in $L^2(\mathrm{d}A_{\tau\wedge t} \otimes \mathbb{P})$. By definition of the stochastic integral (with respect to the martingale $(X_{t\wedge\tau})_{t\geqslant 0}$), this means that

$$\int_0^t Y_s^{(n)} \,\mathrm{d}X_{s\wedge\tau} \to \int_0^t Y_s \,\mathrm{d}X_{s\wedge\tau}$$

in $L^2(\mathrm{d}\mathbb{P})$ as $n \to \infty$, and the first equality is shown.

b) For the second equality, assume first that τ is discrete, i.e. $\tau(\Omega) = \{t_1, \dots, t_n\}$, and that $Y \in \mathcal{E}_T$. Then also $(Y_s \mathbb{1}_{[0,\tau)}(s))_{s \leqslant T} \in \mathcal{E}_T$, and the claim holds in the same way as in a) by writing both sides with a suitable partition P. Let now $Y \in \mathcal{L}^2_T$, but still τ discrete, and let $Y^{(n)} \to Y$ in $\mathcal{L}^2_T$. We already know that the middle expression converges under this approximation. For the right expression, it is enough to show $(Y^{(n)}_s \mathbb{1}_{[0,\tau)}(s))_{s \leqslant T}$ converges to $(Y_s \mathbb{1}_{[0,\tau)}(s))_{s \leqslant T}$ in $L^2(\mu_A)$, which is quite obvious. Let now τ be a stopping time and let $\tau_n \searrow \tau$ as $n \to \infty$ be a sequence of approximating discrete stopping times. The middle expression converges almost surely by path continuity of the stochastic integral (it is in $\mathcal{M}^2_T$). For the rightmost expression, the Itô-isometry gives

$$\Big\| \int_0^t Y_s \mathbb{1}_{[0,\tau^{(n)})}(s)\,\mathrm{d}X_s - \int_0^t Y_s \mathbb{1}_{[0,\tau)}(s)\,\mathrm{d}X_s \Big\|_{L^2} = \mathbb{E}\Big(\int_0^t Y_s^2 \underbrace{(\mathbb{1}_{[0,\tau^{(n)})}(s) - \mathbb{1}_{[0,\tau)}(s))^2}_{\to 0 \text{ a.s.}}\,\mathrm{d}A_s \Big),$$

which converges to zero by dominated convergence. This shows the second claimed equality. □

(3.23) Definition

A stochastic process Y is called **progressively measurable** with respect to a filtration $(\mathcal{F}_t)$ if

$$\forall A \in \mathcal{B}(\mathbb{R}^d), \forall t \leqslant T : \{(s,\omega) \in [0,t] \times \Omega : Y_s(\omega) \in A\} \in \mathcal{B}([0,t]) \otimes \mathcal{F}_t.$$

(3.24) Proposition

a) Every $Y \in \mathcal{E}_T$ is progressively measurable.

b) Let $X \in \mathcal{M}^2_T$ with qvp A. Then

$$\mathcal{L}^2_T = \{Y \in L^2(\mu_A) : Y \text{ has a representative that is progressively measurable}\}$$

Proof: a) is left as an exercise. b) is a bit more technical, but we will not do it here. See Theorem 14.23 of Schilling/Partzsch. □

(3.25) Definition

Let $X \in \mathcal{M}^2_T$ with qvp A.

a) We define

$$\mathcal{L}^0_T(X) := \{Y : Y \text{ progressively measurable}, \int_0^T Y_s^2(\omega)\,\mathrm{d}A_s(\omega) < \infty \text{ for } \mathbb{P}\text{-almost all } \omega\}.$$

b) Let $(\mathcal{F}_t)$ be a filtration and Y be an $(\mathcal{F}_t)$-adapted process. A *localizing sequence* (wrt. the filtration) is a is a sequence (σ_n) of $(\mathcal{F}_t)$-stopping times that fulfil $\sigma_{n+1} \geqslant \sigma_n$ almost surely, and $\lim_{n\to\infty} \sigma_n = \infty$ almost surely. Recall that

$$Y_t(\omega)\mathbb{1}_{[0,\sigma_n(\omega))}(t) = \begin{cases} Y_t(\omega) & \text{if } t \leqslant \sigma_n(\omega) \\ 0 & \text{otherwise.} \end{cases}$$

We define

$$\mathcal{L}^2_{T,\mathrm{loc}}(X) := \{Y : \text{ there exists a localizing sequence such that } (Y\mathbb{1}_{[0,\sigma_n)}) \text{ is in } \mathcal{L}^2_T \text{ for all } n.\}$$

(3.26) Example

With $Y_t = \exp(B_t^2)$, we have $(Y_t) \in \mathcal{L}^2_{T,\mathrm{loc}}$. We see this by choosing $\sigma_n(\omega) := \min\{t \geqslant 0 : |B_t(\omega)| \geqslant n\}$ and checking that this is a localizing sequence.

(3.27) Proposition

We have $\mathcal{L}^2_T(X) \subset \mathcal{L}^0_T(X) \subset \mathcal{L}^2_{T,\mathrm{loc}}(X)$.

Proof: exercise. Use the stopping times $\sigma_n := \inf\{t \geqslant 0 : \int_0^t Y_s^2(\omega)\,\mathrm{d}A_s(\omega) \geqslant n\}$ for the second implication. □

(3.28) The stochastic integral for integrands from $\mathcal{L}^2_{T,\mathrm{loc}}$

Let $X \in \mathcal{M}^2_T$ and $Y \in \mathcal{L}^2_{T,\mathrm{loc}}(X)$, with localizing sequence (σ_n). We define $Y^{(n)} = Y\mathbb{1}_{[0,\sigma_n)}(.)$. The Itô-integral $J_t^{(n)} := \int_0^t Y_s^{(n)}\,\mathrm{d}X_s$ then exists in $L^2(\mathbb{P})$ for all n. Now for each $t \leqslant T$ and all $n, k \in \mathbb{N}$, we have the following equalities (valid in $L^2(\mathbb{P})$):

$$\begin{aligned}
\mathbb{1}_{\{\sigma_n > T\}} J_t^{(n+k)} &= \mathbb{1}_{\{\sigma_n > T\}} \int_0^t Y_s^{(n+k)}\,\mathrm{d}X_s = \mathbb{1}_{\{\sigma_n > T\}} \int_0^{t\wedge\sigma_n} Y_s \mathbb{1}_{[0,\sigma_{n+k})}(s)\,\mathrm{d}X_s \overset{(3.22)}{=} \\
&= \mathbb{1}_{\{\sigma_n > T\}} \int_0^t Y_s \mathbb{1}_{[0,\sigma_{n+k})}(s)\mathbb{1}_{[0,\sigma_n)}(s)\,\mathrm{d}X_s = \mathbb{1}_{\{\sigma_n > T\}} \int_0^t Y_s \mathbb{1}_{[0,\sigma_n)}(s)\,\mathrm{d}X_s \\
&= \mathbb{1}_{\{\sigma_n > T\}} J_t^{(n)}.
\end{aligned}$$

This means that on $\{\sigma_n > T\}$, the sequence $(J^{(n+k)})$ becomes constant after the n-th term. We pick a representative $\omega \mapsto J^{(n)}(\omega)$ for each n and define

$$J_t(\omega) := \begin{cases} J_t^{(n)}(\omega) & \text{for } \omega \in \{\sigma_{n-1} \leqslant T < \sigma_n\}, \\ 0 & \text{for } \omega \in \{\lim_{n\to\infty} \sigma_n < \infty\}. \end{cases}$$

Since $\mathbb{P}(\lim \sigma_n < \infty) = 0$, we have

$$\lim_{n\to\infty} (J_t^{(n)})_{t\leqslant T} = (J_t)_{t\leqslant T} \qquad \mathbb{P}\text{-almost surely}$$

This defines the **generalized stochastic integral** for fixed $t \leqslant T$. You should check that J_t is independent of the localizing sequence in the sense that for two different localizing sequences, the results differ at most on a set of measure zero.

Note that the generalized stochastic integral (J_t) is *not* a martingale, in general, because we cannot guarantee $\mathbb{E}(|J_t|) < \infty$. However, $(J_t^{(n)}) = (J_{t\wedge\sigma_n})$ is a martingale for all n by Theorem (3.21). Such objects are of independent interest, and we define

(3.29) Definition

An $(\mathcal{F}_t)$-adapted, right-continuous stochastic process $(M_t)_{t \geqslant 0}$ is called a **local martingale** if there exists a localizing sequence such that $(M_{t\wedge\sigma_n} \mathbb{1}_{\{\sigma_n<\infty\}})_{t \geqslant 0}$ is a martingale for each n. We write $\mathcal{M}_{\mathrm{loc}}$ for the set of local martingales, and $\mathcal{M}^2_{T,\mathrm{loc}}$ for those local martingales where $(M_{t\wedge\sigma_n} \mathbb{1}_{\{\sigma_n<\infty\}})_{t \leqslant T}$ is in $\mathcal{M}^2_T$ for all n.

(3.30) Properties of generalized stochastic integrals

Let $X \in \mathcal{M}^2_T$ and $Y \in \mathcal{L}^2_{T,\mathrm{loc}}$.

a) The map $Y \mapsto \int_0^t Y_s \,\mathrm{d}X_s$ is linear from $\mathcal{L}^2_{T,\mathrm{loc}}$ to $\mathcal{M}^2_{T,\mathrm{loc}}$.

b) The process $\big((\int_0^t Y_s \,\mathrm{d}X_s)^2 - \int_0^t Y_s^2 \mathrm{d}A_s\big)_{t \leqslant T}$ is a local martingale.

c) $\mathbb{P}\big(\big|\int_0^t Y_s \,\mathrm{d}X_s\big| > \varepsilon\big) \leqslant \frac{C}{\varepsilon^2} + \mathbb{P}\big(\int_0^t |Y_s|^2 \mathrm{d}A_s > C\big)$ for all $\varepsilon > 0, C > 0$.

d) For a stopping time τ, the equality $\int_0^{t\wedge\tau} Y_s \,\mathrm{d}X_s = \int_0^t Y_s \mathbb{1}_{[0,\tau)}(s) \mathrm{d}X_s$ holds $\mathbb{P}$-almost surely.

Warning: There is no equivalent to the Itô-isometry for generalized stochastic integrals.

Proof: a), b) and d) are just chasing of definitions and are left as exercises. For c), consider the stopping time

$$\tau = \inf \Big\{ t \leqslant T : \int_0^t Y_s^2 \,\mathrm{d}A_s > C \Big\},$$

then $\mathbb{E}\big(\int_0^T (Y \mathbb{1}_{[0,\tau)}(s))^2 \,\mathrm{d}A_s\big) \leqslant C$ and thus $Y \mathbb{1}_{[0,\tau)} \in \mathcal{L}^2_T$. We have

$$\mathbb{P}\Big(\big|\int_0^t Y_s \,\mathrm{d}X_s\big| > \varepsilon\Big) \leqslant \mathbb{P}\Big(\big|\int_0^t Y_s \,\mathrm{d}X_s\big| > \varepsilon, \tau > t\Big) + \mathbb{P}(\tau \leqslant t).$$

The second term is equal to $\mathbb{P}\big(\int_0^t |Y_s|^2 \mathrm{d}A_s > C\big)$ and thus corresponds to the second term in the claim. The first term is equal to $\mathbb{P}\big(\big|\int_0^{t\wedge\tau} Y_s \,\mathrm{d}X_s\big| > \varepsilon, \tau > t\big) \leqslant \mathbb{P}\big(\big|\int_0^{t\wedge\tau} Y_s \,\mathrm{d}X_s\big| > \varepsilon)$, and by d), the Chebyshev intequality and the Itô-isometry applied to $Y \mathbb{1}_{[0,\tau)}$, we get

$$\mathbb{P}\Big(\big|\int_0^t Y_s \mathbb{1}_{[0,\tau)}(s) \,\mathrm{d}X_s\big| > \varepsilon\Big) \leqslant \frac{1}{\varepsilon^2} \mathbb{E}\Big(\big(\int_0^t Y_s \mathbb{1}_{[0,\tau)}(s) \,\mathrm{d}X_s\big)^2\Big) = \frac{1}{\varepsilon^2} \mathbb{E}\Big(\int_0^t (Y_s \mathbb{1}_{[0,\tau)}(s))^2 \,\mathrm{d}A_s\Big).$$

We have already seen that the expected value on the right hand side is bounded by C, so the claim follows. □

(3.31) Abbreviation

A right-continuous function f such that $\lim_{t\nearrow t_0} f(t)$ exists for all t_0 is called càdlàg, short for the French expression „continue à droite, limites à gauche“.

(3.32) Theorem

Let Y be an adapted stochastic process with càdlàg paths, let $X \in \mathcal{M}^2_T$ for all T, and assume that for almost all ω, the measure $\mu_{A,\omega} : f \mapsto \int_0^T f(s) \,\mathrm{d}A_s(\omega)$ is absolutely continuous with respect to the Lebesgue measure. Then $Y \in \mathcal{L}^2_{T,\mathrm{loc}}(X)$ for all $T < \infty$, and approximations to the generalized stochastic integral by arbitrary partitions converge uniformly in probability to the generalized stochastic integrals; this means that for all $\varepsilon > 0$,

$$\lim_{|P|\to 0} \mathbb{P}\Big(\sup_{0 \leqslant t \leqslant T} \Big| \sum_{[u,v)\in P\cap[0,t)} Y_u X_{u,v} - \int_0^t Y_s \, \mathrm{d}X_s \Big| > \varepsilon \Big) = 0.$$

Proof: Càdlàg functions are locally bounded and $A_T < \infty$ almost surely, thus $\int_0^T Y_s^2 \, \mathrm{d}A_s < \infty$ almost surely. This means that $Y \in \mathcal{L}_T^0$ and thus $Y \in \mathcal{L}^2_{T,\mathrm{loc}}$ by (3.25). We define the localizing sequence

$$\sigma_n(\omega) := \inf \big\{ s \geqslant 0 : |Y_s(\omega)|^2 > n \big\} \wedge n,$$

and $Y_s^{(n)}(\omega) = Y_s(\omega) \mathbb{1}_{[0,\sigma_n(\omega))}(s)$. As in the proof of (3.30 c), we find that

$$\mathbb{P}\Big(\sup_{0 \leqslant t \leqslant T} \Big| \sum_{[u,v)\in P\cap[0,t)} Y_u X_{u,v} - \int_0^t Y_s \, \mathrm{d}X_s \Big| > \varepsilon \Big) \leqslant$$
$$\leqslant \frac{1}{\varepsilon^2} \mathbb{E}\Big(\sup_{0 \leqslant t \leqslant T} \Big| \sum_{[u,v)\in P\cap[0,t)} Y_u^{(n)} X_{u,v} - \int_0^t Y_s^{(n)} \, \mathrm{d}X_s \Big|^2 \Big) + \mathbb{P}(\sigma_n < T).$$

Since σ_n is a localizing sequence, it converges to infinity almost surely, and thus for each $\delta > 0$ there exists $n \in \mathbb{N}$ with $\mathbb{P}(\sigma_n < T) \leqslant \delta$. The proof will thus be finished once we show that the first term converges to zero for each n when $|P| \to 0$.

To show this, note that $Y^{(n)} \in \mathcal{L}_T^2$ because it is bounded, and so $\int_0^t Y_s^{(n)} \, \mathrm{d}X_s$ exists as a classical Itô-integral for all $t \leqslant T$ and all $n \in \mathbb{N}$. For a partition P of $[0,T]$, let us write

$$Y_s^{(n),P}(\omega) := \sum_{[u,v)\in P} Y_u^{(n)}(\omega) \mathbb{1}_{[u,v)}(s),$$

for the „Riemann approximation“ of $Y^{(n)}$ with respect to P, then

$$\sum_{[u,v)\in P\cap[0,t)} Y_u^{(n)} X_{u,v} = \int_0^t Y_s^{(n),P} \, \mathrm{d}X_s,$$

and we have

$$\mathbb{E}\Big(\sup_{0 \leqslant t \leqslant T} \Big| \sum_{[u,v)\in P\cap[0,t)} Y_u^{(n)} X_{u,v} - \int_0^t Y_s^{(n)} \, \mathrm{d}X_s \Big|^2 \Big) = \mathbb{E}\Big(\sup_{0 \leqslant t \leqslant T} \Big| \int_0^t (Y_s^{(n),P} - Y_s^{(n)}) \, \mathrm{d}X_s \Big|^2 \Big) \overset{\text{Doob}}{\leqslant}$$
$$\leqslant 4\mathbb{E}\Big(\Big(\int_0^T (Y_s^{(n),P} - Y_s^{(n)}) \, \mathrm{d}X_s \Big)^2 \Big) \overset{\text{Itô}}{=} 4\mathbb{E}\Big(\int_0^T \big(Y_s^{(n),P} - Y_s^{(n)} \big)^2 \, \mathrm{d}A_s \Big).$$

The integrand in the final expression is bounded by $4n^2$. For a sequence of partitions (P_m) with $P_m \subset P_{m+1}$, we have $Y_s^{(n),P_m}(\omega) \to Y_s^{(n)}(\omega)$ for all $s \in [0,T]$ and all ω as $m \to \infty$. To see this note that for points of continuity of $s \mapsto Y_s^{(n)}(\omega)$, the convergence obviously holds. Now one can show that a càdlàg function can only have countably many discontinuities, and by assumption, the measure $\mathrm{d}A_t(\omega)$ is absolutely continuous with respect to Lebesgue measure almost surely. Thus, almost surely, $\int \mathbb{1}_{B_\omega} \, \mathrm{d}A_s(\omega) = 0$, where B_ω is the set of $s \in [0,T]$ where the convergence fails. Now Fubinis theorem shows the claimed almost sure pointwise convergence. Dominated convergence now shows that the final expression in the last display converges to zero as $|P| \to 0$, and finishes the proof. $\square$

(3.33) Remark

Up to now, we have mostly considered the existence (and approximability) of stochastic integrals. This is necesseary, but is only interesting if we can do something useful with the objects we obtain. A useful theory should enable us to actually calculate something of interest, for example:

a) We have already seen that $\int_0^t B_s \, \mathrm{d}B_s = \frac{1}{2}B_t^2 - \frac{1}{2}t$. But what about $\int_0^t B_s^k \, \mathrm{d}B_s$ for $k > 1$? Is the solution also given by the classical term $\frac{1}{k+1}B_t^{k+1}$ + „a correction“? If so, what is this correction? Is there closed formula for all k?

b) Eventually, we want to solve stochastic differential equations. In classical ODE theory, a very nice way to solve an ODE is just to guess the solution and then check that it actually is the solution. Let us assume we have a process Y_t and we guess that it fulfils the equation

$$Y_t - Y_0 = \int_0^t f(Y_s) \, \mathrm{d}B_s$$

for some smooth function f. How can we check this?

In classical analysis, the answer to both questions is the chain rule. If we replace $t \mapsto B_t$ by a differentiable function $t \mapsto h(t)$ with $h(0) = 0$, then in problem a), the chain rule give $\partial_t h^{k+1}(t) = (k+1)h^k(t)h'(t)$, and by the fundamental theorem of calculus this gives the answer

$$\int_0^t h^k(s) \, \mathrm{d}h(s) = \int_0^t h^k(s)h'(s) \, \mathrm{d}s = \frac{1}{k+1}h^{k+1}(t).$$

The differential equation from problem b) in classical analysis reads $u'(t) = f(u(t))h'(t)$, where we replaced the process Y by the classical function u and ignored the initial condition. The solution by separation of variables, in the most general form, again relies on the chain rule: assume that we are able to find a primitive (antiderivative) $t \mapsto K(t)$ of $t \mapsto 1/f(t)$, and let us assume that for the (given) function h we can find the solution u to the equation $h(t) = K(u(t))$ for all t, and that it is a differentiable function. Then by the chain rule, this solution u satisfies

$$h'(t) = \partial_t K(u(t)) = \frac{1}{f(u(t))}u'(t),$$

which gives a solution of the ODE by rearranging. Of course, many things can go wrong here, but for example the solution to the equation $u'(t) = \frac{1}{u(t)^2}h'(t)$ can then be computed to be $u(t) = (3h(t))^{1/3}$, and one easiy checks (again by the chain rule) that this is indeed a solution (of course h should be nonnegative here).

The problem is that for stochastic integrals, as we have seen the chain rule is not valid. But the good news is that there is a replacement, which is the Itô-formula. It is the most important formula for stochastic integrals, and we will discuss it now.

(3.34) The Itô-formula (elementary version)

Let $X \in \mathcal{M}_T^2$ with qvp A, $f \in C^2(\mathbb{R})$. Then the map $(s, \omega) \mapsto f'(X_s(\omega))$ is an element of

$\mathcal{L}^2_{T,\mathrm{loc}}(X)$, and for all t, the **Itô-formula**

$$f(X_t) - f(X_0) = \int_0^t f'(X_s)\,\mathrm{d}X_s + \frac{1}{2}\int_0^t f''(X_s)\,\mathrm{d}A_s$$

holds almost surely. For the case $X = B$, the Itô-formula specializes to

$$f(B_t) - f(B_0) = \int_0^t f'(B_s)\,\mathrm{d}B_s + \frac{1}{2}\int_0^t f''(B_s)\,\mathrm{d}s.$$

Proof: Clearly $f'(X) \in \mathcal{L}^2_{T,\mathrm{loc}}$ since it is adapted, continuous in s, and locally bounded. The idea of the proof of the Itô formula is not difficult: first, we write the left hand side of Itôs formula as a telescopic sum along a partition P of $[0,t)$:

$$f(X_t) - f(X_0) = \sum_{[u,v)\in P} (f(X_v) - f(X_u)).$$

Since X is continuous, X_v will be close to X_u for each interval $[u,v)$ once we send $|P| \to 0$. Therefore the next natural step is to approximate each term $(f(X_v) - f(X_u))$ on the right hand side by a Taylor approximation. The only thing we have to be careful about is that we actually expand the function f around the point X_u and not the function $u \mapsto f(X_u)$ around the point u, because the latter will usually not even be differentiable. Also, it will turn out that we have to expand to second order. The result is

$$f(X_v) - f(X_u) = f(X_u) + f'(X_u)X_{u,v} + \frac{1}{2}f''(X_u)X_{u,v}^2 + \Delta_{u,v} - f(X_u),$$

where

$$\Delta_{u,v} = \int_0^1 \mathrm{d}\rho \int_0^\rho \mathrm{d}\nu \big(f''(X_u + \nu X_{u,v}) - f''(X_u)\big) X_{u,v}^2$$

is the Taylor remainder term in a slightly unusual integral form. The explicit „raw form“ that does not need any extra differentiability beyond C^2. Consequently,

$$(*) \qquad f(X_t) - f(X_0) = \sum_{[u,v)\in P} f'(X_u)X_{u,v} + \frac{1}{2}\sum_{[u,v)\in P} f''(X_u)X_{u,v}^2 + \sum_{[u,v)\in P} \Delta_{u,v}.$$

The first term on the right hand side converges to $\int_0^t f'(X_s)\,\mathrm{d}X_s$ in probability as $|P| \to 0$ by Theorem (3.32), yielding the first term of the Itô formula. For the other terms, we first assume, as we did in the proof of Theorem (2.32) that X is uniformly bounded, i.e. $|X_s(\omega)| \leqslant K$ for some K, all s and all ω. Note that this also implies that $\|f'' \circ X\|_\infty = \{\sup\{|f''(X_s(\omega))| : s \leqslant t, \omega \in \Omega\} < \infty$. Also like in Theorem (2.32), let $P_n(t)$ be the partition generated by the dyadic rationals of spacing 2^{-n} and the point t. We know that

$$\lim_{n\to\infty} \sum_{[u,v)\in P_n(t)} f''(X_u(\omega))A_{u,v}(\omega) = \int_0^t f''(X_s(\omega))\,\mathrm{d}A_s(\omega)$$

for all ω, and are thus interested in the L^2-distance between this expression and the second term of $(*)$. We have

$$\mathbb{E}\Big(\big(\sum_{[u,v)\in P_n(t)} f''(X_u)(X_{u,v}^2 - A_{u,v})\big)^2\Big) \leqslant \|f''\circ X\|_\infty^2 \mathbb{E}\Big(\big(\sum_{[u,v)\in P_n(t)} (X_{u,v}^2 - A_{u,v})\big)^2\Big)$$
$$= \|f''\circ X\|_\infty^2 \mathbb{E}\big((A_t^{(n)} - A_t)^2\big),$$

where $A_t^{(n)} = \sum_{[u,v)\in P_n(t)} X_{u,v}^2$ is precisely the approximation from Theorem (2.32 b). The statement of that Theorem says that this converges to zero as $n\to\infty$, and so $\sum_{[u,v)\in P_n} f''(X_u)X_{u,v}^2 \to \int_0^t X_s^2 \,\mathrm{d}A_s$ almost surely along a subsequence. Picking this subesquence gives us the second term in Itôs formula.

For the third term, we set

$$D(\omega,u,v,\nu) := |f''(X_u(\omega) + \nu X_{u,v}(\omega)) - f''(X_u(\omega))|$$

and estimate (again with the partition from above, more precisely with the subsequence from the last step)

$$\Big|\sum_{[u,v)\in P_n(t)} \Delta_{u,v}(\omega)\Big| \leqslant \underbrace{\sup\{D(\omega,u,v,\nu) : |u-v|\leqslant 2^{-n}, 0\leqslant\nu\leqslant 1\}}_{=:\delta_n(\omega)} \sum_{[u,v)\in P_n(t)} X_{u,v}^2,$$

and applying the Cauchy-Schwarz-inequality gives us

$$\mathbb{E}\Big(\Big|\sum_{[u,v)\in P_n(t)} \Delta_{u,v}(\omega)\Big|\Big)^2 \leqslant \mathbb{E}(\delta_n^2)\mathbb{E}\Big(\big(\sum_{[u,v)\in P_n(t)} X_{u,v}^2\big)^2\Big).$$

We have seen in the proof of Theorem (2.32) that the second factor on the right hand side above is bounded independent of n. The integrand in the first factor is bounded by $4|\sup_{|x|\leqslant 3K} f''(x)|^2$ for all n and converges to zero pointwise by uniform continuity of $s\mapsto X_s(\omega)$ and of f on compact subsets of $\mathbb{R}$. It therefore converges to zero by dominated convergence. It follows that $\big|\sum_{[u,v)\in P_n(t)} \Delta_{u,v}(\omega)\big|$ converges to zero in L^1, and thus almost surely along a (further) subsequence. Running through this final subsequence in the expression $(*)$ now gives Itô's formula for bounded X.

For unbounded X, we have to localize again. Let $\tau_n = \inf\{s\geqslant 0 : |X_s|\geqslant n\}$. Since Doobs inequality gives $\mathbb{E}(\sup_{s\leqslant t} X_s^2) \leqslant 4\mathbb{E}(X_t^2) < \infty$, we have $\mathbb{P}(\tau_n\leqslant t)\to 0$ as $n\to\infty$ for all t, and so $\tau_n\to\infty$ almost surely. Now for the bounded martingale $(X_s^{(n)})_{s\leqslant t} := (X_{\tau\wedge s})_{s\leqslant t}$ with qvp $A^{(n)} = (A_{s\wedge\tau_n})_{s\leqslant t}$, we find by the first part of the proof

$$f(X_t^{(n)}) - f(X_0^{(n)}) = \int_0^t f'(X_s^{(n)})\,\mathrm{d}X_s^{(n)} + \frac{1}{2}\int_0^t f''(X_s^{(n)})\,\mathrm{d}A_s^{(n)} =$$
$$= \int_0^{t\wedge\tau_n} f'(X_s)\,\mathrm{d}X_s + \frac{1}{2}\int_0^{t\wedge\tau_n} f''(X_s)\,\mathrm{d}A_s$$

So Itôs formula holds for all $\omega\in\{\tau_n > t\}$ and all n, and thus for all $\omega\in\bigcup_{n\in\mathbb{N}}\{\tau_n > t\} \supset \{\lim_{n\to\infty}\tau_n = \infty\}$. This shows the claim. □

(3.35) Examples

a) For $f(x) = x^2/2$, we get

$$\frac{1}{2}B_t^2 = f(B_t) - f(B_0) = \int_0^t f'(B_s)\,\mathrm{d}B_s + \frac{1}{2}\int_0^t f''(B_s)\,\mathrm{d}s = \int_0^t B_s\,\mathrm{d}B_s + \frac{1}{2}t.$$

Rearranging shows the known formula $\int_0^t B_s\,\mathrm{d}B_s = \frac{1}{2}B_t^2 - \frac{1}{2}t$.

b) The **geometric Brownian motion** $Y_t = \mathrm{e}^{\alpha B_t}$ (with $\alpha > 0$) is an important stochastic process in financial mathematics. It solves the SDE

$$Y_t - Y_0 = \alpha \int_0^t Y_s\,\mathrm{d}B_s + \frac{\alpha^2}{2}\int_0^t Y_s\,\mathrm{d}s.$$

(Exercise using Itôs formula). Formally writing $\mathrm{d}B_s = B_s'\,\mathrm{d}s$ and differentiating this SDE gives

$$Y_t' = \alpha Y_t B_t' + \frac{\alpha^2}{2}Y_t = \frac{\alpha^2}{2}Y_t(1 + \frac{2}{\alpha}\text{„noise“}),$$

compare this to the ODE $f'(t) = \frac{\alpha^2}{2}f(t)$, which models the exponential growth of the price of an asset e.g. under constant inflation or with constant value gain. So Y_t is the analogue when there is in addition some market movement (noise) that changes the rate of price growth either up or down.

Note that we can *only* get growing solutions with $\mathrm{e}^{\alpha B_t}$, it does not help to put $\mathrm{e}^{-\alpha B_t}$ since (by symmetry of Brownian motion, or by calculation) the result will be the same. So what about the very important ODE $f' = -\frac{\alpha^2}{2}f$ when we perturb it by some noise? We will see the answer later.

(3.36) The elementary Itô formula in d dimensions

Let B be a d-dimensional Brownian motion, and $f \in C^{1,2}(\mathbb{R}_0^+ \times \mathbb{R}^d, \mathbb{R})$, i.e. once continuously differentiable in the first variable and twice in the second. Then

$$f(t, B_t) - f(0, B_0) = \int_0^t [\partial_s f](s, B_s)\,\mathrm{d}s + \int_0^t [\nabla_x f](s, B_s)\cdot\mathrm{d}B_s + \frac{1}{2}\int_0^t [\Delta_x f](s, B_s)\,\mathrm{d}s,$$

where $\nabla_x f = (\partial_{x_1} f, \dots, \partial_{x_d} f)$, and $\Delta_x = \sum_{i=1}^d \partial_x^2$.

Proof: Analogous to the proof of (3.34), but somewhat more notation-intense... □

(3.37) Remark

Compare Theorem (3.36) to Theorem (2.16): the latter tells us that

$$M_t = f(t, B_t) - f(t, B_0) - \int_0^t \big([\tfrac{1}{2}\Delta_x f](s, B_s) + [\partial_s f](s, B_s)\big)\,\mathrm{d}s$$

is a martingale; Theorem (3.36) actually tells us what martingale it is!

(3.38) Differential Notation

a) SDE are often written in *differential notation*:

$$\mathrm{d}Y_t = b(t, Y_t)\,\mathrm{d}t + \sigma(t, Y_t)\,\mathrm{d}B_t \qquad (*)$$

Integrating $(*)$ from 0 to T then gives the familiar form

$$Y_T - Y_0 = \int_0^T \mathrm{d}Y_t = \int_0^T b(Y_t, t)\,\mathrm{d}B_t + \int_0^T \sigma(t, B_t)\,\mathrm{d}B_t.$$

Stochastic processes of the form $(*)$ are called **Itô processes**, for the precise definition see below.

b) Itô's formula in differential form reads: for $Y_t = f(t, B_t)$, B Brownian motion,

$$\mathrm{d}Y_t = [\partial_t f](t, B_t)\,\mathrm{d}t + \frac{1}{2}[\Delta_x f](t, B_t)\,\mathrm{d}t + [\nabla f](t, B_t) \cdot \mathrm{d}B_t.$$

If $X \in \mathcal{M}^2$ with qvp A, $Y_t = f(t, X_t)$, in one dimension we get

$$\mathrm{d}Y_t = [\partial_t f](t, X_t)\,\mathrm{d}t + \frac{1}{2}[\partial_x^2 f](t, X_t)\,\mathrm{d}A_t + [\partial_x f](t, X_t)\,\mathrm{d}X_t.$$

A multi-dimensional analogue exists, but needs the concept of quadratic covariation which we will not cover in this lecture.

(3.39) Itô-processes

Let B be a d-dimensional Brownian motion with its filtration $(\mathcal{F}_t)$. Let $(\sigma_t)_{t \geqslant 0}$ be an $\mathbb{R}^{m \times d}$-valued stochastic process, and let $(b_t)_{t \geqslant 0}$ be an $\mathbb{R}^m$-valued stochastic process. Assume that both processes are progressively measurable and have locally bounded paths. The $\mathbb{R}^m$-valued stochastic process Y with

$$\mathrm{d}Y_t = \sigma_t\,\mathrm{d}B_t + b_t\,\mathrm{d}t \equiv \Big(\sum_{j=1}^d (\sigma_t)_{i,j}\,\mathrm{d}B_t^{(j)} + (b_t)_i\,\mathrm{d}t \Big)_{i=1,\dots,m}$$

is called the **Itô process** with **drift** b and **diffusion coefficient** σ.

(3.40) Theorem: general Itô formula

Let Y be an $\mathbb{R}^m$-valued Itô process with m-dimensional drift b and $m \times d$-dimensional diffusion coefficient σ, and let $f \in C^{1,2}(\mathbb{R}_0^+ \times \mathbb{R}^m, \mathbb{R})$. For $Z_s(\omega) := f(Y_s(\omega))$ we get

$$\mathrm{d}Z_t = \partial_t f(t, Y_t)\,\mathrm{d}t + \frac{1}{2}\mathrm{tr}\big[\sigma_t^* [D^2 f](t, Y_t)\sigma_t\big]\,\mathrm{d}t + [\nabla_x f](t, Y_t) \cdot \mathrm{d}Y_t,$$

where σ^* is the transpose of σ, $D^2 f$ is the $m \times m$-matrix with entries $\partial_i \partial_j f$ (i.e. the Hessian of f), and

$$[\nabla_x f](t, Y_t)\,\mathrm{d}Y_t = \underbrace{[\nabla_x f](t, Y_t)}_{\in \mathbb{R}^m} \cdot \underbrace{[\sigma_t \mathrm{d}B_t]}_{\in \mathbb{R}^m} + [\nabla_x f](t, Y_t) \cdot b_t\,\mathrm{d}t.$$

For $m = 1$, this simplifies to

$$\mathrm{d}Z_t = [\partial_t f](t, Y_t)\,\mathrm{d}t + [\partial_x f](t, Y_t)\,\mathrm{d}Y_t + \frac{1}{2}\mathrm{tr}(\sigma_t^* \sigma_t)[\partial_x^2 f](t, Y_t)\,\mathrm{d}t,$$

where $\mathrm{tr}(\sigma_t^*\sigma_t)$ is just a fancy way of writing the squared norm $\sum_{i=1}^d (\sigma_t)_i^2$ of the d-dimensional vector σ_t.

Proof (very short sketch): The proof is not so different from the one of the elementary Itô formula: one has to do a Taylor expansion of f around the points Y_u of a partition, and collect terms of first and second order.
a) Terms containing $(\mathrm{d}B_t^{(i)})^2$ converge to terms with $\mathrm{d}t$.
b) Terms containing $\mathrm{d}B_t^{(i)}\mathrm{d}B_t^{(j)}$ vanish when $i \neq j$.
c) Terms containing $\mathrm{d}B_t^{(i)}\mathrm{d}t$ or $(\mathrm{d}t)^2$ also vanish.
For purposes of calcualation, the following table summarizes these facts in a useful way:

	$\mathrm{d}B_t^{(i)}$	$\mathrm{d}t$
$\mathrm{d}B_t^{(j)}$	$\mathrm{d}t\delta_{i,j}$	0
$\mathrm{d}t$	0	0

(3.41) Exercise

Consider the linear SDE

$$dY_t = b(t)Y_t\,\mathrm{d}t + \sigma(t)Y_t\,\mathrm{d}B_t$$

(note that b and σ do not depend on ω). Use the Itô formula with the Ansatz

$$Y_t = \exp\Big(\int_0^t q(s)\,\mathrm{d}s + \int_0^t r(s)\,\mathrm{d}B_s\Big) \quad (*)$$

to show that $(*)$ is a solution to the SDE when $r(t) = \sigma(t)$ and $q(t) = b(t) - \sigma(t)^2$.

(3.42) The Ornstein-Uhlenbeck porcess

Consider the linear SDE

$$\mathrm{d}Y_t = -\alpha Y_t\,\mathrm{d}t + \sigma\,\mathrm{d}B_t, \qquad (\alpha, \sigma > 0).$$

(Why is this not a special case of the previous example?) The corresponding ODE is

$$\partial_t f(t) = -\alpha f(t) + \sigma\partial_t h(t)$$

with a smooth function h, and by the variation of constants formula it has the solution

$$f(t) = \sigma\int_0^t \mathrm{e}^{-\alpha(t-s)}\,(\partial_s h)(s)\,\mathrm{d}s + \mathrm{e}^{-\alpha t}\,f(0) = \sigma\int_0^t \mathrm{e}^{-\alpha(t-s)}\,\mathrm{d}h(s) + \mathrm{e}^{-\alpha t}\,f(0).$$

Observe that in the final integral, it is no problem to replace $\mathrm{d}h(s)$ by $\mathrm{d}B_s$, since the integrand is smooth and thus the Young integral exists. Therefore, it is a safe bet to guess that the solution to the SDE is

$$Y_t = \sigma\int_0^t \mathrm{e}^{-\alpha(t-s)}\,\mathrm{d}B_s + \mathrm{e}^{-\alpha t}\,Y_0.$$

Of course, we have to verify that this indeed gives a solution to the SDE, and we do this by using the Itô formula: we define $Z_t := \int_0^t e^{\alpha s}\,dB_s$ and $g(t,x) = e^{-\alpha t}(\sigma x + Y_0)$. Then $Y_t = g(t, Z_t)$, and

$$\begin{aligned} dY_t &= \partial_t g(t,Z_t)\,dt + \partial_x g(t,Z_t)\,dZ_t + \underbrace{\partial_x^2 g(t,Z_t)}_{=0}(dZ_t)^2 \\ &= -\alpha g(t,Z_t)\,dt + \sigma\, e^{-\alpha t}\,dZ_t = -\alpha Y_t\,dt + \sigma\, e^{-\alpha t}\, e^{\alpha t}\,dB_t. \end{aligned}$$

So Y indeed solves the SDE.

(3.43) Levy's characterisation of Brownian motion

Let X be an $(\mathcal{F}_t)$-adapted, real-valued stochastic process with continuous paths, and with $X_0 = 0$ almost surely. Then the following two statements are equivalent:

(i): $(X_t)_{t \geqslant 0}$ is a Brownian motion.

(ii): $(X_t)_{t \geqslant 0}$ and $(X_t^2 - t)_{t \geqslant 0}$ are martingales.

Proof: We already know that (i) implies (ii). To show the converse, we check the axioms (B0)-(B4) of Brownian motion. (B0) (which is $X_0 = 0$ a.s) and (B4) (which is continuous paths) hold by assumption. (B1)-(B3) say that X has stationary, independent, Gaussian increments, and all of this will be shown once we prove that (ii) implies

$$(*) \qquad \mathbb{E}(e^{i\xi X_{s,t}}\,\mathbb{1}_A) = e^{-\frac{1}{2}(t-s)^2\xi^2}\,\mathbb{P}(A) \qquad \forall \xi \in \mathbb{R}, A \in \mathcal{F}_s.$$

Why is $(*)$ enough? Because the choice $A = \Omega$ gives the Gaussian distribution (B3) of the increment and its stationarity (B2), and for the independent increment property (B1), we replace $e^{-\frac{1}{2}(t-s)^2\xi^2}$ by $\mathbb{E}(e^{i\xi X_{s,t}})$ (which is true by $(*)$) and obtain

$$\mathbb{E}(e^{i\xi X_{s,t}}\,\mathbb{1}_A) = \mathbb{E}(e^{i\xi X_{s,t}})\mathbb{P}(A) \qquad \forall \xi \in \mathbb{R}, A \in \mathcal{F}_s.$$

Since any bounded measurable function can be approximated by linear combinations of complex exponentials (this is the reason why characteristic functions determine the distribution!), we conclude that $\mathbb{E}(g(X_{s,t})\mathbb{1}_A) = \mathbb{E}(g(X_{s,t}))\mathbb{P}(A)$ for all $A \in \mathcal{F}_s$ and all bounded measurable functions. This means that $X_{s,t} \perp\!\!\!\perp \mathcal{F}_s$, which shows the claimed independence.

To show $(*)$, first note that (ii) implies that the qvp of X is $A_t = t$. Itô's formula applied to $f(x) = e^{i\xi x}$ (with $f'(x) = i\xi\, e^{i\xi x}$ and $f''(x) = -\xi^2 f(x)$) then gives

$$f(X_t) - f(X_s) = i\xi \int_s^t e^{i\xi X_r}\,dX_r - \frac{\xi^2}{2}\int_s^t e^{i\xi X_r}\,\underbrace{dA_r}_{=dr}.$$

We multiply this with $e^{i\xi X_s}\,\mathbb{1}_A$ with $A \in \mathcal{F}_s$ and take expectations on both sides, leading to

$$\underbrace{\mathbb{E}(e^{i\xi X_{s,t}}\,\mathbb{1}_A)}_{=:\varphi(t)} - \mathbb{P}(A) = \mathbb{E}\Big(\underbrace{e^{i\xi X_s}\,\mathbb{1}_A}_{\in m\mathcal{F}_s}\underbrace{\mathbb{E}\big(i\xi \int_s^t e^{i\xi X_r}\,dX_r \,\big|\, \mathcal{F}_s\big)}_{=0,\text{ martingale}}\Big) - \frac{\xi^2}{2}\int_s^t \underbrace{\mathbb{E}(e^{i\xi X_{s,r}}\,\mathbb{1}_A)}_{=\varphi(r)}\,dr.$$

We conclude that the function φ solves the ODE $\varphi'(t) = -\frac{\xi^2}{2}\varphi(t)$ with initial condition $\varphi(s) = \mathbb{P}(A)$. This ODE has the unique solution $\varphi(t) = e^{-\xi^2(t-s)/2}\,\mathbb{P}(A)$, which shows that $(*)$ holds. □

(3.44) Remark

Question: What is the difference between a one-dimensional Brownian motion and a general (local) martingale M?

Answer: The speed at which it runs through its paths.

Explanation: For each path, the quantity $A_t(\omega) = \lim_{|P|\to 0} \sum_{[u,v)\in P} M_{u,v}^2(\omega)$ (the limit actually might not exist pathwise, but does exist in L^2) can be thought of the „distance" that the path has covered up to time t. For smooth functions, that distance would be given by the total variation $\lim_{|P|\to 0} \sum_{[u,v)\in P} |f(v) - f(u)|$, but for continuous martingales this quantity is always infinite (unless M is constant), and so the quadratic variation is the next best quantity. More precisely, it measures the total variance (in the sense of the central limit theorem) that the martingale has picked up by time t.

For Brownian motion, we have $A_t(\omega) = t$ for all ω, so all paths run „at the same speed", which is given by $A'(t) = 1$.

For a general martingale M, $A_t(\omega)$ usually depends on the path belonging to ω. So we can try to individually slow down each path when $A_t'(\omega) > 1$, and to speed it up when $A'(t) < 1$. When we do this path by path, we should get back Brownian motion. The next theorem shows that this strategy works, at least when the variance of M_t diverges when t goes to infinity.

(3.45) Theorem (Döblin 1940, Dambis, Dubins, Schwartz 1965)

Let M be a continuous square integrable martingale for the right-continuous filtration $(\mathcal{F}_t)$. Let A be the qvp of M. Assume that $\lim_{t\to\infty} A_t = \infty$ $\mathbb{P}$-almost surely. Let $\tau_s(\omega)$ be the generalized inverse of $A_t(\omega)$, i.e.

$$\tau_s(\omega) := \inf\{t \geqslant 0 : A_t(\omega) > s\}$$

Then

a) τ_s is an $(\mathcal{F}_t)$-stopping time for all s, and $\tau_s \geqslant \tau_r$ when $s \geqslant r$.

b) Let $\mathcal{F}_{\tau_s} := \{B \in \mathcal{F} : B \cap \{\tau_s < t\} \in \mathcal{F}_t \,\forall t\}$ be the stopped sigma-algebra for the stopping time τ_s, and $\mathcal{G}_s := \sigma(\mathcal{F}_{\tau_r} : r \leqslant s)$. Then the stochastic process $(M_{\tau_s})_{s\,\geqslant\,0}$ is a $(\mathcal{G}_s)$-martingale, and its qvp is given by $A_{\tau_s} = s$.

c) The stochastic process $\tilde{B}_s := M_{\tau_s}$ is a Brownian motion for the filtration $(\mathcal{G}_s)_{s\,\geqslant\,0}$.

d) For all $\omega \in \Omega$ and all $r \geqslant 0$, we have $M_t(\omega) = \tilde{B}_{A_t(\omega)}(\omega)$.

Proof: a) The stopping time property follows from Lemma 2.22 since $A_t \in \mathrm{m}\mathcal{F}_t$ and $\mathcal{F}_t = \mathcal{F}_{t+}$. The inequality is clear from the definition.

b) We have $\mathbb{P}(\tau_t < \infty) = \mathbb{P}(\lim_{r\to\infty} A_r = \infty) = 1$. For those ω with $\tau_t(\omega) < \infty$, the equality $A_{\tau_t(\omega)}(\omega) = t$ by path continuity of A. Since $\mathbb{E}(A_{\tau_t}) = t < \infty$, Theorem 2.33 gives $\mathbb{E}(M_{\tau_t} \,|\, \mathcal{F}_{\tau_s}) = M_{\tau_s}$ and $\mathbb{E}(M_{\tau_t}^2 - A_{\tau_t} \,|\, \mathcal{F}_{\tau_s}) = M_{\tau_s}^2 - A_{\tau_s}$, so $(M_{\tau_s})_{s\,\geqslant\,0}$ and $(M_{\tau_s}^2 - A_{\tau_s})_{s\,\geqslant\,0} = (M_{\tau_s}^2 - s)_{s\,\geqslant\,0}$ are $(\mathcal{G}_s)$-martingales. The uniqueness of the qvp completes the claim.

c) This follows from b) and Theorem 3.43 if we can show that $\tilde{B}$ has continuous paths. The map $s \mapsto \tau_s(\omega)$ is not necessarily continuous; if $s \mapsto A_s(\omega)$ is constant on some interval $[a, b]$, then $\tau_\cdot(\omega)$ has a jump of size $b - a$ at $s = A_a(\omega)$. However, in this case also $s \mapsto M_s(\omega)$ is constant on $[a, b]$ (why?), and so $s \mapsto \tilde{B}_s(\omega) = M_{\tau_s(\omega)}(\omega)$ is still continuous.

d) We have $\tilde{B}_{A_t(\omega)}(\omega) = M_{\tau_{A_t(\omega)}(\omega)}(\omega)$ and $\tau_{A_t(\omega)}(\omega) = \inf\{r \geqslant 0 : A_r(\omega) > A_t(\omega)\}$. If $A_.(\omega)$ is strictly increasing on an interval $[t,b]$, then $\tau_{A_t(\omega)}(\omega) = t$, showing the claim. If $A_.(\omega)$ is constant on an interval $[t,b]$ and strictly increasing after that, then $\tau_{A_t(\omega)}(\omega) = b$. However, then $M_.(\omega)$ is also constant on $[t,b]$, and thus $M_{\tau_{A_t(\omega)}(\omega)}(\omega) = M_b(\omega) = M_t(\omega)$, showing the claim also in this case. □

Remark: The theorem even holds for the case where M is only a local martingale. We did not develop enough theory of local martingales to give an effortless proof, so we skip it.

(3.46) Preparations

A function $f : \mathbb{C} \to \mathbb{C}$ is complex differentiable (analytic) in $z = x + \mathrm{i}y \in \mathbb{C}$ if (with $f(z) = u(x,y) + \mathrm{i}v(x,y)$ for functions $u, v : \mathbb{R}^2 \to \mathbb{R}$) the **Cauchy-Riemmann differential equations** (CRE)

$$\partial_x u(x,y) = \partial_y v(x,y), \qquad \partial_y u(x,y) = -\partial_x v(x,y)$$

hold. The function

$$f' : \mathbb{C} \to \mathbb{C}, \qquad x + \mathrm{i}y \mapsto \partial_x[u + \mathrm{i}v](x,y) = -\mathrm{i}\partial_y[u + \mathrm{i}v](x,y)$$

is the derivative of f. As a consequence of the CRE, we have $\partial_x^2 u + \partial_y^2 u = \partial_x^2 v + \partial_y^2 v = 0$.

Complex Brownian motion is the stochastic process $B_t = B_t^{(1)} + \mathrm{i}B_t^{(2)}$, where $B^{(1)}$ and $B^{(2)}$ are independent one-dimensional Brownian motions. Let f be complex differentiable. Writing $f(z) = u(x,y) + \mathrm{i}v(x,y)$, we apply the two-dimensional Itô formula separately to the real and imaginary part of f and obtain for $U_t + \mathrm{i}V_t = Y_t = f(B_t) = u(B_t^{(1)}, B_t^{(2)}) + \mathrm{i}v(B_t^{(1)}, B_t^{(2)})$ the equality

$$\mathrm{d}Y_t = \mathrm{d}U_t + \mathrm{i}\mathrm{d}V_t = \big(\nabla[u + \mathrm{i}v](B_t^{(1)}, B_t^{(2)})\big) \cdot (B_t^{(1)}, B_t^{(2)})^{\mathrm{t}} + \frac{1}{2}\big(\Delta[u + \mathrm{i}v](B_t^{(1)}, B_t^{(2)})\big)\,\mathrm{d}t.$$

By the CRE, the second term above vanishes, and the first is equal to $f'(B_t)\mathrm{d}B_t$. So we obtain the Itô formula without Itô term,

$$\mathrm{d}Y_t = f'(B_t)\mathrm{d}B_t.$$

In other words for *complex* Brownian motion, stochastic integrals behave just like classical ones! This is an important ingredient in the next theorem.

(3.47) Theorem

Let $D \subset \mathbb{C}$ be a domain (open, connected subset) in $\mathbb{C}$, $f : D \to \mathbb{C}$ analytic. Let B be a complex Brownian motion, starting in $x \in D$. Define

$$\tau_D := \inf\{t \geqslant 0 : B_t \notin D\}, \qquad \text{and} \qquad \xi_t(\omega) := \int_0^t \big|f'(B_s(\omega))\big|^2\,\mathrm{d}s.$$

There exists a complex Brownian motion $\tilde{B}$, starting in $f(x) \in \mathbb{C}$, such that

$$f(B_t)\mathbb{1}_{\{t<\tau_D\}} = \tilde{B}_{\xi_t}\mathbb{1}_{\{t<\tau_D\}}.$$

If in addition f is bijective from D onto $\tilde{D} = f(D)$, then we have

$$f(B_{t\wedge\tau_D}) = \tilde{B}_{\xi_t\wedge\tilde{\tau}_{\tilde{D}}} \qquad \text{with} \qquad \tilde{\tau}_{\tilde{D}} := \inf\{t \geqslant 0 : \tilde{B}_t \notin \tilde{D}\}.$$

Proof: Let $B_t = B_t^{(1)} + \mathrm{i}\,\mathrm{d}B_t^{(2)}$, and $\hat{B}_t = (B_t^{(1)}, B_t^{(2)})$. We have seen above that with $Y_t = U_t + \mathrm{i}V_t = f(B_t)$, we have

$$\mathrm{d}U_t + \mathrm{i}\mathrm{d}V_t = \mathrm{d}Y_t = f'(B_t)\mathrm{d}B_t = \partial_x u(\hat{B}_t)\mathrm{d}B_t^{(1)} + \partial_y u(\hat{B}_t)\mathrm{d}B_t^{(2)} + \mathrm{i}\big(\partial_x v(\hat{B}_t)\mathrm{d}B_t^{(1)} + \partial_y v(\hat{B}_t)\mathrm{d}B_t^{(2)}\big).$$

Therefore, U and V are martingales. We will now show that they have the same quadratic variation process. U is the Itô process with $\mathrm{d}U_t = \sigma(\hat{B}_t)\mathrm{d}\hat{B}_t$, i.e. with with drift 0 and 2×1 diffusion matrix $\sigma(\hat{B}_t) = (\partial_x u(\hat{B}_t), \partial_y u(\hat{B}_t))^{\mathrm{t}}$. We apply the general Itô formula (3.40) (for the simple case $m = 1$) with the function $f(x) = x^2$ and obtain

$$\mathrm{d}(U^2)_t = 2U_t\mathrm{d}U_t + \big((\partial_x u(\hat{B}_t))^2 + (\partial_y u(\hat{B}_t))^2\big)\,\mathrm{d}t.$$

Since $\int_0^t U_s \mathrm{d}U_s$ is a martingale, this shows that the qvp of the process U is given by the process

$$\omega \mapsto \Big(\int_0^t |f'(B_s(\omega))|^2\,\mathrm{d}s\Big)_{t\geqslant 0} = \Big(\int_0^t \big([(\partial_x u)^2](\hat{B}_s(\omega)) + [(\partial_y u)^2](\hat{B}_s(\omega))\big)\,\mathrm{d}s\Big)_{t\geqslant 0}.$$

The same calculation for the process V shows

$$\mathrm{d}(V^2)_t = 2V_t\mathrm{d}V_t + \big((\partial_x v(\hat{B}_t))^2 + (\partial_y v(\hat{B}_t))^2\big)\,\mathrm{d}t.$$

$\int_0^t V_s \mathrm{d}V_s$ is a martingale, and by the CRE, $(\partial_x u)^2 + (\partial_y u)^2 = (\partial_x v)^2 + (\partial_y v)^2$. So the qvp of V is exactly the same process (not just equal in distribution). This means that we can make U and V into Brownian motions *by the same time change* in the spirit of Theorem (3.45)!

To do this, let us first clarify that we only need to consider the case where D is compact and where f is analytic on an open subset containing D. Namely, otherwise we define

$$D_n := B(0,n) \cap \big\{z \in D : \inf\{|z - \tilde{z}| : \tilde{z} \in \partial D\} > 1/n\big\}.$$

Then each $\overline{D}_n$ has the required properties, and once we have proved the claim for $\overline{D}_n$ we let $n \to \infty$. Path continuity of Brownian motion then gives the general claim.

Assume therefore that D is compact and where f is analytic on an open subset containing D. Then f' is uniformly bounded on D, and thus $(U_{t\wedge\tau_D})_{t\geqslant 0}$ and $(V_{t\wedge\tau_D})_{t\geqslant 0}$ are L^2-martingales by the Itô isometry. They share the qvp $\tilde{A}_t = \int_0^{t\wedge\tau_D} |f'(B_s)|^2\,\mathrm{d}s$. We can not directly apply (3.45) because $\lim_{t\to\infty} \tilde{A}_t < \infty$ almost surely. We repair this as follows: Let B'_t be another complex Brownian motion, independent of B, and let $(\tilde{\mathcal{F}}_{t+})$ be the completed σ-algebra generated by both Brownian motions. Then B and B' are $(\tilde{\mathcal{F}}_{t+})$-martingales. We define

$$M_t(\omega) := f(B_{t\wedge\tau_D(\omega)}(\omega)) + B'_t(\omega) - B'_{t\wedge\tau_D(\omega)}(\omega);$$

in words, we follow $(f(B_t))$ until the time τ_D when B exits D, and after that time we follow the Brownian motion B' which is however first „glued“ to the value of $f(B_{\tau_D})$. The glueing is achieved by adjusting the current position of B' by the term $f(B_{\tau_D}) - B'_{\tau_D}$.

The real and imaginary parts of all terms are martingales by the optional stopping theorem, and they still have the same quadratic variation process $\bar{A}$ with $\bar{A}_t(\omega) = \tilde{A}_{t\wedge\tau_D(\omega)}(\omega) + t - \tilde{A}_{\tau_D(\omega)}(\omega)$. Now we can apply (3.45) and find Brownian motions $\tilde{B}^{(1)}$ and $\tilde{B}^{(2)}$ such that with $M_t(\omega) = (\tilde{B}^{(1)}_{\bar{A}_t(\omega)}(\omega), \tilde{B}^{(2)}_{\bar{A}_t(\omega)}(\omega))$. Since $\bar{A}_t \mathbb{1}_{\{t<\tau_D\}} = A_t = \xi_t$ and since $M = Y$ on $\{t < \tau_D\}$, we have

$$f(B_t)\mathbb{1}_{\{t<\tau_D\}} = \big(\tilde{B}^{(1)}_{\xi} + \mathrm{i}\tilde{B}^{(2)}_{\xi}\big)\mathbb{1}_{\{t<\tau_D\}},$$

which is almost the first claimed equality. What is still missing is that $\tilde{B}^{(1)}$ and $\tilde{B}^{(2)}$ are *independent* Brownian motions. In order to see this, we apply the general Itô formula for the function $f(x,y) = xy$ and the $\mathbb{R}^2$-valued Itô-process

$$\mathrm{d}\begin{pmatrix} U_t \\ V_t \end{pmatrix} = \begin{pmatrix} \partial_x u(\hat{B}_t) & \partial_y u(\hat{B}_t) \\ \partial_x v(\hat{B}_t) & \partial_y v(\hat{B}_t) \end{pmatrix} \mathrm{d}\hat{B}_t =: \sigma(\hat{B}_t)\mathrm{d}\hat{B}_t.$$

We have

$$Df = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \Longrightarrow \qquad \mathrm{tr}\Big(\sigma^*(\hat{B}_t) Df(\hat{B}_t)\sigma(\hat{B}_t)\Big) = \mathrm{tr}\begin{pmatrix} 2u_x v_x & * \\ * & 2u_y v_y \end{pmatrix} = 0,$$

since $u_x v_x = -u_y v_y$ by the CRE (we wrote $u_x = \partial_x u$ here to cut down on notation). This means that $\mathrm{d}(UV)_t = U_t \mathrm{d}V_t + V_t \mathrm{d}U_t$, in other words the process UV is a martingale. This means that $\mathbb{E}(U_{s,t}V_{s,t}) = 0$, and then by optional stopping also $\mathbb{E}(\tilde{B}^{(1)}_{s,t}\tilde{B}^{(2)}_{s,t}) = 0$. So the increments of the two Brownian motions are uncorrelated, and since they are Gaussian, they are independent. The independence of the full processes follows from the general fact that for two Brownian motions that are adapted to the same σ-algebra, the independence of the increments $(\tilde{B}^{(1)}_{s,t}, \tilde{B}^{(2)}_{s,t})$ for all $s < t$ is sufficient for the independence of the full processes. The proof is a standard exercise: consider the characteristic function of the type $\mathbb{E}(\exp(\sum_{i=1}^n a_i \tilde{B}^{(1)}_{s_i,t_i} \sum_{j=1}^m a_i \tilde{B}^{(2)}_{s_j,t_j}))$ and use successive conditioning and the assumed inedependence to show that this is equal to the product of the corresponding characteristic functions for $B^{(1)}$ and $B^{(2)}$.

Let us now show the second statement. If f is bijective from D to $\tilde{D}$, then $z \notin D$ if and only if $f(z) \notin \tilde{D}$ for all $z \in \mathbb{C}$. Consequently,

$$t < \tau_D \quad \Leftrightarrow \quad B_s \in D \ \forall s \leqslant t \quad \Leftrightarrow \quad \tilde{B}_{\xi_s} = f(B_{s \wedge \tau_D}) \in \tilde{D} \ \forall s \leqslant t \quad \Leftrightarrow \quad \xi_s < \tilde{\tau}_{\tilde{D}}.$$

So the claim follows. □

(3.48) Remark

Theorem (3.47) implies that two-dimensional (or complex) Brownian paths are **conformally covariant**, in the following sense: For a Brownian motion B starting in $x \in D$, the map $\omega \mapsto \{B_s(\omega) : s \leqslant \tau_D(\omega)\}$ defines a random subset of D, namely the „trace“ of the path of B. Conformal covariance means that when f is bijective from D onto $V = f(D)$, then the distribution of the random set defined by $\omega \mapsto \{f(B_s(\omega)) : s \leqslant \tau_D(\omega)\}$ is the same as the distribution of the map $\omega \mapsto \{\tilde{B}_s(\omega) : s \leqslant \tau_{\tilde{D}}(\omega)\}$, where $\tilde{B}$ is a Brownian motion started in $f(x) \in V$. This is a similar statement as the ones we gave at the beginning of Chapter 2, but more complicated because it involves a time change. A special case occurs when $D = \tilde{D}$, then the distribution of the random set is invariant under the map $f : D \mapsto \tilde{D}$. The conformal covariance of Brownian motion is a corner stone of the theory of Schramm-Löwner-evoultions, which produced the first two fields medals for probability theory in 2006 and 2010.

(3.49) Definition

Let B be a d-dimensional $(\mathcal{F}_t)$-Brownian motion, $b : [0,\infty) \times \mathbb{R}^n \to \mathbb{R}^n$ and $\sigma : [0,\infty) \times \mathbb{R}^n \to$

$\mathbb{R}^{n\times d}$ be measurable. We say that a stochastic process X is a **strong solution** of the SDE

$$\mathrm{d}X_t = b(t, X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}B_t, \qquad X_0 = x \in \mathbb{R}^n, t < T < \infty$$

if X is $(\mathcal{F}_t)$-progressively measurable, and

$$X_t = x + \int_0^t b(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s, X_s)\,\mathrm{d}B_s \qquad \forall t \leqslant T,$$

$\mathbb{P}$-almost surely.

(3.50) Remark

a) If b and σ are Lipschitz-continuous, then a solution to the SDE exists and is unique (up to sets of measure zero). We will prove this fact in the context of rough paths in the next chapter. There, we will need a bit mor regularity of b, σ (they need to be in C^2), but the benefit is that we get a much better concept of a solution (see c) below). For the classical proof of existence of solutions to SDE see e.g. Schilling/Partzsch.

b) It is possible to show that for two solutions X^x and X^y with different (possibly random) starting points $X_0^x = x$, $X_0^y = y$ of the SDE, there exists $C, k > 0$ such that

$$\mathbb{E}(\|X^x - X^y\|_{\infty,T}^2) \leqslant C\,\mathrm{e}^{kT}\,\mathbb{E}(|x-y|^2)$$

holds. Here, $\|X\|_{\infty,T} = \sup\{|X_t| : t \leqslant T\}$. This means that the solution of the SDE is continuous in the initial condition (in the stated norm): small changes of initial condition mean small changes of the solution.

c) Strong solutions of SDE are, however, *not* continuous in the „noise that drives them“. This means that if for two Brownian motion paths $B(\omega)$ and $B(\omega')$ we have $\|B.(\omega) - B.(\omega')\|_{\infty,T} < \varepsilon$, we have no guarantee that the solutions of the SDE for the same pair ω, ω' are close to each other. This is not convenient, since in many cases the noise is supposed to be a „perturbation“ of the ODE without noise, and it means e.g. that the solution with no noise at all may look very different from the solution with a very tiny bit of noise. But it is unavoidable: one can show that there exists no norm on the space of continuous functions such that the map that maps a Brownian motion path to the solution of the SDE for this path is continuous in this norm.

In the chapter about rough paths, on the other hand, we will meet another type of solution (not the strong solution above!) that has this continuity property. This is one of the big advantages of using rough paths.

Our last item in this chapter is about the Markov property for solutions of SDE. In Definition (2.41) we only introduced *time-homogenous* Markov processes in order not to introduce too many difficulties at once. Since SDE are quite often time-inhomogenous, we now extend this definition. It is a useful exercise to go through all our earlier statements on Markov processes and check that they still hold (with the hopefully obvious modifications) for time-inhomogenous Markov processes. The intuition of the Definition below is that $\mathbb{P}^{x,t}$ is the probability measure for the process X started in x at time t, but that from the point of view of the process itself, time starts at 0. In other words, $\mathbb{P}^{x,t}(f(X_s))$ means that the process (in a global point of view) has run for time s after being started at time t in x and is then plugged into f and integrated. Of course, the definition reduces to Definition (2.41) if the $\mathbb{P}^{x,t}$ are the same for all t.

(3.51) Definition

Let

- $(\Omega, \mathcal{F})$ be a measurable space with a filtration $(\mathcal{F}_t)$,
- E be a metric space and $\mathcal{E}$ its Borel-σ-algebra,
- $(X_t)_{t \geqslant 0}$ be a family of functions from Ω to E, such that each X_t is $\mathcal{F}_t$-$\mathcal{E}$-measurable,
- $(\mathbb{P}^{x,t})_{x \in E, t \geqslant 0}$ be a family of probability measures on $(\Omega, \mathcal{F})$ such that $\mathbb{P}^{x,t}(X_0 = x) = 1$ for all $x \in \mathbb{E}$.

The pair $((X_s)_{s \geqslant 0}, (\mathbb{P}^{x,t})_{x \in E, t \geqslant 0})$ is called

a) **a weak Markov process** if for all $f \in C_b(E, \mathbb{R})$, all $x \in E$ and all $s, r \geqslant 0$, we have

$$\mathbb{E}^{x,s}(f(X_{t+r}) \,|\, \mathcal{F}_r)(\bar\omega) = \mathbb{E}^{X_r(\bar\omega), s+r}(f(X_t)) \qquad \text{for } \mathbb{P}^{x,s}\text{-almost all } \bar\omega \in \Omega.$$

b) **a Markov process** if for all measurable, bounded $F : (E^{\mathbb{R}_0^+}, \mathcal{E}^{\otimes \mathbb{R}_0^+}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, all $x \in E$ and all $s, r \geqslant 0$, we have

$$\mathbb{E}^{x,s}\Big(F \circ \theta_r\big((X_t)_{t \geqslant 0}\big) \,\Big|\, \mathcal{F}_r\Big)(\bar\omega) \equiv \mathbb{E}^{x,s}\Big(F\big((X_{t+r})_{t \geqslant 0}\big) \,\Big|\, \mathcal{F}_r\Big)(\bar\omega) = \mathbb{E}^{X_r(\bar\omega), s+r}(F((X_t)_{t \geqslant 0})),$$

for $\mathbb{P}^{x,s}$-almost all $\bar\omega \in \Omega$.

c) **a strong Markov process** if for all $\mathcal{B}(\mathbb{R}) \otimes \mathcal{E}^{\otimes \mathbb{R}_0^+} - \mathcal{B}(\mathbb{R})$-measurable, bounded functions $F : \mathbb{R}_0^+ \times \mathbb{E}^{\mathbb{R}_0^+} \to \mathbb{R}$, all $x \in E$ and all $(\mathcal{F}_t)$-stopping times τ, and all $s \geqslant 0$, we have

$$\mathbb{E}^{x+s}\Big(F\big(\tau, (\theta_\tau X)_t\big) \,\Big|\, \mathcal{F}_\tau\Big)(\bar\omega) = \mathbb{E}^{X_{\tau(\bar\omega)}(\bar\omega), s+\tau(\bar\omega)}\Big(F\big(\tau(\bar\omega), (X_t)_{t \geqslant 0}\big)\Big),$$

for $\mathbb{P}^x$-almost all $\bar\omega \in \{\tau < \infty\}$.

(3.52) Theorem

Let b, σ be as in (3.49), and assume they are Lipschitz continuous. Then the solution of the SDE

$$\mathrm{d}Z_t = b(t, Z_t)\,\mathrm{d}t + \sigma(t, Z_t)\,\mathrm{d}B_t \qquad (*)$$

is a weak Markov process.

More precisely: For $x \in \mathbb{R}^n$ and $s \geqslant 0$, let $(Z^x_{t;s})_{t \geqslant 0}$ be the strong solution of the SDE

$$\mathrm{d}Z^x_{t;s} = b(t+s, Z^x_{t;s})\,\mathrm{d}t + \sigma(t+s, Z^x_{t;s})\,\mathrm{d}B_t, \qquad Z^x_{0;s} = x.$$

Let $\mathbb{P}^{x,s}$ be the distribution of $(Z^x_{t;s})_{t \geqslant 0}$, i.e. the probability measure on $C(\mathbb{R}, \mathbb{R}^n)$ such that for all $t_1, \ldots, t_m \geqslant 0$, and all $A_1, \ldots, A_m \in \mathcal{B}(\mathbb{R}^n)$,

$$\mathbb{P}^{x,s}(\pi_{t_i} \in A_i \ \forall i \leqslant n) = \mathbb{P}(Z^x_{t_i;s} \in A_i \ \forall i \leqslant n),$$

with $\pi_t(\omega) = \omega(t)$ the point evaluation. Then the family $(\mathbb{P}^{x,s})_{x \in \mathbb{R}^n, s \geqslant 0}$ together with the maps $(\pi_t)_{t \geqslant 0}$, the complete filtration $(\mathcal{F}_t)_{t \geqslant 0}$ with $\mathcal{F}_t = \bigcap_{u > t} \sigma(\pi_r : r \leqslant u)$ and the target space $E = \mathbb{R}^n$ is a weak Markov process in the sense of Definition (3.51).

Proof: We have to show that for all s, t and all bounded continuous f,

$$\mathbb{E}^{x,s}(f(\pi_{t+r}) \,|\, \mathcal{F}_r)(\bar\omega) = \mathbb{E}^{\pi_r(\bar\omega), s+r}(f(\pi_t)) \qquad \mathbb{P}^{x,s}\text{-almost surely.} \qquad (*)$$

The central formula is the identity

$$Z^x_{t+r;s}(\bar\omega) = Z^x_{r;s}(\bar\omega) + \int_r^{t+r} b(u+s, Z_{u;s}(\bar\omega))\,\mathrm{d}u + \Big[\int_r^{t+r} \sigma(u+s, Z_{u;s})\,\mathrm{d}B_u\Big](\bar\omega),$$

which follows from the additivity of integrals. Let $\tilde Z^y_{t+r;s}$ be the solution of the SDE

$$\tilde Z^y_{t;s+r} = y + \int_0^t b(u+s+r, \tilde Z^y_{u;s+r})\,\mathrm{d}u + \int_0^t \sigma(u+s+r, \tilde Z^y_{u;s+r})\,\mathrm{d}\tilde B_u,$$

where the Brownian motion $\tilde B_u = B_{r+u}$ has increments that are independent of $(B_v)_{v \leqslant r}$. Then by the above identity and the uniqeness of SDE solutions, the distribution of $Z^x_{t+r;s}$ is the same as the distribution of $\omega \mapsto \tilde Z^{Z^x_{r;s}(\omega)}_{t;s+r}(\omega)$. Writing $\phi(y,r,t,(B_{r,u}(\omega))_{u \geqslant r}) := \tilde Z^y_{t;s+r}(\omega)$ for the solution map, we then get

$$\begin{aligned}\mathbb{E}^{x,s}(f(\pi_{t+r}) \,|\, \mathcal{F}_r)(\bar\omega) &= \mathbb{E}(f(Z^x_{t+r;s}) \,|\, \mathcal{F}_r)(\bar\omega) = \mathbb{E}(f \circ \phi(Z^x_{r;s}, r, t, (B_{r,u})_{u \geqslant r}) \,|\, \mathcal{F}_r)(\bar\omega) \\ &= \mathbb{E}(f \circ \phi(Z^x_{r;s}(\bar\omega), r, t, (B_{r,u})_{u \geqslant r})) = \mathbb{E}(f(\tilde Z^{Z^x_{r;s}(\bar\omega)}_{t;s+r})) = \mathbb{E}^{Z^x_{r;s}(\bar\omega), s+r}(f(\pi_t)). \qquad (**)\end{aligned}$$

The equality of first and second line holds because the map $\bar\omega \mapsto Z^x_s(\bar\omega)$ is $\mathcal{F}_r$-measurable and $\omega \mapsto (B_{r,u}(\omega))_{u \geqslant r}$ is independent of $\mathcal{F}_r$; see Proposition (3.23) of the Probability Theory lecture notes. Since π_r has distribution $Z^x_{r;s}$ under $\mathbb{P}^{x,s}$, we have for all $(\mathcal{F}_r)$-measurable A that

$$\mathbb{E}^{x,s}(\mathbb{E}^{x,s}(f(\pi_{t+r}) \,|\, \mathcal{F}_r)\mathbb{1}_A) \overset{(**)}{=} \mathbb{E}^{x,s}(\mathbb{E}^{\pi_r, s+r}(f(\pi_t))\mathbb{1}_A).$$

Since $\omega \mapsto \mathbb{E}^{\pi_r(\omega), s+r}(f(\pi_t))$ is clearly $\mathcal{F}_r$-measurable, this implies $(*)$. □

4. Rough Paths

(4.1) Another look at Young integrals

For $X \in C^\alpha$, $Y \in C^\beta$ with $\gamma := \alpha + \beta > 1$, Theorem (3.4) states that the limit

$$\int_0^T Y_s\,\mathrm{d}X_s := \lim_{|P| \to 0} \sum_{[u,v) \in P} \underbrace{Y_u X_{u,v}}_{=:\Xi_{u,v}}$$

exists. Let us recall the central estimate that makes this true: the quantity

$$\|\delta\Xi\|_\gamma := \sup_{r: u<r<v} \frac{1}{|v-u|^\gamma} |\Xi_{u,v} - \Xi_{u,r} - \Xi_{r,v}|$$

is finite. Here is another point of view on the situation: if $\Xi_{u,v}$ would be a proper integral, then it would be additive, i.e. $\Xi_{u,v} - \Xi_{u,r} - \Xi_{r,v}$ would be equal to zero. But it is not an integral, only an approximation. The condition on $\delta\Xi$ just means that the „deviation from additivity“ $(\delta\Xi)_{u,v} = \sup_{r: u<r<v} \Xi_{u,v} - \Xi_{u,r} - \Xi_{r,v}$ vanishes faster than linearly as $|u - v| \to 0$.

(4.2) Beyond Young integrals

For $X \in C^\alpha$ and $Y \in C^\beta$ with $\alpha + \beta < 1$, the supremum over $r \in (u,v)$ of the quantity

$$(\delta\Xi)_{u,r,v} = \Xi_{u,v} - \Xi_{u,r} - \Xi_{r,v} = Y_u X_{u,v} - Y_u X_{u,r} - Y_r X_{r,v} = -Y_{u,r} X_{r,v}$$

is only guaranteed to vanish like $|u-v|^{\alpha+\beta}$, which is not fast enough. To see this, take $r = u+(v-u)/2$. This breaks the proof of Theorem (3.4). To see how this can be repaired, let us first consider the special case when $Y_u = f(X_u)$ for a C^2-function f. Then $Y \in C^\alpha$, and so we are looking at the cases $\alpha \leqslant 1/2$. The approximation $\Xi_{u,v} = f(X_u)X_{u,v}$ to the (not yet well-defined!) integral $\int_u^v f(X_r)\,\mathrm{d}X_r$ is then not good enough, since its „deviation from additivity" does not shrink fast enough as $|u-v| \to 0$. The key idea is to improve this by adding a further term: by expanding f around the point X_u, we find

$$\int_u^v f(X_r)\,\mathrm{d}X_r \approx \int_u^v \big(f(X_u) + f'(X_u)X_{u,r}\big)\,\mathrm{d}X_r = f(X_u)X_{u,v} + f'(X_u)\underbrace{\int_u^v X_{u,r}\,\mathrm{d}X_r}_{=:\mathbb{X}_{u,v}}.$$

The improvement of approximation is comparable to a „one sided trapezoid rule" in numerical analysis. The first term on the right hand side is the Riemann sum term, the next one is the next order improvement. Of course, still all the integrals do not make any sense - but we remark already now that if we somehow succeed in making sense of the single quantity $\mathbb{X}_{u,v}$ then we have a (hopefully) improved approximation of $\int_u^v f(X_r)\,\mathrm{d}X_r$ for all $f \in C^2$. Before we worry about what $\mathbb{X}_{u,v}$ might be, let us convince ourselves that finding it is worth the effort.

For this purpose, let us go back to the case $\alpha > 1/2$ for the moment. Then all integrals above make sense as Young integrals, and we can easily calculate (exercise!)

$$\int_u^v X_r\,\mathrm{d}X_r = \frac{1}{2}(X_v^2 - X_u^2).$$

This means that

$$\mathbb{X}_{u,v} = \int_u^v (X_r - X_u)\,\mathrm{d}X_r = \int_u^v X_r\,\mathrm{d}X_r - X_uX_{u,v} = \frac{1}{2}(X^2)_{u,v} - X_uX_{u,v}.$$

This quantity is again not additive, but we can specify precisely the defect to additivity:

$$\begin{aligned}\mathbb{X}_{u,v} + \mathbb{X}_{v,w} &= \frac{1}{2}(X^2)_{u,w} - X_uX_{u,v} - X_vX_{v,w} = \\ &= \frac{1}{2}(X^2)_{u,w} - X_uX_{u,w} + X_uX_{v,w} - X_vX_{v,w} = \mathbb{X}_{u,w} - X_{u,v}X_{v,w}.\end{aligned}$$

This means that

$$\mathbb{X}_{u,w} = \mathbb{X}_{u,v} + \mathbb{X}_{v,w} + X_{u,v}X_{v,w} \qquad \text{(A)}$$

and important identity that we will meet again later. On the other hand, we have

$$\mathbb{X}_{u,v} = \frac{1}{2}(X_v + X_u)X_{u,v} - X_uX_{u,v} = \frac{1}{2}(X_{u,v})^2, \qquad \text{(B)}$$

which means that $\sup_{u,v\in[0,T]} \frac{1}{|v-u|^{2\alpha}}\mathbb{X}_{u,v} < \infty$, i.e. $\mathbb{X}$ is twice as regular as X. Of course, what we really want is that if we re-define Ξ as

$$\Xi_{u,v} = f(X_u)X_{u,v} + f'(X_u)\mathbb{X}_{u,v},$$

then the deviation from additivity of the improved approximation $\Xi_{u,v}$ vanishes faster than for the original approximation $f(X_u)X_{u,v}$. This is indeed the case:

$$\begin{aligned}\Xi_{u,v} + \Xi_{v,w} &= f(X_u)X_{u,w} + (f(X_v) - f(X_u))X_{v,w} + f'(X_u)\mathbb{X}_{u,v} + f'(X_v)\mathbb{X}_{v,w} = \\ &= f(X_u)X_{u,w} + (f(X))_{u,v}X_{v,w} + f'(X_u)\underbrace{(\mathbb{X}_{u,v} + \mathbb{X}_{v,w})}_{\overset{(A)}{=}\mathbb{X}_{u,w} - X_{u,v}X_{v,w}} + (f'(X))_{u,v}\mathbb{X}_{v,w} = \\ &= \underbrace{f(X_u)X_{u,v} + f'(X_u)\mathbb{X}_{u,w}}_{=\Xi u,w} + (f(X_v) - f(X_u) - f'(X_u)X_{u,v})X_{v,w} + (f'(X))_{u,v}\mathbb{X}_{v,w}.\end{aligned}$$

By Taylors theorem and the mean value theorem, there exist $\tilde{r}, \tilde{s} \in [u, v]$ with

$$R^{f(X)}_{u,v} := f(X_v) - f(X_u) - f'(X_u)X_{u,v} = \frac{1}{2}f''(\tilde{r})(X_{u,v})^2, \qquad (f'(X))_{u,v} = f''(\tilde{s})X_{u,v},$$

which means that

$$\sup_{v:u<v<w} |\Xi_{u,w} - \Xi_{u,v} - \Xi_{v,w}| \leqslant \|f''\|_\infty \sup_{v:u<v<w} \Big|\frac{1}{2}(X_{u,v})^2 X_{v,w} + X_{u,v}\mathbb{X}_{v,w}\Big| \sim |u - w|^{3\alpha},$$

where for the second term above we used equality (B). Therefore, the new $\Xi_{u,v}$ deviates from additivity only by oder 3α, compared to the previous order 2α. Using it to approximate the integral as in (4.1) then repairs the proof of Theorem (3.4) and restores convergence if $\alpha > 1/3$. This is good enough for Brownian motion! The task that remains is to make sense of the expression $\mathbb{X}_{u,v} = \int_u^v X_{u,r}\,\mathrm{d}X_r$.

(4.3) Example

Let B be a one-dimensional Brownian motion. Then we can define

$$\mathbb{X}_{u,v}(\omega) := \Big[\int_u^v B_{u,s}\,\mathrm{d}B_s\Big](\omega),$$

where on the right hand side we use the Itô integral. Itô's formula then gives

$$\mathbb{X}_{u,v} = \int_{u,v} B_s\,\mathrm{d}B_s - B_u B_{u,v} = \frac{1}{2}(B_v^2 - B_u^2) - \frac{1}{2}(v-u) - B_u B_{u,v} = \frac{1}{2}(B_{u,v})^2 - \frac{1}{2}(v-u),$$

which means that $\mathbb{X}_{u,v}(\omega) \in C^{2\alpha}$ almost surely for any $\alpha < 1/2$. This regularity is what is needed from equation (B) of (4.2). On the other hand,

$$\begin{aligned}\mathbb{X}_{u,w} - \mathbb{X}_{u,v} - \mathbb{X}_{v,w} &= \frac{1}{2}\big((B_{u,w})^2 - (B_{u,v})^2 - (B_{v,w})^2\big) = -B_u B_w + B_u B_v + B_v B_w - B_v^2 = \\ &= B_{u,v}B_{v,w},\end{aligned}$$

so equality (A) from (4.2) holds. In other words, the Itô-integral seems like a valid choice for the quantity $\mathbb{X}$. There are some hidden pitfalls in this reasoning, though (can you spot them), which we will discuss and remove below. The Stratonovich integral, on the other hand, is also a valid choice, as is any other of the classical stochastic integrals (why?). For general square integrable martingales X with qvp A, the correct choice is

$$\mathbb{X}_{u,v}(\omega) = \Big[\int_u^v X_{u,s}\,\mathrm{d}X_s\Big](\omega) \overset{\text{Itô}}{=} \frac{1}{2}(X_{u,v}(\omega))^2 - \frac{1}{2}A_{u,v}(\omega).$$

Since $A_{u,v}$ is additive, the same calculation as above gives that (4.2) (A) holds. For (B), one needs to use Theorem (3.45) and change in order to obtain a Brownain motion again - we omit the details.

Instead of trying to find $\mathbb{X}_{u,v}$, we will take an axiomatic approach in the following: the philosophy is to just postulate the existence of some $\mathbb{X}$ with the property (B) and the 2α-regularity that comes from (A) in the context of (4.2). We will then call the pair $\big((X_u)_{0\leqslant u\leqslant T}, (\mathbb{X}_{u,v})_{0\leqslant u,v\leqslant T}\big)$ a rough path, and show that then all integrals, differential equations etc. make sense. We will want to do this in higher dimensions and therfore we will need some preparations.

(4.4) Tensor products

In the following, U, V, W will always denote Hilbert spaces - in fact we will need only finite dimensional Hilbert spaces (aka $\mathbb{R}^n$), but the theory actually works even for Banach spaces (but is less clean there). Let (e_i) be an orthonormal basis (ONB) of U and (f_i) be an ONB of V. The tensor product $U\otimes V$ of U and V is the Hilbert space with ONB $(e_i\otimes f_j)_{i,j}$ (which, as a set, can be identified with the set of ordered pairs $\{(e_i, f_j) : i, j\}$) and scalar product

$$\langle e_i\otimes f_j, e_k\otimes f'_\ell\rangle_{U\otimes V} = \langle e_i, e_k\rangle_U\langle f_j, f_\ell\rangle_V.$$

For two vectors $u=\sum_i \alpha_i e_i\in U$, $v=\sum_j \beta_j f_j\in V$, the vector $\sum_{i,j}\alpha_i\beta_j e_i\otimes f_j\in U\otimes V$ is the tensor product $u\otimes v$ of u and v.

Two special cases are instructive: if U, V are vector spaces of functions $f : x\mapsto f(x)$, $g : y\mapsto g(y)$, then $f\otimes g$ is (can be identified with) the function $f\otimes g : (x,y)\mapsto f(x)g(y)$. In the even more special case where $x\in\{1,\dots,n\}$ and $y\in\{1,\dots,m\}$, $U\simeq\mathbb{R}^n$ and $V\simeq\mathbb{R}^m$. Then we write $u(i)=u_i$, $v(j)=v_j$, and identify $u\otimes v$ with an $n\times m$-matrix:

$$u\otimes v = (u_iv_j)_{1\leqslant i\leqslant n, 1\leqslant j\leqslant m}\in\mathbb{R}^{n\times m}.$$

This formula gets a separate line because this identification will be used in all that follows. It also shows that for finite dimensional $U\simeq\mathbb{R}^n, V\simeq\mathbb{R}^m$, we have $U\otimes V\simeq\mathbb{R}^{n\times m}$.

We will use another identification of two vector spaces frequently: the map $U\times V\to U\otimes V$, $(u,v)\mapsto u\otimes v$ is bilinear, and therefore

$$\{\text{bilinear maps } U\times V\to W\}\equiv\mathcal{L}(U\times V, W)\simeq\mathcal{L}(U\otimes V, W)\equiv\{\text{linear maps } U\otimes V\to W\}$$

(4.5) Definition

Recall Definition (1.34): for $f:\mathbb{R}_0^+\to\mathbb{R}^d$, $D\subset\mathbb{R}_0^+$ and $\alpha>0$ we defined $\|f\|_{D,\alpha}=\sup_{s,t\in D,s\neq t}|t-s|^{-\alpha}|f_{s,t}|$ and $C^\alpha(D,\mathbb{R}^d)=\{f:\|f\|_{D,\alpha}<\infty\}$. We will now abuse this notation and define

$$\|f\|_{T,\alpha} := \|f\|_{[0,T],\alpha}\qquad\text{for } T>0.$$

Also, if $f : U\to V$ where U is a subset of $U\subset\mathbb{R}^n$, V a Hilbert space, and $D\subset U$, then the definition $\|f\|_{D,\alpha}$ still makes sense by replacing $|t-s|$ by $\|u-u'\|$ for $u,u'\in D$, and $|f_{s,t}|$ by $\|f_{s,t}\|_V$. We will continue writing $|f|$ instead of $\|f\|_V$ in this case.

For $F:[0,T]^2\to V$, we define

$$\|F\|_{T,\alpha,\text{diag}} := \sup\{|t-s|^{-\alpha}F(s,t) : s<t, s,t\in[0,T]\},$$

and $C_2^\alpha([0,T],V) = \{F : [0,T]^2 \to V : \|F\|_{\alpha,T,\text{diag}} < \infty\}$. Note that $C_2^\alpha([0,T],V) \neq C^\alpha([0,T]^2,V)$, actually the two are very different. Functions from $C_2^\alpha([0,T],V)$ are necessarily zero on the diagonal $s = t$, which is not the case for functions from $C^\alpha([0,T]^2,V)$; on the other hand, they have no restrictions for any points away from the diagonal.

(4.6) Definition

Let $X : [0,T] \to V$, $\mathbb{X} : [0,T]^2 \to V \otimes V$, and consider the pair $\boldsymbol{X} = (X, \mathbb{X})$.
a) **$\boldsymbol{X}$ satisfies Chen's relation** if for all $0 \leqslant u \leqslant v \leqslant w \leqslant T$, we have

$$\mathbb{X}_{u,w} = \mathbb{X}_{u,v} + \mathbb{X}_{v,w} + X_{u,v} \otimes X_{v,w}.$$

b) Let $\alpha \in (1/3, 1/2]$. $\boldsymbol{X} = (X, \mathbb{X})$ is called a **α-Hölder rough path** (or α-rough path) if
(i): $X \in C^\alpha([0,T],V)$,
(ii): $\mathbb{X} \in C_2^{2\alpha}([0,T], V \otimes V)$,
(iii): Chen's relation holds.
We write $\mathcal{C}^\alpha([0,T],V) \subset C^\alpha([0,T],V) \times C_2^{2\alpha}([0,T], V \otimes V)$ for the set of α-rough paths.

(4.7) Remark

a) Chen's relation is the axiomatization of equality (A) from (4.2), and $\mathbb{X}_{u,v}$ is the abstract replacement of the quantity $\int_u^v X_{u,s}\, \mathrm{d}X_s$. It is far from unique: for any pair $(X, \mathbb{X})$ that fulfils the reqiurements of Definition (4.6) and any function $f \in C^{2\alpha}([0,T], V \otimes V)$, the pair $(X, (\mathbb{X}_{u,v} + f_{u,v})_{u,v \in [0,T]})$ also fulfils them (exercise!). This in particular covers all Itô-, Stratonovich-, and other stochastic integrals, but is actually much more flexible than that.

b) While $\mathcal{C}^\alpha([0,T],V)$ is a subset of the vector space $C^\alpha([0,T],V) \times C_2^{2\alpha}([0,T], V \otimes V)$, it not a *subspace*. Since Chen's relation is not linear (not conserved under addition), $\mathcal{C}^\alpha$ is not a vector space!

c) The case $\alpha > 1/2$ is possible but uninteresting, since then we can do Young integrals. The case $\alpha \leqslant 1/3$ is interesting, but not necessary for the case of Brownian motion. If one wants to treat $\alpha \leqslant 1/3$, one has to add higher order correction terms (in the spirit of (4.2)) to the approximations of the integral. Basically, one assumes more regularity of f there and does further Taylor expansions. The details are a bit messy, and we will not cover this case.

d) Chen's relation means that the values of $(X_t \mathbb{X}_{0,t})_{t \leqslant T}$ determine all values of $\boldsymbol{X}_{s,t}$ for $s, t \leqslant T$ (exercise!). Therefore even though it seems as if $\boldsymbol{X}$ is a function of two variables, $\boldsymbol{X} = (X, \mathbb{X})$ is indeed a „path“, i.e. a function of a one-dimensional variable with values in $V \times (V \otimes V)$.

Our next aim is to define $\int_0^t f(X_s) \mathrm{d}\boldsymbol{X}_s := \lim_{|P| \to 0} \sum_{[u,v) \in P} (f(X_u) X_{u,v} + f'(X_u) \mathbb{X}_{u,v})$ in the spirit of (4.2). Before we do this, we need to clarify what the expression $f'(X_u)\mathbb{X}_{u,v}$ is in higher dimensions.

(4.8) Derivatives in higher dimensions

For $f \in C^2(V, \mathcal{L}(V,W))$ and $X \in V$, we have $f(X_u) X_{u,v} \in W$. The total derivative of f is the map

$$Df : V \to \mathcal{L}(V, \mathcal{L}(V, W)) \simeq \mathcal{L}(V \otimes V, W).$$

that satisfies

$$\underbrace{[Df(v)]}_{\in \mathcal{L}(V\otimes V, W)}(e_i \otimes e_j) = \underbrace{[\partial_i f(v)]}_{\in \mathcal{L}(V,W)}(e_j) \in W$$

for all basis vectors e_i, e_j of V. Since $Df \in C^1(V, \mathcal{L}(V, \mathcal{L}(V, W)))$, we can define

$$D^2 f : V \to \mathcal{L}(V, \mathcal{L}(V, \mathcal{L}(V, W))) \simeq \mathcal{L}(V \otimes V \otimes V, W)$$

as $D^2 f = D(Df)$ in the same way as above, with W replaced by $\mathcal{L}(V, W)$. This means that

$$[Df(v)](e_i \otimes e_j \otimes e_k) = [\partial_i \partial_j f(v)](e_k) \in W.$$

The case $V = \mathbb{R}^d$ reduces the familiar case where e_i and e_j are used to pick the matrix element of the Hesse matrix. Since the target space of f is $\mathcal{L}(V, W)$, the additional basis vector e_k is needed to produce an element of W.

For $\boldsymbol{X} = (X, \mathbb{X}) \in \mathcal{C}^\alpha([0, T], V)$ and $f \in C^2(V, \mathcal{L}(V, W))$, we will frequently use the following abbreviations:

$$\begin{aligned} Y_s &:= f(X_s) \in \mathcal{L}(V, W), \\ Y'_s &:= Df(X_s) \in \mathcal{L}(V \otimes V, W) \simeq \mathcal{L}(V, \mathcal{L}(V, W)), \\ R^Y_{s,t} &:= Y_{s,t} - Y'_s X_{s,t} \in \mathcal{L}(V, W), \\ \Xi_{s,t} &:= Y_s X_{s,t} + Y'_s \mathbb{X}_{s,t} \in W. \end{aligned}$$

Here, $R^Y_{s,t}$ should be viewed as the „second order Taylor term“, and $\Xi_{s,t}$ as the approximation to an integral over the intervall $[s, t]$ when $|t - s|$ is small.

(4.9) Lemma

Let $f \in C^2_b$, $T > 0$. Writing $\|\cdot\|_\alpha$ instead of $\|\cdot\|_{T,\alpha}$ for brevity, we have

a) $Y \in C^\alpha([0, T], \mathcal{L}(V, W))$ with $\|Y\|_\alpha \leqslant \|Df\|_\infty \|X\|_\alpha$,

b) $Y' \in C^\alpha([0, T], \mathcal{L}(V \otimes V, W))$ with $\|Y'\|_\alpha \leqslant \|D^2 f\|_\infty \|X\|_\alpha$,

c) $R^Y \in C^{2\alpha}_2([0, T], \mathcal{L}(V, W))$ with $\|R^Y\|_{2\alpha,\mathrm{diag}} \leqslant \frac{1}{2}\|D^2 f\|_\infty \|X\|^2_\alpha$,

d) We have

$$\Xi_{u,w} - \Xi_{u,v} - \Xi_{v,w} = -R^Y_{u,v} X_{v,w} - Y'_{u,v} \mathbb{X}_{v,w}, \qquad (*)$$

and

$$\|\delta \Xi\|_{3\alpha} := \sup_{0 \leqslant u<v<w \leqslant T} \frac{|\Xi_{u,w} - \Xi_{u,v} - \Xi_{v,w}|}{|u - w|^{3\alpha}} \leqslant \|R^Y\|_{2\alpha,\mathrm{diag}} \|X\|_\alpha + \|Y'\|_\alpha \|\mathbb{X}\|_{2\alpha,\mathrm{diag}}.$$

Proof: a) we have

$$Y_v - Y_u = \int_0^1 \frac{\mathrm{d}}{\mathrm{d}h} f(X_u + hX_{u,v}) \,\mathrm{d}h = \int_0^1 \underbrace{[Df](X_u + hX_{u,v})}_{|.| \leqslant \|Df\|_\infty} X_{u,v} \,\mathrm{d}h.$$

Dividing by $|v - u|^\alpha$ and taking the supremum shows the claim.

b) The same as above for $Y'_v - Y'_u$.

c) We have

$$
\begin{aligned}
R^Y_{u,v} = Y_{u,v} - Y'_u X_{u,v} &\overset{\text{a)}}{=} \int_0^1 \big(Df(X_u + hX_{u,v}) - Df(X_u)\big)\,\mathrm{d}h\ X_{u,v} \\
&= \int_0^1 \mathrm{d}h \int_0^h \mathrm{d}s \Big[\frac{\mathrm{d}}{\mathrm{d}s}\big[Df(X_u + sX_{u,v})\big]\Big] X_{u,v} = \\
&= \int_0^1 \mathrm{d}h \int_0^h \mathrm{d}s\, \big[D^2 f(X_u + sX_{u,v})\big](X_{u,v} \otimes X_{u,v}).
\end{aligned}
$$

Since $X_{u,v} \otimes X_{u,v} \in C_2^{2\alpha}$ and the integral is bounded by $\frac{1}{2}\|D^2 f\|_\infty$, the claim follows.

d) The calculations we did at the end of (4.2) lead (in the same notation!) the equality $(*)$, and the claimed estimate then follows from a), b) and c). □

(4.10) Notation

Let $T, \alpha, \beta > 0$. We write

$$
\Delta_T := \{(s,t) \in [0,T]^2 : 0 \leqslant s \leqslant t \leqslant T\}.
$$

For $\Xi : \Delta_T \to W$ we define

$$
\|\delta\Xi\|_\beta := \sup\Big\{\frac{|\Xi_{u,w} - \Xi_{u,v} - \Xi_{v,w}|}{|u-w|^\beta} : 0 \leqslant u < v < w \leqslant T\Big\}, \qquad \|\Xi\|_{\alpha,\beta} := \|\Xi\|_\alpha + \|\delta\Xi\|_\beta,
$$

and

$$
C_2^{\alpha,\beta}([0,T],W) := \{\Xi : \Delta_T \to W : \|\Xi\|_{\alpha,\beta} < \infty, \Xi(t,t) = 0\ \forall t \in [0,T]\}.
$$

(4.11) Sewing Lemma

Let $0 < \alpha < 1 < \beta$, and $T > 0$.

a) For all $\Xi \in C_2^{\alpha,\beta}([0,T],W)$ and all $s < t \in \Delta_T$, the limit

$$
(\mathcal{I}\Xi)_{s,t} := \lim_{|P| \to 0} \sum_{[u,v) \in P \cap [s,t)} \Xi_{u,v}
$$

exists, and fulfills

$$
\big|(\mathcal{I}\Xi)_{s,t} - \Xi_{s,t}\big| \leqslant C|t-s|^\beta \qquad (*)
$$

with the constant $C = \|\delta\Xi\|_\beta 2^\beta \zeta(\beta)$, where ζ is the Riemann zeta function.

b) The map $t \mapsto (\mathcal{I}\Xi)_{0,t}$ is in $C^\alpha([0,T],W)$, and

$$
\|\mathcal{I}\Xi\|_\alpha \leqslant \max\big\{2^\beta T^{\beta-\alpha}\zeta(\beta), 1\big\}\|\Xi\|_{\alpha,\beta}.
$$

c) The map $\Xi \mapsto \big((\mathcal{I}\Xi)_{0,t}\big)_{t \leqslant T}$ is the unique linear map $C_2^{\alpha,\beta} \to C^\alpha$ sucht that $(*)$ holds for some $C < \infty$.

Proof: Recall that $\delta\Xi_{u,r,v} := \Xi_{u,v} - \Xi_{u,r} - \Xi_{r,v}$ for $u < r < v$. By the assumption $\Xi \in C_2^{\alpha,\beta}$, the inequality

$$
\sup_{r \in [u,v)} |\delta\Xi_{u,r,v}| \leqslant \|\delta\Xi\|_\beta |u-v|^\beta
$$

holds for all $u < v$. This means that the inequality (A) from the proof of Theorem (3.4) holds. Following the proof of that theorem from that point on (word by word!), we obtain a).

For b), we estimate

$$\left|(\mathcal{I}\Xi)_{s,t}\right| \leqslant \left|(\mathcal{I}\Xi)_{s,t} - \Xi_{s,t}\right| + \left|\Xi_{s,t}\right| \overset{(*)}{\leqslant} \|\delta\Xi\|_\beta 2^\beta \zeta(\beta)|t-s|^\beta + \|\Xi\|_\alpha |t-s|^\alpha.$$

Since $\alpha < \beta$, we have $|t-s|^\beta \leqslant T^{\beta-\alpha}|t-s|^\alpha$, and b) follows.

c) is left as an exercise. □

(4.12) Definition

Let $\boldsymbol{X} \in \mathcal{C}^\alpha([0,T], V)$. The **rough path norm** pf $\boldsymbol{X}$ is

$$\|\!|\boldsymbol{X}|\!\|_\alpha := \|X\|_\alpha + \sqrt{\|\mathbb{X}\|_{2\alpha,\text{diag}}}.$$

For $\boldsymbol{X}, \boldsymbol{Y} \in \mathcal{C}^\alpha([0,T], V)$, the **rough path distance** of $\boldsymbol{X}$ and $\boldsymbol{Y}$ is given by

$$\varrho_\alpha(\boldsymbol{X}, \boldsymbol{Y}) := \|X - Y\|_\alpha + \|\mathbb{X} - \mathbb{Y}\|_{2\alpha,\text{diag}},$$

where $\|\mathbb{X} - \mathbb{Y}\|_{2\alpha,\text{diag}} := \sup_{0 \leqslant s < t \leqslant T} \frac{1}{|t-s|^{2\alpha}}|\mathbb{X}_{s,t} - \mathbb{Y}_{s,t}|$.

(4.13) Remark

a) The square root in the definition of $\|\!|\cdot|\!\|_\alpha$ makes the expression homogenous in the following sense: the only way to „multiply" a rough path by a constant c so that Chen's relation still holds is to define $c\boldsymbol{X} = (cX, c^2\mathbb{X})$. Then the definition of our norm guarantees $\|\!|c\boldsymbol{X}|\!\|_\alpha = c\,\|\!|\boldsymbol{X}|\!\|_\alpha$. Note that no square root appears in the definition of ϱ_α, which seems odd at first - would we not like an equality like $\varrho_\alpha(\boldsymbol{X}, \boldsymbol{Y}) = \|\!|\boldsymbol{X} - \boldsymbol{Y}|\!\|_\alpha$ to hold? The answer is that this is not relevant, because $\boldsymbol{X} - \boldsymbol{Y}$ is in general not a rough path: Chen's relation is not linear! On the other hand, $c\boldsymbol{X}$ *is* a rough path for $c \in \mathbb{R}$, which is why we want to keep the homogeniety of the norm intact.

b) Note on the other hand that for two rough paths $\boldsymbol{X}$ and $\boldsymbol{Y}$, the expression $\|X - Y\|_\alpha + \sqrt{\|\boldsymbol{X} - \boldsymbol{Y}\|_{2\alpha,\text{diag}}}$ does make sense exactly as given in Definition (4.12 b). $\boldsymbol{X} - \boldsymbol{Y}$ is then viewed as a function of two variables, as in Definition (4.5). So ϱ_α is the „distance" that is derived from the „norm" $(X, F) \mapsto \|X\|_\alpha + \|F\|_{2\alpha,\text{diag}}$ on the vector space $C^\alpha \oplus C_2^{2\alpha}$, and in this sense the omission of the square root in this expression is the sensible choice. Note however the quotation marks: the rough path norm is not actually a norm on $C^\alpha \oplus C_2^{2\alpha}$ because it is equal to zero on all constant functions. The same problem exists for ρ_α. It is not hard to see that the map $\boldsymbol{X} \mapsto \|\!|\boldsymbol{X}|\!\|_\alpha + |X_0|$ is a norm, and that $(\boldsymbol{X}, \boldsymbol{Y}) \mapsto \rho_\alpha(\boldsymbol{X}, \boldsymbol{Y}) + |Y_0 - X_0|$ is a metric.

d) Below and in the remainder of the lecture notes, we will use the notation $\|F\|_{C_b^k} = \sum_{j=0}^k \|D^j F\|_\infty$.

(4.14) Theorem

Let $\boldsymbol{X} = (X, \mathbb{X}) \in \mathcal{C}^\alpha([0,T], V)$, $\alpha \in (1/3, 1/2]$ and $F \in C_b^2(V, \mathcal{L}(V, W))$. We define

$$\Xi_{u,v} := F(X_u)X_{u,v} + [DF](X_u)\mathbb{X}_{u,v}.$$

a) The limit

$$\int_s^t F(X_r)\,\mathrm{d}\boldsymbol{X}_r := \lim_{|P|\to 0} \sum_{[u,v)\in P\cap[s,t)} \Xi_{u,v}$$

exists for all $s<t$, and

$$\Big|\int_s^t F(X_r)\,\mathrm{d}\boldsymbol{X}_r - \Xi_{s,t}\Big| \leqslant C(\alpha)\|F\|_{C_b^2}\Big(\|X\|_\alpha^3 + \|X\|_\alpha \|\mathbb{X}\|_{2\alpha,\mathrm{diag}}\Big)|t-s|^{3\alpha},$$

with $C(\alpha) = 2\max\{2^{3\alpha}T^{2\alpha}\zeta(3\alpha), 1\}$.
b) The map $t \mapsto \int_0^t F(X_r)\mathrm{d}\boldsymbol{X}_r$ is in C^α, and

$$\Big\|\int_0^\cdot F(X_r)\,\mathrm{d}\boldsymbol{X}_r\Big\|_\alpha \leqslant 12C(\alpha)\|F\|_{C_b^2}\big(\, \|\!|\boldsymbol{X}|\!\|_\alpha \vee (\|\!|\boldsymbol{X}|\!\|_\alpha^{1/\alpha}\max\{T^{1-\alpha},1\})\big).$$

Proof: a) will follow from the Sewing Lemma (4.11) once we show that $\Xi \in C_2^{\alpha,3\alpha}$. We have

$$|\Xi_{u,v}| \leqslant \|F\|_\infty |X_{u,v}| + \|DF\|_\infty |\mathbb{X}_{u,v}| \leqslant \|F\|_{C_b^2}\big(\|X\|_\alpha |v-u|^\alpha + \|\mathbb{X}\|_{2\alpha,\mathrm{diag}}|v-u|^{2\alpha}\big),$$

and

$$\|\delta\Xi\|_{3\alpha} \overset{(4.9d)}{\leqslant} \|R^Y\|_{2\alpha,\mathrm{diag}}\|X\|_\alpha + \|Y'\|_\alpha \|\mathbb{X}\|_{2\alpha,\mathrm{diag}} \overset{(4.9b,c)}{\leqslant} \|D^2F\|_\infty\big(\frac{1}{2}\|X\|_\alpha^3 + \|X\|_\alpha\|\mathbb{X}\|_{2\alpha,\mathrm{diag}}\big).$$

The sewing lemma now yields the claim.
b) We estimate $|\int_s^t F(X_r)\,\mathrm{d}\boldsymbol{X}_r| \leqslant |\int_s^t F(X_r)\,\mathrm{d}\boldsymbol{X}_r - \Xi_{s,t}| + |\Xi_{s,t}|$. For the first term, we use a) and bound it by $C(\alpha)\|F\|_{C_b^2}(\|X\|_\alpha^3 + \|X\|_\alpha\|\mathbb{X}\|_{2\alpha,\mathrm{diag}})|t-s|^{3\alpha}$. The second term is bounded by $\|F\|_\infty\|X\|_\alpha|t-s|^\alpha + \|DF\|_\infty\|\mathbb{X}\|_{2\alpha,\mathrm{diag}}|t-s|^{2\alpha}$. Since $C(\alpha)\geqslant 1$ and $\|\boldsymbol{X}\|_{2\alpha,\mathrm{diag}} \leqslant \|\!|\boldsymbol{X}|\!\|_\alpha^2$, we obtain the intequality

$$\Big|\int_s^t F(X_r)\,\mathrm{d}\boldsymbol{X}_r\Big| \leqslant 2C(\alpha)\|F\|_{C_b^2}\sum_{i=1}^3 \|\!|\boldsymbol{X}|\!\|_\alpha^i\,|t-s|^{i\alpha}.$$

If $|t-s|^\alpha \leqslant \|\!|\boldsymbol{X}|\!\|_\alpha^{-1}$, then we can divide the above inequality by $|t-s|^\alpha$ and obtain

$$\frac{1}{|t-s|^\alpha}\Big|\int_s^t F(X_r)\,\mathrm{d}\boldsymbol{X}_r\Big| \leqslant 6C(\alpha)\|F\|_{C_b^2}\,\|\!|\boldsymbol{X}|\!\|_\alpha\,. \qquad (*)$$

If $|t-s|^\alpha > \|\!|\boldsymbol{X}|\!\|_\alpha^{-1}$, then $h := \|\!|\boldsymbol{X}|\!\|_\alpha^{-1/\alpha}$ fulfils $1 < \frac{|t-s|}{h}$. We set $t_j = (s+jh)\wedge t$, have $|t_j - t_{j-1}| \leqslant \|\!|\boldsymbol{X}|\!\|_\alpha^{-1/\alpha} = h$, and thus get

$$\begin{aligned}\Big|\int_0^t F(X_r)\,\mathrm{d}\boldsymbol{X}_r\Big| &\leqslant \sum_{0\,\leqslant\, j<(t-s)/h}\Big|\int_{t_j}^{t_{j+1}} F(X_r)\,\mathrm{d}\boldsymbol{X}_r\Big| \overset{(*)}{\leqslant} \\ &\leqslant \underbrace{\big((t-s)/h+1\big)}_{\leqslant\,2(t-s)/h}\cdot 6C(\alpha)\|F\|_{C_b^2}\underbrace{\|\!|\boldsymbol{X}|\!\|_\alpha\,h^\alpha}_{=1} \leqslant 12C(\alpha)\|F\|_{C_b^2}\,\|\!|\boldsymbol{X}|\!\|_\alpha^{1/\alpha}\,|t-s|.\end{aligned}$$

When we divide this by $|t-s|^\alpha$, together with the case $(*)$ we obtain

$$\frac{1}{|t-s|^\alpha}\Big|\int_0^t F(X_r)\,\mathrm{d}\boldsymbol{X}_r\Big| \leqslant 6C(\alpha)\|F\|_{C_b^2}(\|\!|\boldsymbol{X}|\!\|_\alpha \vee \|\!|\boldsymbol{X}|\!\|_\alpha^{1/\alpha}\max\{T^{1-\alpha},1\}),$$

which shows the claim. □

We have just shown the existence of the rough integral for every given rough path. In order to make the connection to the theory of Brownian motion, we need to show that almost surely, a Brownian motion path B can be enhanced to a rough path. This means that we need to define what $\mathbb{B}$ should be, check that $(B, \mathbb{B})$ fulfils Chen's relation, and that the rough path norm of $(B, \mathbb{B})$ is finite for $\alpha \in (1/3, 1/2]$. We start by defining $\mathbb{B}$.

(4.15) Definition

Let B be a d-dimensional Brownian motion. For every $\omega \in \Omega$ and all $0 \leqslant s < t \leqslant T$, we define

$$\mathbb{B}^{\text{Itô}}_{s,t}(\omega) := \Big(\int_0^t B_r \otimes \mathrm{d}B_r \Big)(\omega) - \Big(\int_0^s B_r \otimes \mathrm{d}B_r \Big)(\omega) - B_s(\omega) \otimes B_{s,t}(\omega).$$

(4.16) Lemma

a) $\mathbb{B}^{\text{Itô}}$ fulfils Chen's relation for all s, t, ω.

b) For all $s < t$ we have $\mathbb{B}^{\text{Itô}}_{s,t} = \int_s^t B_{s,r} \otimes \mathrm{d}B_r$ almost surely.

Proof: Let us write $\mathbb{B}$ instead of $\mathbb{B}^{\text{Itô}}$ for simplicity. Define $J_v = \int_0^v B_r \otimes \mathrm{d}B_r$. By the definition of $\mathbb{B}$, we have

$$\begin{aligned} \mathbb{B}_{u,v} + \mathbb{B}_{v,w} + B_{u,v} \otimes B_{v,w} &= J_v - J_u - B_u \otimes B_{u,v} + J_w - J_v - B_v \otimes B_{v,w} + B_{u,v} \otimes B_{v,w} = \\ &= J_w - J_u - (B_u \otimes B_{u,w} - B_u \otimes B_{v,w}) - B_v \otimes B_{v,w} + B_{u,v} \otimes B_{v,w} = \mathbb{B}_{u,w}. \end{aligned}$$

This proves a). For b), we use that Itô-integrals are additive almost surely, and get

$$\Big(\int_s^t B_{s,r} \otimes \mathrm{d}B_r \Big)_{i,j} \equiv \int_s^t B^i_{s,r} \, \mathrm{d}B^j_r = \int_0^t B^i_r \, \mathrm{d}B^j_r - \int_0^s B^i_r \, \mathrm{d}B^j_r - B^i_s \int_s^t \mathrm{d}B^j_r = \mathbb{B}^{i,j}_{s,t}$$

almost surely. □

(4.17) Remark

Definition (4.15) is made so that Chen's relation holds for *all* s, t and ω. The equality in (4.16 b), on the other hand, extends to all s, t simultaneously by path continuity of Brownian motion and its integrals. So, we could equally well define $\mathbb{B}^{\text{Itô}}_{s,t} := \int_s^t B_{s,r} \otimes \mathrm{d}B_r$, in this case Chen's relation would only hold on a set of probability one, but would hold on that set for all s, t simultaneously.

We have made sure that Chens relation holds for $\mathbb{B}^{\text{Itô}}$; we also know that the first component of the pair $(B, \mathbb{B}^{\text{Itô}})$ is in C^α for all $\alpha \in (1/3, 1/2)$. What we still need is that $\mathbb{B}^{\text{Itô}}$ is in $C_2^{2\alpha}$ for the same range of α. This will be our next task.

(4.18) Kolmogorov-Chentsov-Theorem, rough path version

Let $(X_t)_{0 \leqslant t \leqslant T}$ be a V-valued stochastic process, and let $(\mathbb{X}_{s,t})_{0 \leqslant s < t \leqslant T}$ be a family of $V \otimes V$-valued random variables on the same probability space as X. Assume that

(i): The pair $(X(\omega), \mathbb{X}(\omega))$ fulfils Chen's relation for all ω.

(ii): There exists $q \geqslant 2$, $\beta > 1/q$ and $C < \infty$ with

$$\mathbb{E}(|X_{s,t}|^q) \leqslant C|t-s|^{\beta q} \qquad \text{and} \qquad \mathbb{E}(|\mathbb{X}_{s,t}|^{q/2}) \leqslant C|t-s|^{\beta q}$$

for all $s < t$. Then
a) with $D = \{k2^{-n} : k \in \mathbb{N}, n \in \mathbb{N}\} \cap [0,T]$, for all $\alpha \in [0, \beta - \frac{1}{q})$ we have

$$\mathbb{E}\big(\|X\|^q_{D,\alpha}\big) < \infty, \qquad \text{and} \qquad \mathbb{E}\big(\|\mathbb{X}\|^{q/2}_{D,2\alpha,\text{diag}}\big) < \infty.$$

b) there exists a version of $(X, \mathbb{X})$ that fulfils $(X(\omega), \mathbb{X}(\omega)) \in \mathcal{C}^\alpha([0,T])$ for all ω.

Proof: The statement $\mathbb{E}(\|X\|^q_{D,\alpha}) < \infty$ was already proved in Theorem (1.36). The other statement is proved in a similar manner. We write

$$\begin{aligned} D_n &:= \{2^{-n}k : k \in \mathbb{N}\} \cap [0,T], \\ \mathbb{K}_n(\omega) &:= \max\{|\mathbb{X}_{t,t+2^{-n}}(\omega)| : t \in D_n\}, \\ K_n(\omega) &:= \max\{|X_{t,t+2^{-n}}(\omega)| : t \in D_n\}. \end{aligned}$$

Then

$$\begin{aligned} \mathbb{E}(\mathbb{K}_n^{q/2}) &\leqslant \sum_{t \in D_n} \mathbb{E}\big(|\mathbb{X}_{t,t+2^{-n}}|^{q/2}\big) \overset{(ii)}{\leqslant} 2^n T C 2^{-n\beta q} = CT(2^{-n})^{\beta q - 1}, \qquad (*) \\ \mathbb{E}(K_n^q) &\leqslant CT(2^{-n})^{\beta q - 1}. \end{aligned}$$

As in the proof of Theorem (1.36) we have that for given $s < t$ with $|t-s| < 1/2$, there exists $j \in \mathbb{N}$ with $2^{-j} < t-s \leqslant 2^{-j+1}$ and a chain $s = t_0, t_1, \ldots, t_n = t$ of points from D such that two consecutive points have distance 2^{-m} for some $m \geqslant j$ and each distance appears at most twice in the chain. For arbitrary $s < t \in [0,T]$ we can then find a similar chain (by concatenation) such that each distance

The trick in the proof of (1.36) was to write $X_{s,t} = \sum_{i=1}^n X_{t_{i-1},t_i}$ as a telescopic sum and then use the triangle inequality. This does not work here directly because $\mathbb{X}_{s,t}$ is not additive, but Chen's relation helps us again. We have

$$\mathbb{X}_{s,t} = \sum_{i=1}^n \mathbb{X}_{t_{i-1},t_i} + \sum_{i=1}^n X_{t_{i-1},t_i} \otimes X_{t_i,t},$$

which is proved by iterative application of Chen's relation (exercise!). Writing $X_{t_i,t} = \sum_{k=i+1}^n X_{t_{k-1},t_k}$, we then obtain (with $K_\ell(\omega) = \max\{|X_{t,t+2^{-n}}(\omega)| : t \in D_n\}$)

$$\begin{aligned} |\mathbb{X}_{s,t}| &\leqslant \sum_{i=1}^n |\mathbb{X}_{t_{i-1},t_i}| + \sum_{i=1}^n |X_{t_{i-1},t_i}| \max_{0 \leqslant k \leqslant n} |X_{t_k,t}| \leqslant \sum_{i=1}^n |\mathbb{X}_{t_{i-1},t_i}| + \Big(\sum_{i=1}^n |X_{t_{i-1},t_i}|\Big)^2 \\ &\leqslant 2\sum_{\ell=j}^\infty \mathbb{K}_\ell + \Big(2\sum_{\ell=j}^\infty K_\ell\Big)^2 \overset{|t-s|>2^{-j}}{\leqslant} 2|t-s|^{2\alpha} \sum_{\ell=j}^\infty 2^{2\ell\alpha}\mathbb{K}_\ell + 4|t-s|^{2\alpha}\Big(\sum_{\ell=j}^\infty 2^{\ell\alpha} K_\ell\Big)^2 \end{aligned}$$

for all s,t with $|t-s|<1/2$. For general $s,t\in[0,T]$, we can choose at most $2T+1$ points $s=s_0<s_1\ldots<s_n=t$ with $s_{i+1}-s_i<1/2$ for all i. Then, as above,

$$\begin{aligned}|\mathbb{X}_{s,t}| &\leqslant \sum_{i=1}^{n}|\mathbb{X}_{s_{i-1},s_i}| + \Big(\sum_{i=1}^{n}|X_{s_{i-1},s_i}|\Big)^2 \\ &\leqslant |t-s|^{2\alpha}\Big((4T+2)\sum_{\ell=j}^{\infty}2^{2\ell\alpha}\mathbb{K}_\ell + (8T+4)\Big(\sum_{\ell=j}^{\infty}2^{\ell\alpha}K_\ell\Big)^2 + (4T+2)^2\Big(\sum_{\ell=j}^{\infty}2^{\ell\alpha}K_\ell\Big)^2\Big).\end{aligned}$$

Dividing by $|t-s|^{2\alpha}$ and taking the supremum over $\{(s,t):s,t\in D\cap[0,T]:|s-t|<1/2\}$, we get

$$\begin{aligned}\mathbb{E}\big(\|\mathbb{X}\|_{D,2\alpha,\mathrm{diag}}^{q/2}\big)^{2/q} &\leqslant \tilde{C}\,\Big\|\sum_{\ell=0}^{\infty}2^{2\ell\alpha}\mathbb{K}_\ell + \Big(\sum_{\ell=0}^{\infty}2^{\ell\alpha}K_\ell\Big)^2\Big\|_{L^{q/2}} \\ &\leqslant \tilde{C}\Big(\sum_{\ell=0}^{\infty}2^{2\ell\alpha}\|\mathbb{K}_\ell\|_{L^{q/2}} + \Big\|\Big(\sum_{\ell=0}^{\infty}2^{\ell\alpha}K_\ell\Big)^2\Big\|_{L^{q/2}}\Big),\end{aligned}$$

where the constant $\tilde{C}$ depends on T as a quadratic polynomial. The first term above is equal to

$$\tilde{C}\sum_{\ell=0}^{\infty}2^{2\ell\alpha}\mathbb{E}(\mathbb{K}_\ell^{q/2})^{2/q} \overset{(*)}{\leqslant} \tilde{C}T^{\frac{2}{q}}C^{2/q}\sum_{\ell=0}^{\infty}2^{2\ell\alpha}2^{-\ell(\beta q-1)\frac{2}{q}} = \tilde{C}T^{\frac{2}{q}}C^{2/q}\sum_{\ell=0}^{\infty}2^{-2\ell(\beta-1/q-\alpha)},$$

and thus finite for $\alpha<\beta-1/q$. The second term is equal to $\big\|\sum_{\ell=0}^{\infty}2^{\ell\alpha}K_\ell\big\|_{L^q}^2$ which is finite under the same condition by the first part of the proof, see the proof of Theorem (1.36) for reference. This finishes the proof of a). The proof of b) is literally the same as the one for Theorem (1.38). □

What remains is to show that $\mathbb{B}^{\text{Itô}}$ fulfils assumption (4.18 (ii)).

(4.19) Proposition

For all $T>0$, $q=4k$ with $k\in\mathbb{N}$, there exists $C_q<\infty$ with

$$\mathbb{E}(|\mathbb{B}_{s,t}|^{q/2}) \leqslant C_q|t-s|^{q/2} \qquad \forall s,t\in[0,T].$$

As a consequence, the second inequality of (4.18 (ii)) holds for all $\beta<1/2$.

Proof: We start with some preparations. First of all, we use that

$$\mathbb{B}_{s,t} = \int_s^t B_{s,r}\otimes \mathrm{d}B_r \sim \int_0^{t-s} B_r\otimes\mathrm{d}B_r,$$

where the first equality holds almost surely. Second, we use the sum norm for the matrix $\mathbb{B}_{s,t}\in\mathbb{R}^{d\times d}$, and then the triangle inequality gives

$$\|\mathbb{B}_{s,t}\|_{L^{q/2}} \leqslant \sum_{i,j=1}^{d}\Big\|\int_0^{t-s}B_r^{(i)}\mathrm{d}B_r^{(j)}\Big\|_{L^{q/2}}.$$

So we only need to see that each term of that sum is bounded by a constant times $(|t-s|^{q/2})^{2/q} = |t-s|$. For $i = j$, this follows from the Itô formula and the fact that for a centered Gaussian random variable Y with variance σ^2, we have $\mathbb{E}(Y^{2k}) = (\sigma^2)^k (2k)!!$. We have for $q/2 = 2k$

$$\Big\| \int_0^{t-s} B_r^{(i)} \mathrm{d}B_r^{(i)} \|_{L^{q/2}} = \frac{1}{2} \|(B_{t-s}^{(i)})^2 - (t-s)\|_{L^{2k}} \leqslant \frac{1}{2} \|(B_{t-s}^{(i)})^2\|_{L^{2k}} + \|(t-s)\|_{L^{2k}} =$$
$$= \big[(4k)!!|t-s|^{2k}\big]^{\frac{1}{2k}} + |t-s|.$$

For $i \neq j$ (wlog let $i = 1, j = 2$), let $\mathcal{F}^{(1)}$ be the σ-algebra generated by $(B_t^{(1)})_{t \geqslant 0}$. We claim that for all $\alpha > 0$ and all $t > 0$, the equality

$$\mathbb{E}\Big(\mathrm{e}^{\mathrm{i}\alpha \int_0^t B_s^{(1)} \mathrm{d}B_s^{(2)}} \,\Big|\, \mathcal{F}^{(1)} \Big)(\bar{\omega}) = \exp\Big(-\frac{\alpha^2}{2} \int_0^t (B_s^{(1)}(\bar{\omega}))^2 \,\mathrm{d}s \Big) \qquad (*)$$

holds for almost all $\bar{\omega}$. (In other words, the distribution of $\int_0^t B_s^{(1)} \,\mathrm{d}B_s^{(2)}$ conditional on $\mathcal{F}^{(1)}$ is Gaussian with mean zero and variance $\int_0^t (B_s^{(1)}(\bar{\omega}))^2 \,\mathrm{d}s$.) To see the claimed equality, we recall that $\sum_{[u,v) \in P} B_u^{(1)} B_{u,v}^{(2)}$ converges in L^2 to $\int_0^t B_s^{(1)} \mathrm{d}B_s^{(2)}$ when $|P| \to 0$, and thus almost surely along a subsequence. Along this subsequence, then also $\exp(\mathrm{i}\alpha \sum_{[u,v) \in P} B_u^{(1)} B_{u,v}^{(2)})$ converges to $\exp(\mathrm{i}\alpha \int_0^t B_s^{(1)} \mathrm{d}B_s^{(2)})$ almost surely, and since the complex exponential function is bounded, dominated convergence implies the convergence in L^2. Since conditional expectation is continuous (it is a projection), we conclude that for the subsequence (P_n) of partitions

$$\lim_{n \to \infty} \mathbb{E}\Big(\mathrm{e}^{\mathrm{i}\alpha \sum_{[u,v) \in P_n} B_u^{(1)} B_{u,v}^{(2)}} \,\Big|\, \mathcal{F}^{(1)} \Big) = \mathbb{E}\Big(\mathrm{e}^{\mathrm{i}\alpha \int_0^t B_s^{(1)} \mathrm{d}B_s^{(2)}} \,\Big|\, \mathcal{F}^{(1)} \Big)$$

in L^2, and thus almost surely along a further subsequence (also called (P_n)). For each n, the fact that $B^{(1)}$ is $\mathcal{F}^{(1)}$-measurable and $B^{(2)}$ is independent from $\mathcal{F}^{(1)}$ implies that, almost surely,

$$\mathbb{E}\Big(\mathrm{e}^{\mathrm{i}\alpha \sum_{[u,v) \in P_n} B_u^{(1)} B_{u,v}^{(2)}} \,\Big|\, \mathcal{F}^{(1)} \Big)(\bar{\omega}) = \mathbb{E}\Big(\mathrm{e}^{\mathrm{i}\alpha \sum_{[u,v) \in P_n} B_u^{(1)}(\bar{\omega}) B_{u,v}^{(2)}} \Big) = \mathrm{e}^{-\frac{\alpha^2}{2} \sum_{[u,v) \in P_n} (B_u^{(1)}(\bar{\omega}))^2 (v-u)},$$

where the last inequality comes from the Gaussian and independent increments of $B^{(2)}$. The sum in the exponent on the right hand side is a Riemann sum and converges to $\int_0^t [B_s^{(1)}(\bar{\omega})]^2 \,\mathrm{d}s$, which shows the claimed equality $(*)$.

What remains to do is to take expectaion in $(*)$, then differentiate both sides $2k = q/2$ times with respect to α, and finally to evaluate at $\alpha = 0$. On the left, this gives $\mathbb{E}((\mathbb{B}_{0,t-s})^{q/2})$. On the right, the fact that we evaluate at $\alpha = 0$ means that what we get is the term where precisely k derivatives hit the exponential (each producing a prefactor of $\int_0^t (B_s^{(1)}(\bar{\omega}))^2 \,\mathrm{d}s$ under the integral), while the other k derivatives act on the existing prefactor from earlier differentiations. This means that

$$\mathbb{E}\Big(\big(\int_0^{t-s} B_r^{(1)} \mathrm{d}B_r^{(2)} \,\Big|\, \mathcal{F}^{(1)} \big)^{2k} \Big) = C_{2k} \mathbb{E}\Big(\big(\int_0^{t-s} [B_r^{(1)}]^2 \,\mathrm{d}r \big)^k \Big) \leqslant$$
$$\leqslant C_{2k} |t-s|^k \mathbb{E}\Big(\sup\{|B_r|^2 : r \leqslant |t-s|\}^k \Big) = C_{2k} |t-s|^k \mathbb{E}\big(\max_{r \leqslant |t-s|} (B_r)^{2k} \big) =$$
$$\overset{(2.48a)}{=} C_{2k} |t-s|^k \mathbb{E}\big(|B_{t-s}|^{2k} \big) = C_{2k} |t-s|^k (2k)!! |t-s|^k.$$

This proves the claim. □

(4.20) Corollary

For all $\alpha \in (1/2, 1/2]$ and all $T > 0$, we have

$$\boldsymbol{B}^{\text{Itô}} := (B, \mathbb{B}^{\text{Itô}}) \in \mathcal{C}^\alpha([0,T], \mathbb{R}^d).$$

Proof: just combine (4.16), (4.18) and (4.19). □

(4.21) Remark

a) We have now succeeded to give a *pathwise* interpretation of the Itô integral $\int_0^t f(B_s)\,\mathrm{d}B_s$, at least for $f \in C^2$. In the past, we were very careful to always write $\left[\int_0^t f(B_s)\,\mathrm{d}B_s\right](\omega)$ to emphasize that for the Itô integral is an L^2 limit, and all paths need to work together to make sense of it. For rough integrals, we can pick ($\mathbb{P}$-almost) any Brownian path, compute its second rough path component $\boldsymbol{B}^{\text{Itô}}$ as in (4.15) (not that this is still an Itô integral, i.e. no pathwise integration here!), and then compute all integrals $\int_0^t f(B_s(\omega))\,\mathrm{d}\boldsymbol{B}_s^{\text{Itô}}(\omega)$ as rough integrals, i.e. pathwise.

b) But what have we actually achieved by this? First of all, instead of having $\int_0^t f(B_s)\,\mathrm{d}B_s$ as a different L^2-limit for each f, we have a *single* L^2 limit which is then used to make pathwise sense of stochastic integrals for all $f \in L^2$. Moreover, we will soon see that with this single L^2 limit, we can even define pathwise integrands for much more general inegrands than those of the form $f(B_s)$ - in particular, we will not need the integrands to be adapted. This is an important advantage, since the need to have adapted integrands has always been a serious limitation of stochastic integration; there have been some ad hoc ways around this, but no systematic one.

c) The only ambiguity left in the stochastic integral is thus the choice of the second rough path component $\boldsymbol{B}$. Here, $\boldsymbol{B}^{\text{Itô}}$ is not the most natural choice for many applications in rough paths; the reason is that the theory of rough paths does not rely on martingale properties as heavily as the classical theory does, and it is therefore attractive to make another choice for $\boldsymbol{B}$, namely the one that obeys the chain rule. Rough paths that obey the chain rule are important in general, and we now study them.

(4.22) Definition

For $\mathbb{X} \in V \otimes V$, the **symmetric part** of $\mathbb{X}$ is the element of $V \otimes V$ with components

$$\text{Sym}(\mathbb{X})^{i,j} := \frac{1}{2}(\mathbb{X}^{i,j} + \mathbb{X}^{j,i}),$$

where $\mathbb{X}^{i,j}$ is the component of $\mathbb{X}$.

(4.23) Definition

A rough path $\boldsymbol{X}$ is called **geometric** if $\text{Sym}(\mathbb{X}_{s,t}) = \frac{1}{2} X_{s,t} \otimes X_{s,t}$ for all $s < t$. We write

$$\mathcal{C}_{\mathrm{g}}^\alpha := \{\boldsymbol{X} \in \mathcal{C}^\alpha : \boldsymbol{X} \text{ is geometric}\}.$$

(4.24) Remarks

a) Consider a differentiable function X with its natural second component $\mathbb{X}_{s,t} = \int_s^t X_{s,r} \otimes X'_r \, \mathrm{d}r$. Then $(X, \mathbb{X})$ is a geometric rough path for all $\alpha \leqslant 1$ (proof: exercise). This explain why the equation $\mathrm{Sym}(\mathbb{X}_{s,t}) = \frac{1}{2} X_{s,t} \otimes X_{s,t}$ is also sometimes called the chain rule.

b) Geometric rough paths are „almost“ the closure of „smooth“ paths under the rough path distance. More precisely, let

$$\begin{aligned} \bar{\varrho}_\alpha : \quad & \big(C^\alpha([0,T], V) \times C_2^{2\alpha}([0,T], V \otimes V)\big)^2 \to \mathbb{R}_0^+, \\ & \big((X, \mathbb{X}), (Y, \mathbb{Y})\big) \mapsto \|Y - X\|_\alpha + \|\mathbb{Y} - \mathbb{X}\|_{2\alpha,\mathrm{diag}} + |Y_0 - X_0| \end{aligned}$$

be the metric on $C^\alpha([0,T], V) \times C_2^{2\alpha}([0,T], V \otimes V)$ that corresponds to the rough path distance ϱ_α; see also remark (4.13 b). Let $\mathcal{C}_{\mathrm{g}}^{0,\alpha}([0,T])$ be the closure of the subset

$$\Big\{\big(X, \big(\int_s^t X_{s,r} \otimes X'_r \, \mathrm{d}r\big)_{0 \leqslant s < t \leqslant T}\big) : X \in C^1\Big\} \subset C^\alpha([0,T], V) \times C_2^{2\alpha}([0,T], V \otimes V)$$

with respect to this metric, i.e. the rough paths that can be approximated by smooth functions and their natural second components. Then we have for all $0 < \alpha < \beta < 1$

$$\mathcal{C}_{\mathrm{g}}^{\beta} \subset \mathcal{C}_{\mathrm{g}}^{0,\alpha} \subsetneqq \mathcal{C}_{\mathrm{g}}^{\alpha} = \overline{\mathcal{C}_{\mathrm{g}}^{\alpha}}^{\bar{\varrho}_\alpha}.$$

In words: the set of α-geometric rough paths complete under $\bar{\varrho}_\alpha$, but is *strictly larger* than the set of approximable α rough paths. It is however smaller than the set of approximable rough paths when we are allowed to approximate C_{g}^{β} in the slighty weaker $\bar{\varrho}_\alpha$ metric.

The proof of these important relations will be given in a series of exercises on the problem sheets. Only the strict inclusion is too difficult, here we refer to the literature.

(4.25) Example

By Lemma (4.16 b) and the Itô formula, we have

$$\begin{aligned} [\mathrm{Sym}(\mathbb{B}_{s,t}^{\mathrm{Itô}})]^{i,j} &= \frac{1}{2}\Big(\int_s^t B_{s,r}^{(i)} \mathrm{d}B_r^{(j)} + \int_s^t B_{s,r}^{(j)} \mathrm{d}B_r^{(i)}\Big) \\ &= \frac{1}{2}\big(B_{s,t}^{(i)} B_{s,t}^{(j)} - \delta_{i,j}|t-s|\big) = \frac{1}{2}(B_{s,t} \otimes B_{s,t} - I|t-s|)^{i,j}. \end{aligned}$$

Therefore, $\boldsymbol{B}^{\mathrm{Itô}} \notin \mathcal{C}_{\mathrm{g}}^{\alpha}$.

(4.26) Definition

The geometric rough path

$$\boldsymbol{B}(\omega) \equiv \boldsymbol{B}^{\mathrm{Strat}}(\omega) := \big(B(\omega), \mathbb{B}^{\mathrm{Itô}}(\omega) + (\tfrac{1}{2}(t-s)I)_{s,t}\big)$$

is called **Stratonovich-Brownian motion**, or simply **Brownian rough path**.

(4.27) Remark

The calculation in Example (4.25) shows that Stratonovich-BM $\boldsymbol{B}$ is indeed a geometric rough path. Therefore, by Remark (4.24 b), it can be approximated (in α-rough path distance for any $\alpha < 1/2$) by sequences of the form $(B^{(n)}, (\int_s^t B^{(n)}_{s,r} \otimes \mathrm{d}B^{(n)}_r)$, where $B^{(n)}$ is a mollification of B. In the remark, this mollification is chosen to be in C^1; in practice, a different and very concrete sequence of $B^{(n)}$ is quite useful, and it can be checked by hand that this converges to Stratonovich BM. We do this next.

(4.28) Definition

Let P_n be the partition of $[0,T]$ generated by the dyadic rationals with nearest neighbour distance 2^{-n}. For $f \in C([0,T],V)$, the **dyadic piecewise linear approximation** of f at level n is the function $f^{(n)}$ with

$$f^{(n)}_t := \sum_{[u,v)\in P_n\cap[0,t)} \frac{1}{|v-u|}\big((v-t)f_u + (t-u)f_v\big)\mathbb{1}_{[u,v)}(t).$$

The following proposition serves as a preparation for Theorem (4.31) below, but is quite interesting on its own.

(4.29) Proposition

Let B be a d-dimensional Brownian motion, D_n be the set of dyadic rationals with spacing 2^{-n}, and $\mathcal{F}^{(n)} = \sigma(B_t : t \in D_n)$. Then for all $t \geqslant 0$, we have

$$B^{(n)}_t = \mathbb{E}(B_t \,|\, \mathcal{F}^{(n)}) \qquad \text{almost surely.}$$

Moreover, for $i, j \leqslant d$ with $i \neq j$ and all $t \geqslant 0$, we have

$$\int_0^t (B^i_s)^{(n)}\,\mathrm{d}(B^j_s)^{(n)} = \mathbb{E}\Big(\int_0^t B^i_s\,\mathrm{d}B^j_s\,\Big|\mathcal{F}^{(n)}\Big) \qquad \text{almost surely.}$$

Proof: To see the first claimed equality, first note that $B^{(n)}_t$ is $\mathcal{F}^{(n)}$-measurable. Let $[u,v) \in P_n$ with $u \leqslant t < v$. Then with

$$Y := B_t - B^{(n)}_t = B_t - \frac{v-t}{v-u}B_u - \frac{t-u}{v-u}B_v,$$

what remains to be checked is the orthogonality condition $\mathbb{E}(YZ) = 0$ for all $\mathcal{F}^{(n)}$-measurable bounded Z. By a monotone class argument, we can restrict our attention to Z of the form $Z = \prod_{w\in D_n\cap[0,T)} f_w(B_w)$ for bounded functions f_w and fixed T. We can also assume $t > u$ since otherwise $Y = 0$ and the claim is trivial. Finally, by treating each component separately, we need only investigate one-dimensional Brownian motion. By the explicit formula (1.23) for the fidis of Brownian motion, we then find bounded functions F and G with

$$\mathbb{E}(YZ) = \int_{\mathbb{R}} \mathrm{d}x F(x) \int_{\mathbb{R}} \mathrm{d}y G(y) \int_{\mathbb{R}} \mathrm{d}z\, \underbrace{\frac{\mathrm{e}^{-\frac{(x-z)^2}{2(t-u)}}}{\sqrt{2\pi(t-u)}}\frac{\mathrm{e}^{-\frac{(y-z)^2}{2(v-t)}}}{\sqrt{2\pi(v-t)}}}_{=:\Phi(z)}\Big(z - \frac{v-t}{v-u}x - \frac{t-u}{v-u}y\Big).$$

The functions F and G appear by integrating out all variables in the formula from (1.23) except the ones corresponding to B_u, B_t and B_v. We thus need to show that the $\mathrm{d}z$-integral in the above expression is equal to zero for all choices of u, v, t, x and y. This can be checked directly (e.g. by completing the square), but this is extremely painful. A little trick helps: we have

$$\partial_z \Phi(z) = \Big(\frac{x-z}{t-u} + \frac{y-z}{v-t}\Big)\Phi(z) = \Big(\frac{x}{t-u} + \frac{y}{v-t} - z\frac{v-u}{(t-u)(v-t)}\Big)\Phi(z) =$$
$$= \frac{v-u}{(t-u)(v-t)}\Big(\frac{v-t}{v-u}x + \frac{t-u}{v-u}y - z\Big)\Phi(z).$$

The right hand side is (up to a z-independent factor) our integrand of interest, and the integral over $\int \mathrm{d}z \partial_z \Phi(z)$ is equal to zero by the fundamental theorem of calculus. This shows the claim.

For the second equality, first note that for arbitrary functions f, g, $f^{(n)}$ and $g^{(n)}$ are Lipschitz (and piecewise differentiable), and thus the Young integral $\int_0^t f_s^{(n)} \otimes \mathrm{d}g_s^{(n)}$ exists and is given by

$$\int_0^t f_s^{(n)} \otimes \mathrm{d}g_s^{(n)} = \sum_{[u,v)\in P_n \cap [0,t)} \int_u^v \Big(\frac{(v-t)f_u + (t-u)f_v}{v-u}\Big) \otimes \frac{g_v - g_u}{v-u}\,\mathrm{d}t =$$
$$= \sum_{[u,v)\in P_n \cap [0,t)} \frac{1}{2}(f_v + f_u) \otimes (g_v - g_u).$$

Note that this means that

$$\int_0^t B_s^{(n)} \otimes \mathrm{d}B_s^{(n)} = \sum_{[u,v)\in P_n \cap [0,t)} \frac{1}{2}(B_v + B_u) \otimes B_{u,v}$$

is precisely the Stratonovich approximation to the stochastic integral. We also have

$$\mathbb{E}\Big(\int_0^t (B_s^i)^{(n)} \mathrm{d}(B_s^j)^{(n)} \,\Big|\, \mathcal{F}^{(n-1)}\Big) = \mathbb{E}\Big(\sum_{[u,v)\in P_n \cap [0,t)} \frac{1}{2}(B_v^i + B_u^i)(B_v^j - B_u^j) \,\Big|\, \mathcal{F}^{(n-1)}\Big)$$
$$= \mathbb{E}\Big(\sum_{[u,v)\in P_n \cap [0,t)} \frac{1}{2}((B_v^{(n)})^i + (B_u^{(n)})^i)((B_v^{(n)})^j - (B_u^{(n)})^j) \,\Big|\, \mathcal{F}^{(n-1)}\Big)$$
$$= \sum_{[u,v)\in P_n \cap [0,t)} \frac{1}{2}((B_v^{(n-1)})^i + (B_u^{(n-1)})^i)((B_v^{(n-1)})^j - (B_u^{(n-1)})^j) = (*).$$

The final equality holds because we already know that $((B_t^{(n)})^i)_n$ is a martingale for all t, and because of the following fact: if (M_n) is an $(\mathcal{F}_n)$-martingale and N_n is a $(\mathcal{G}_n)$-martingale, and if $\mathcal{F}_n$ and $\mathcal{G}_n$ are independent for all n, then $(M_n N_n)$ is a $(\mathcal{F}_n \otimes \mathcal{G}_n)$-martingale (check that this is precisely the situation we have above!). To see this, check that that $\mathbb{E}(M_n N_n \mathbb{1}_A \mathbb{1}_B) = \mathbb{E}(M_k N_k \mathbb{1}_A \mathbb{1}_B)$ for all $A \in \mathcal{F}_k$ and $B \in \mathcal{G}_k$. By the usual approximation argument this then holds with $\mathbb{1}_C$ for $C \in \mathcal{F}_k \otimes \mathcal{G}_k$ instead of $\mathbb{1}_A \mathbb{1}_B$. This shows $\mathbb{E}(M_n N_n \,|\, \mathcal{F}_k \otimes \mathcal{G}_k) = M_k N_k$, hence the claim.

Now again for general functions f, g, for $[u, w) \in D_{n-1}$ and for (the unique) $v \in (u, w) \cap D_n$, we have

$$f_v^{(n-1)} = \frac{1}{w-u}((w-v)f_u + (v-u)f_w),$$

and therefore

$$g_v^{(n-1)} - g_u^{(n-1)} = \frac{v-u}{w-u} g_{u,w}, \qquad g_w^{(n-1)} - g_v^{(n-1)} = \frac{w-v}{w-u} g_{u,w},$$

and

$$f_u^{(n-1)} + f_v^{(n-1)} = \big(1 + \frac{w-v}{w-u}\big) f_u + \frac{v-u}{w-u} f_w, \qquad f_v^{(n-1)} + f_w^{(n-1)} = \frac{w-v}{w-u} f_u + \big(1 + \frac{v-u}{w-u}\big) f_w.$$

Now an elementary but tedious calculation gives

$$(f_u^{(n-1)} + f_v^{(n-1)}) \otimes (g_v^{(n-1)} - g_u^{(n-1)}) + (f_v^{(n-1)} + f_w^{(n-1)}) \otimes (g_w^{(n-1)} - g_v^{(n-1)}) = (f_w + f_u) \otimes (g_w - g_u).$$

Applying this to $(*)$, we see that the contribution of each pair of intervals $[u, v)$ and $[v, w)$ with $u, v \in P_{n-1}$ and $v \in P_n$ combines to give precisely the term that belongs to the interval $[u, w)$ in the formula for $\int_0^t (B_s^i)^{(n-1)} \mathrm{d}(B_s^j)^{(n-1)}$. We thus have

$$\mathbb{E}\Big(\int_0^t (B_s^i)^{(n)} \mathrm{d}(B_s^j)^{(n)} \,\Big|\, \mathcal{F}^{(n-1)} \Big) = \int_0^t (B_s^i)^{(n-1)} \mathrm{d}(B_s^j)^{(n-1)}.$$

Iterating this using the tower property, we find that for all n and all $m \geqslant n$,

$$\mathbb{E}\Big(\int_0^t (B_s^i)^{(m)} \mathrm{d}(B_s^j)^{(m)} \,\Big|\, \mathcal{F}^{(n-1)} \Big) = \int_0^t (B_s^i)^{(n-1)} \mathrm{d}(B_s^j)^{(n-1)}.$$

The random variable $\int_0^t (B_s^i)^{(m)} \mathrm{d}(B_s^j)^{(m)}$ converges to the Stratonovich integral $\int_0^t B_s^i \mathrm{d} B_s^j$ in L^2, which coincides with the Itô integral since we are only looking at off-diagonal terms. Since conditional expectation is a projection, and thus continuous in L^2, the left hand side converges to $\mathbb{E}(\int_0^t B_s^i \mathrm{d} B_s^j \,|\, \mathcal{F}^{(n-1)})$. This shows the claim. □

Remark: Note that we have not claimed the second identity for the diagonal terms $i = j$. It is wrong there, because for a one-dimensional Brownian motion, $\int_0^t B_s^{(n)} \mathrm{d} B_s^{(n)} = \frac{1}{2} (B_t^{(n)})^2$ by the telescopic sum. Since $(B_t^{(n)})_{n \in \mathbb{N}}$ is a non-trivial martingale, $((B_t^{(n)})^2)_{n \in \mathbb{N}}$ can not be a martingale. But $(\mathbb{E}(X \,|\, \mathcal{F}^{(n)}))_{n \in \mathbb{N}}$ is always a martingale, so $\int_0^t B_s^{(n)} \mathrm{d} B_s^{(n)}$ cannot be of that form.

Another ingredient for the theorem below is

(4.30) Lemma

For rough paths $(\boldsymbol{X}^{(n)})_{n \in \mathbb{N}}$ and $\boldsymbol{X}$ assume that $X_t^{(n)} \to X_t$ and $\mathbb{X}_{0,t} \to \mathbb{X}_{0,t}$ for all t. Assume further that for some $\beta < 1$,

$$\sup_{n \in \mathbb{N}} \|X^{(n)}\|_\beta < \infty \qquad \text{and} \qquad \sup_{n \in \mathbb{N}} \|\mathbb{X}^{(n)}\|_{2\beta} < \infty \qquad (**).$$

Then $\boldsymbol{X} \in \mathcal{C}^\beta$, and $\rho_\alpha(\boldsymbol{X}^{(n)}, \boldsymbol{X}) \to 0$ for all $\alpha < \beta$.

Proof: will be given in the exercises.

(4.31) Theorem

Let B be a d-dimensional Brownian motion. Then for all $\alpha < 1/2$, we have

$$\mathbb{P}\Big(\varrho_\alpha\Big(\big(B^{(n)}, \int_0^{\cdot} B^{(n)}_s \otimes \mathrm{d}B^{(n)}_s\big), \big(B, \mathbb{B}^{\mathrm{Strat}}\big)\Big) \overset{n\to\infty}{\longrightarrow} 0\Big) = 1.$$

Proof: Recall Theorem (4.68) from Probability Theory: for an integrable random variable X and a filtration $(\mathcal{F}_n)$, the sequence $(\mathbb{E}(X \,|\, \mathcal{F}_n))_{n\in\mathbb{N}}$ is a uniformly integrable martingale and converges almost surely to $\mathbb{E}(X \,|\, \mathcal{F}_\infty)$, with $\mathcal{F}_\infty = \sigma(\bigcup_{n\in\mathbb{N}} \mathcal{F}_n)$. Write $\mathbb{B}^{(n)}_t = \int_0^t B^{(n)}_s \otimes \mathrm{d}B^{(n)}_s$. Proposition (4.29) then shows that $(B^{(n)}_t)_{n\in\mathbb{N}}$ and $((\mathbb{B}^{(n)}_t)^{i,j})_{n\in\mathbb{N}}$ (for $i \neq j$) are uniformly integrable martingales and converge almost surely to B_t and $\mathbb{B}^{i,j}_t := \int_0^t B^i_s \mathrm{d}B^j_s$, respectively. Since $(\mathbb{B}^{(n)}_t)^{i,i} = \frac{1}{2}((B^{(n)}_t)^i)^2$, it converges almost surely as well. Note that the limit is not the Itô integral $\frac{1}{2}(B^i_t)^2 + \frac{1}{2}t$, however, but the Stratonovich integral $\frac{1}{2}(B^i_t)^2$. By the usual argument of continuous paths, we can then find a set of measure one such that the convergence holds simultaneously on all t for each ω from that set.

For arbitrary $\beta < 1/2$. we will now find a (possibly different) set of measure one such that on that set, $\sup_{n\in\mathbb{N}} \|B^{(n)}\|_\beta < \infty$ and $\sup_{n\in\mathbb{N}} \|\mathbb{B}^{(n)}\|_{2\beta} < \infty$. Then on the intersection of the two sets of measure one, Lemma (4.30) will show $\rho_\alpha(\boldsymbol{B}^{(n)}, \boldsymbol{B}) \to 0$ for all $\alpha < \beta$ (and thus for all $\alpha < 1/2$), and therefore show the claim.

We will show the claims separately for each component of $(B^{(n)})$ and of $(\mathbb{B}^{(n)})$. We will start with $(\mathbb{B}^{(n)})^{i,j}$ and use ideas from the proof of the Kolmogorov-Chentsov theorem. Let $q = 4k$ with $k \in \mathbb{N}$. Since $((\mathbb{B}^{(n)}_t)^{i,j})_{n\in\mathbb{N}}$ is a martingale, the process $((\mathbb{B}^{(n)}_{u,v})^{i,j})_{n\in\mathbb{N}}$ is also a martingale, and the process $\big(\frac{1}{|v-u|^{2\beta}}\big|(\mathbb{B}^{(n)}_{u,v})^{i,j})\big|^{q/2}\big)_{n\in\mathbb{N}}$ is a submartingale as a convex function of a martingale. Since suprema of several submartingales are also submartingales, the process $(S_n)_{n\in\mathbb{N}}$ with

$$S_n := \sup_{u,v\in D, u\neq v} \Big(\frac{1}{|v-u|^{2\beta}}\big|(\mathbb{B}^{(n)}_{u,v})^{i,j})\big|\Big)^{q/2}$$

is a submartingale. Then Doobs maximal inequality gives for all $N \in \mathbb{N}$

$$\mathbb{P}\Big(\sup_{n\leqslant N} \sup_{u,v\in D, u\neq v} \frac{1}{|v-u|^{2\beta}}\big|(\mathbb{B}^{(n)}_{u,v})^{i,j})\big| > C\Big) = \mathbb{P}(\sup_{n\leqslant N} S_n > C^{q/2}) \leqslant \frac{1}{C^{q/2}}\mathbb{E}(|S_N|). \qquad (*)$$

For each N and all u, v, Proposition (4.29) and the conditional Jensen inequality yield

$$\Big|\frac{1}{|v-u|^{2\beta}}(\mathbb{B}^{(N)}_{u,v})^{i,j}\Big|^{q/2} \leqslant \mathbb{E}\Big(\Big|\frac{1}{|v-u|^{2\beta}}\mathbb{B}^{i,j}_{u,v}\Big|^{q/2} \,\Big|\, \mathcal{F}^{(N)}\Big) \leqslant \mathbb{E}\Big(\sup_{u\neq v}\Big|\frac{1}{|v-u|^{2\beta}}\mathbb{B}^{i,j}_{u,v}\Big|^{q/2} \,\Big|\, \mathcal{F}^{(N)}\Big),$$

which in turn gives

$$\mathbb{E}(|S_N|) \leqslant \mathbb{E}\Big(\sup_{u\neq v}\mathbb{E}\Big(\Big|\sup_{u\neq v}\frac{1}{|v-u|^{2\beta}}\mathbb{B}^{i,j}_{u,v}\Big|^{q/2} \,\Big|\, \mathcal{F}^{(N)}\Big)\Big) = \mathbb{E}\Big(\Big|\sup_{u\neq v}\frac{1}{|v-u|^{2\beta}}|\mathbb{B}^{i,j}_{u,v}\Big|^{q/2}\Big).$$

We know from Theorem (4.18) and Proposition (4.19) that the last quantity is finite for all $\beta < 1/2$. We can thus take the limit $N \to \infty$ in $(*)$, use continuity from below of measures on the left hand side, and obtain

$$\mathbb{P}(\sup_{n\in\mathbb{N}} \|(\mathbb{B}^{(n)})^{i,j}\|_{2\beta} > C) \leqslant \frac{K}{C^{q/2}}$$

for some constant K. This shows that $\sup_{n\in\mathbb{N}} \|(\mathbb{B}^{(n)})^{i,j}\|_{2\beta}$ is finite almost surely.

The same argument works for the components of $\sup_{n\in\mathbb{N}} \|(B^{(n)})^{i}\|_{\beta}$, using Theorem (1.36) this time. The details are left as an exercise. Once this is shown, the remaining claim for $\sup_{n\in\mathbb{N}} \|(\mathbb{B}^{(n)})^{i,i}\|_{2\beta}$ follows immediately because $(\mathbb{B}^{(n)}_{s,t})^{i,i} = (B^i_{s,t})^2$. □

Our next aim is to integrate more general functions than $s \mapsto f(X_s)$ against a rough path $\mathbb{X}$. This will be necessary e.g. for treating rough differential equations: they have the form

$$Y_t = \int_0^t F(Y_s)\,\mathrm{d}\boldsymbol{X}_s$$

so Y_s does not only depend on the value X_s but also on all X_r with $r \leqslant s$. The idea is to replace the objects $Y_s = F(X_s)$, $Y'_s = DF(X_s)$ and $R^Y_{s,t} = Y_{s,t} - Y'_s X_{s,t}$ by more general expressions so that the proof of Lemma (4.11) still works, and see where this gets us.

(4.32) Definition

For $X \in C^\alpha([0,T],V)$, we say that $Y \in C^\alpha([0,T],\mathcal{L}(V,W))$ is **controlled by** X if there exists $Y' \in C^\alpha([0,T],\mathcal{L}(V,\mathcal{L}(V,W))$ such that with

$$R^Y_{s,t} = Y_{s,t} - Y'_s X_{s,t} \in \mathcal{L}(V,W),$$

R^Y is an element of $C_2^{2\alpha}([0,T],\mathcal{L}(V,W))$. Any Y' with this property is called **Gubinelli derivative** of Y with respect to X.

(4.33) Remark

a) For $X \in C^\alpha$ and $f \in C_b^2([0,T],\mathcal{L}(V,W))$, Lemma (4.9) shows that with $Y'_s := Df(X_s)$, Y' is a Gubinelli derivative of X.

b) Set $Y_s = X_s$ for all s, then Y is controlled by X. More precisely, set $W = \mathbb{R}$ and then identify V with $\mathcal{L}(V,\mathbb{R})$ via the scalar product, then $Y \in \mathcal{L}(V,W)$ as required. With $Y'_s = \mathrm{id}_V \in \mathcal{L}(V,\mathcal{L}(V,\mathbb{R})) \simeq \mathcal{L}(V,V)$, we then find $R^Y_{s,t} = X_s - \mathrm{id}_V X_s = 0$, which shows the claim. In other words, X is controlled by itself.

c) In more generality, we can define when some $Y \in C^\alpha([0,T],U)$ for an arbitrary Hilbert space U is controlled by $C \in C^\alpha([0,T],V)$ by just replacing $\mathcal{L}(V,W)$ with U everywhere. Since we will not need this for rough integration and since there are already too many vector spaces floating around, we do not do it.

d) The requirement $R^Y \in C_2^{2\alpha}([0,T],\mathcal{L}(V,W))$ means that Y'_s is the *correct factor* that is needed at point s so that $Y'_s X_{s,t}$ cancels some of the irregularity of $Y_{s,t}$ for t very close to s. In the classical case $Y' = Df(X)$, this cancellation is provided by the Taylor expansion, in general we just have to assume it axiomatically.

e) Gubinelli derivatives are possibly not unique. If X and Y are already in $C^{2\alpha} \subset C^\alpha$, then any smooth Y' will do the job. On the other hand, if

$$\limsup_{t\to s} \frac{1}{|t-s|^{2\alpha}} |\phi(X_{s,t})| = \infty \qquad \text{for all } \phi \in \mathcal{L}(X,\mathbb{R}),$$

then one can show that Y' is uniquely determined by Y for all $Y \in C^\alpha$, even smooth ones. Such functions X are sometimes called **truly α-rough**. One can also show that Brownian motion is truly α-rough for any $\alpha > 1/4$. We will not discuss this topic further, see the book of Friz and Hairer, Sections 6.2 and 6.3. Also, a possible non-uniqueness of Y has no consequences for anything we will be doing.

(4.34) Definition

For $X \in C^\alpha([0,T],V)$ define

$$\mathcal{Y}_X := \{Y \in C^\alpha([0,T],\mathcal{L}(V,W)) : Y \text{ is controlled by } X\}$$

We will choose a specific Gubinelli derivative Y' for each $Y \in \mathcal{Y}_X$, using the axiom of choice if X is not truly rough, and with this choice define

$$\mathcal{D}_X^{2\alpha} := \{(Y,Y') : Y \in \mathcal{Y}_X\}, \qquad Y_X := (Y,Y'),$$

and

$$\|Y_X\|_{X,2\alpha} := \|Y'\|_\alpha + \|R^Y\|_{2\alpha,\text{diag}}.$$

(4.35) Proposition

a) $\|.\|_{X,2\alpha}$ is a seminorm. The map $Y_X \mapsto |Y_0| + |Y_0'| + \|Y_X\|_{X,2\alpha} =: \|Y_X\|_{X,2\alpha,0}$ is a norm. $\mathcal{D}_X^{2\alpha}$ is complete under this norm.

b) There exists $C < \infty$ with

$$\|Y\|_\alpha \leqslant C(1 + \|X\|_\alpha)\big(|Y_0'| + T^\alpha \|Y_X\|_{X,2\alpha}\big) \qquad \text{for all } Y \in \mathcal{D}_X^{2\alpha}.$$

In particular, this means that the inclusion of the term $\|Y\|_\alpha$ into the seminorm $\|Y_X\|_{X,2\alpha}$ is not necessary, since the norms $\|\cdot\|_{X,2\alpha,0}$ would be equvalent with or without this inculsion.

Proof: Exercise.

(4.36) Theorem

Let $T > 0$, $\alpha \in (1/3, 1/2]$, $\boldsymbol{X} = (X, \mathbb{X}) \in \mathcal{C}^\alpha([0,t],V)$, $Y_X = (Y,Y') \in \mathcal{D}_X^{2\alpha}([0,T],\mathcal{L}(V,W))$. Then:

a) The rough integral

$$\int_s^t Y_r \,\mathrm{d}\boldsymbol{X}_r := \lim_{|P|\to 0} \sum_{[u,v)\in P} (Y_u X_{u,v} + Y_u' \mathbb{X}_{u,v})$$

exists for all $0 < s < t < T$ and is independent of the approximation partitions (it does depend on the choice of Y' if there is one). We have

$$\Big| \int_s^t Y_r \,\mathrm{d}\boldsymbol{X}_r - (Y_s X_{s,t} + Y_s' \mathbb{X}_{s,t}) \Big| \leqslant 2^{3\alpha} \zeta(3\alpha) \big(\|X\|_\alpha \|R^Y\|_{2\alpha,\text{diag}} + \|Y'\|_\alpha \|\mathbb{X}\|_{2\alpha,\text{diag}}\big) |t-s|^{3\alpha}.$$

b) Set $Z_t = \int_0^t Y_r \,\mathrm{d}\boldsymbol{X}_r$. The map Z is an element of C^α, and

$$\|Z\|_\alpha \leqslant 2 \max\{2^{3\alpha} T^{2\alpha} \zeta(3\alpha), 1\} \big(\|X\|_\alpha \|R^Y\|_{2\alpha,\text{diag}} + \|Y'\|_\alpha \|\mathbb{X}\|_{2\alpha,\text{diag}}\big).$$

c) **Fundamental Theorem of Calculus for Rough Paths:** The map $t \mapsto \int_0^t Y_r \,\mathrm{d}\boldsymbol{X}_r$ is an element of $\mathcal{Y}_X$, and Y is a Gubinelli derivative for it. We write

$$Z_X := (Z, Z') := \Big(\int_0^{\cdot} Y_s \,\mathrm{d}\boldsymbol{X}_s, Y \Big) \in \mathcal{D}_X^{2\alpha}([0,T], W).$$

We have

$$\|Z_X\|_{X,2\alpha,0} \leqslant |Y_0| + \|Y\|_\alpha + \|Y'\|_\infty \|\mathbb{X}\|_{2\alpha} + \tilde{C}(\alpha)(T^\alpha \vee 1)\big(\|X\|_\alpha \|R^Y\|_{2\alpha,\mathrm{diag}} + \|Y'\|_\alpha \|\mathbb{X}\|_{2\alpha,\mathrm{diag}}\big),$$

with $\tilde{C}(\alpha) = 3\zeta(3\alpha)2^{3\alpha}\max\{T^{2\alpha}, 1\}$. In particular, the linear map

$$\mathcal{D}_X^{2\alpha}([0,T], \mathcal{L}(V,W)) \to \mathcal{D}_X^{2\alpha}([0,T], W), \qquad (Y, Y') \mapsto (Z, Z')$$

is continuous with respect to the norm $\|.\|_{X,2\alpha,0}$.

Proof: With $\Xi_{s,t} = Y_s X_{s,t} + Y'_s \mathbb{X}_{s,t}$, we find exactly as in (4.9 d) (or (4.2)) that $\Xi_{u,w} - \Xi_{u,v} - \Xi_{v,w} = -R^Y_{u,v} X_{v,w} - Y'_{u,v}\mathbb{X}_{v,w}$. Thus $\|\delta\Xi\|_{3\alpha} \leqslant \|R^Y\|_{2\alpha,\mathrm{diag}}\|X\|_\alpha + \|Y'\|_\alpha \|\mathbb{X}\|_{2\alpha,\mathrm{diag}}$, which is finite by assumption. Thus the Sewing Lemma (4.11) applies and its part a) gives claims a), and its part b) gives claim b).

For c), we prove the claimed inequality, all other statements follow from it. By definition of $\|.\|_{X,2\alpha}$, we have $\|Z_X\|_{X,2\alpha,0} = |Z_0| + |Z'_0| + \|Z\|_\alpha + \|R^Z\|_{2\alpha}$. We investigate the terms separately. We have $Z' = Y$, and

$$R^Z_{s,t} = Z_{s,t} - Z'_s X_{s,t} = \int_s^t Y_r \,\mathrm{d}\boldsymbol{X}_r - Y_s X_{s,t} = \int_s^t Y_r \,\mathrm{d}\boldsymbol{X}_r - \Xi_{s,t} + Y'_s \mathbb{X}_{s,t},$$

and thus part a) gives

$$\|R^Z\|_{2\alpha} \leqslant 2^{3\alpha}\zeta(3\alpha)\max\{1, T^\alpha\}\big(\|X\|_\alpha \|R^Y\|_{2\alpha,\mathrm{diag}} + \|Y'\|_\alpha \|\mathbb{X}\|_{2\alpha,\mathrm{diag}}\big) + \|Y'\|_\infty \|\mathbb{X}\|_{2\alpha,\mathrm{diag}}.$$

By part b) we have

$$\|Z\|_\alpha \leqslant 2\zeta(3\alpha)2^{3\alpha}\max\{T^{2\alpha}, 1\}\big(\|X\|_\alpha \|R^Y\|_{2\alpha,\mathrm{diag}} + \|Y'\|_\alpha \|\mathbb{X}\|_{2\alpha,\mathrm{diag}}\big).$$

Together with $Z_0 = 0$ and $Z'_0 = Y_0$, we obtain the desired inequality. □

(4.37) Definition

For $X, \tilde{X} \in C^\alpha$, $Y_X \in \mathcal{D}_X^{2\alpha}$ and $\tilde{Y}_{\tilde{X}} \in \mathcal{D}_{\tilde{X}}^{2\alpha}$, the controlled rough path distance is defined by

$$d_{X,\tilde{X},2\alpha}(Y_X, \tilde{Y}_{\tilde{X}}) := \|Y' - \tilde{Y}'\|_\alpha + \|R^Y - R^{\tilde{Y}}\|_{2\alpha,\mathrm{diag}}.$$

Warning: it is possible that $\mathcal{D}_X^{2\alpha} \cap \mathcal{D}_{\tilde{X}}^{2\alpha} = \{0\} \in C^\alpha([0,T], V) \times C^\alpha([0,T], \mathcal{L}(V,W))$. The reason is that we need R^Y and $R^{\tilde{Y}}$ to be 2α-Hölder continuous on the diagonal, which means that the small scale structures of $Y'_s X_{s,t}$ and $Y_{s,t}$ have to cancel some of each other's roughness. If the two paths X and $\tilde{X}$ are too different, this will not be possible. Therefore, it is again not a good idea to try and define $d_{X,\tilde{X},2\alpha}$ just as the norm on the intersection of the two spaces.
Also, note that again, $d_{X,\tilde{X},2\alpha}$ is only a semi-distance.

Remark: As is the case for $\|\cdot\|_{X,2\alpha}$, the inclusion of a term $\|Y - \tilde{Y}\|_\alpha$ into the semidistance $d_{X,\tilde{X},2\alpha}$ is not necessary, as the next lemma shows.

(4.38) Lemma

In the situation of (4.37), we have

$$\|Y-\tilde{Y}\|_\alpha \leqslant \|Y'\|_\infty \|X-\tilde{X}\|_\alpha + \|\tilde{X}\|_\alpha |\tilde{Y}'_0 - Y'_0| + T^\alpha (1+\|X\|_\alpha) d_{X,\tilde{X},2\alpha}(Y_X, \tilde{Y}_{\tilde{X}}).$$

Proof: exercise.

Next, we link the general rough integral back to general classical Itô and Stratonovich integrals.

(4.39) Proposition

Let B be a d-dimensional Brownian motion, $(\mathcal{F}_t)$ its filtration, and let $\boldsymbol{B} \equiv \boldsymbol{B}^{\text{Itô}}$ be the Itô rough path. Let (Y, Y') be a pair of stochastic processes on the same probability space Ω as B, and assume that $(Y(\omega), Y'(\omega)) \in \mathcal{D}^{2\alpha}_{B(\omega)}$ for all $\omega \in \Omega$, so that the rough integral $\int_0^t Y_s(\omega)\, \mathrm{d}\boldsymbol{X}_s(\omega)$ exists for all ω and all t.

a) If we assume that Y is $(\mathcal{F}_t)$-adapted, then the Itô integral $\int_0^t Y_s \mathrm{d}B_s$ exists almost surely.

b) If we assume that both Y and Y' are $(\mathcal{F}_t)$-adapted, then almost surely and simultaneously for all t,

$$\int_0^t Y_s \,\mathrm{d}\boldsymbol{B}_s = \int_0^t Y_s \,\mathrm{d}B_s.$$

Proof: a) Since $Y \in C^\alpha$ and is adapted by assumption, Theorem (3.32) gives the existence of the generalized Itô-integral.

b) The same theorem also gives the convergence of the approximating expressions $\sum_{[u,v)\in P\cap[0,t)} Y_u B_{u,v}$ in probability, and so the convergence is almost sure along a subsequence of partitions. On the other hand, we have

$$\lim_{|P|\to 0} \sum_{[u,v)\in P\cap[0,t)} \big(Y_u(\omega) B_{u,v}(\omega) + Y'_u(\omega)\mathbb{B}_{u,v}(\omega)\big) = \int_0^t Y_s(\omega)\,\mathrm{d}\boldsymbol{B}_s(\omega)$$

for all $\omega \in \Omega$ and all t. On the subsequence of partitions and the set of measure one where the Itô-integral is the pointwise limit of its approximations, we thus find

$$\Delta_t(\omega) := \Big(\int_0^t Y_s\,\mathrm{d}B_s\Big)(\omega) - \int_0^t Y_s(\omega)\,\mathrm{d}\boldsymbol{B}_s(\omega) = \lim_{|P|\to 0} \sum_{[u,v)\in P\cap[0,t)} Y'_u(\omega)\mathbb{B}_{u,v}(\omega) =: \lim_{|P|\to 0} \Delta_t^P(\omega),$$

where the limit exists as the difference of two existing limits. We want to show that it is equal to zero.

For this, assume first that there exists $K < \infty$ such that $|Y'(\omega)| \leqslant K$ all ω, and fix a partition P generated by the points $0 = t_0 < t_1 < \ldots < t_n = t$. Then for $i < n$, we have

$$\mathbb{E}(\Delta^P_{t_{i+1}} \,|\, \mathcal{F}_{t_i}) = \sum_{[u,v)\in P\cap[0,t_i)} Y'_u \mathbb{B}_{u,v} + Y'_{t_i} \underbrace{\mathbb{E}(\mathbb{B}_{t_i,t_{i+1}} \,|\, \mathcal{F}_{t_i})}_{=0}$$

almost surely. Here we used that Y' is adapted by assumption, and that $\mathbb{B}_{u,v} = \int_u^v B_{u,r} \otimes \mathrm{d}B_r$ almost surely, and thus is $\mathcal{F}_v$-measurable. The equality above implies that $(\Delta^P_{t_i})_{i\leqslant n}$ is a martingale, which in turn implies

$$\mathbb{E}\big((\Delta_t^P)^2\big) = \mathbb{E}\Big(\big(\sum_{[u,v)\in P\cap[0,t)} Y'_u \mathbb{B}_{u,v}\big)^2\Big) = \sum_{[u,v)\in P\cap[0,t)} \mathbb{E}\Big(\big(Y'_u \mathbb{B}_{u,v}\big)^2\Big) \leqslant K^2 \sum_{[u,v)\in P\cap[0,t)} \mathbb{E}(\mathbb{B}_{u,v}^2).$$

By Brownian scaling (2.4), we have

$$\mathbb{E}(\mathbb{B}_{u,v}^2) = \mathbb{E}\Big(\Big(\int_0^{v-u} B_s \otimes \mathrm{d}B_s\Big)^2\Big) = \mathbb{E}\Big(\Big((v-u)\int_0^1 B_s \otimes \mathrm{d}B_s\Big)^2\Big) = c(v-u)^2$$

for some $c > 0$, which implies that

$$\mathbb{E}\big((\Delta_t^P)^2\big) \leqslant cK^2 \sum_{[u,v)\in P} (v-u)^2 \leqslant cK^2 T|P| \overset{|P|\to 0}{\longrightarrow} 0.$$

This shows that Δ_t^P converges to zero almost surely along a subsequence, hence $\Delta_t(\omega) = 0$ almost surely. Since it is continuous as the difference of two continuous expressions, this holds simultaneously of all t, almost surely. This shows the claim for uniformly bounded Y'.

For general Y we use localization. We define $\sigma_n(\omega) = \inf\{t \geqslant 0 : |Y'_t(\omega)| \geqslant n\}$, and define $(Y'_t)^{(n)}(\omega) = Y'_t(\omega)\mathbb{1}_{[0,\sigma_n(\omega))}(t)$. Since we did not change (Y_t), the Itô-integral $\int_0^t Y_s \,\mathrm{d}B_s$ still exists as the limit along the same subsequence of partitions and on the same set of probability one as before. Also, the same calculation as above gives $\lim_{|P|\to 0} \Delta_t^{P,n}(\omega) = 0$ on a set of probability one, where $\Delta_t^{P,n}$ is defined like Δ_t^P with Y' replaced by $(Y')^{(n)}$. On the intersection of those two sets of probability one, the limit

$$J_n(\omega) := \lim_{|P|\to 0}\Big(\sum_{[u,v)\in P} Y_u(\omega) B_{u,v}(\omega) + \Delta_t^{P,n}(\omega)\Big)$$

exists and equals the Itô integral - note that the right hand side is not a rough integral as we defined it, since $(Y')^{(n)}$ is possibly discontinuous and thus $(Y, Y'^{(n)}) \notin \mathcal{D}_X^{2\alpha}$. More precisely, we have convergence for the particular sequence of partitions generated by choosing the subsequence to get from convergence in probability to almost sure convergence, but have no guarantee for other sequences of partitions. However, for all ω with $\sigma_n(\omega) > T$, we have $(Y'_t)^{(n)}(\omega) = Y'_t(\omega)$, and so $J_n(\omega) = \int_0^t Y_s(\omega)\,\mathrm{d}\boldsymbol{B}_s(\omega)$. So the equality holds on the set $\bigcup_{n\in\mathbb{N}}\{\sigma_n \geqslant T\}$, which is a set of probability one as $\|Y'(\omega)\|_\infty < \infty$ for all ω due to path continuity. □

(4.40) Stratonovich integrals

For $Y \in \mathcal{L}^2_{T,\mathrm{loc}}$, we say that the Stratonovich integral exists if the limit

$$\int_0^t Y_s \circ \mathrm{d}B_s := \int_0^t Y_s \,\mathrm{d}B_s + \frac{1}{2}\lim_{|P|\to 0} \sum_{[u,v)\in P} Y_{u,v} B_{u,v}$$

exists with respect to convergence in probability, where the first term on the right ist the generalized Itô integral. Similar arguments as above show that for adapted processes $(Y_s)_{s\leqslant T}$, $(Y'_s)_{s\leqslant T}$ with $(Y(\omega), Y'(\omega)) \in \mathcal{D}^{2\alpha}_{B(\omega)}$ for all ω, the Stratonovich integral exists and equals the rough integral with respect to Stratonovich Brownian motion:

$$\int_0^t Y_s \circ \mathrm{d}B_s = \int_0^t Y_s \,\mathrm{d}\boldsymbol{B}_s^{\mathrm{Strat}} \qquad \text{almost surely.}$$

Our next aim is an Itô-formula for rough integrals. For this, we introduce the following notion:

(4.41) Definition

Let $\boldsymbol{X} = (X, \mathbb{X}) \in \mathcal{C}^\alpha$. The map

$$[\boldsymbol{X}] : [0,T]^2 \to V \otimes V, \qquad (s,t) \mapsto [\boldsymbol{X}]_{s,t} := X_{s,t} \otimes X_{s,t} - 2\mathrm{Sym}\mathbb{X}_{s,t}$$

is called the **bracket** of $\boldsymbol{X}$.

(4.42) Lemma

a) For $u < v < w$, $[\boldsymbol{X}]_{u,w} = [\boldsymbol{X}]_{u,v} + [\boldsymbol{X}]_{v,w}$, i.e. the bracket is the increment of a function $r \mapsto [\boldsymbol{X}]_r = [\boldsymbol{X}]_{0,r}$. This function is an element of $C^{2\alpha}$.

b) For two rough patss $\boldsymbol{X} = (X, \mathbb{X})$ and $\tilde{\boldsymbol{X}} = (X, \tilde{\mathbb{X}})$ with the same first component, the map $(s,t) \mapsto \mathrm{Sym}(\mathbb{X}_{s,t} - \tilde{\mathbb{X}}_{s,t})$ is additive (i.e. the increment of a function). It is given by $\frac{1}{2}([\tilde{\boldsymbol{X}}]_{s,t} - [\boldsymbol{X}]_{s,t})$.

c) $\boldsymbol{X}$ is a geometric rough path if and only of $[\boldsymbol{X}] = 0$.

Proof: a) By Chen's relation

$$\begin{aligned}\mathrm{Sym}(\mathbb{X}_{u,v}) + \mathrm{Sym}(\mathbb{X}_{v,w}) &= \mathrm{Sym}(\mathbb{X}_{u,v} + \mathbb{X}_{v,w}) = \mathrm{Sym}(\mathbb{X}_{u,w} - X_{u,v} \otimes X_{v,w}) \\ &= \mathrm{Sym}(\mathbb{X}_{u,w}) - \frac{1}{2}(X_{u,v} \otimes X_{v,w} + X_{v,w} \otimes X_{u,v}).\end{aligned}$$

This gives

$$[\boldsymbol{X}]_{u,v} + [\boldsymbol{X}]_{v,w} = X_{u,v} \otimes X_{u,v} + X_{v,w} \otimes X_{v,w} + X_{u,v} \otimes X_{v,w} + X_{v,w} \otimes X_{u,v} - 2\mathrm{Sym}(\mathbb{X}_{u,w}) = [\boldsymbol{X}]_{u,w}.$$

The regularity of the bracket is clear from its definition.

b) is immediate from a), and c) is by definition of geometric rough paths. □

(4.43) Itô's formula for rough integrals

Let $\boldsymbol{X} \in \mathcal{C}^\alpha$, $f \in C_b^3([0,T], \mathcal{L}(V,W))$ and $\alpha > 1/3$. Then for all $s < t$,

$$f(X_t) - f(X_s) = \int_s^t Df(X_r)\, \mathrm{d}\boldsymbol{X}_r + \frac{1}{2} \int_s^t D^2 f(X_r) \mathrm{d}[\boldsymbol{X}]_r,$$

where the last expression is a Young integral.

Proof: As usual, we consider the telescopic sum

$$f(X_t) - f(X_s) = \sum_{[u,v) \in P \cap [s,t)} (f(X_v) - f(X_u)). \qquad (*)$$

Taylor expansion gives

$$\big| f(X_v) - f(X_u) - Df(X_u) X_{u,v} - \frac{1}{2} D^2 f(X_u)(X_{u,v} \otimes X_{u,v}) \big| \leqslant \|D^3 f\|_\infty |X_{u,v}|^3.$$

Since $|X_{u,v}|^3 \leqslant (v-u)^{3\alpha}$ vanishes faster than linearly, the correction term disappears when taking the limit $|P| \to 0$ in $(*)$.

On the other hand, the rough integral $\int_s^t Df(X_s)\,\mathrm{d}\boldsymbol{X}_s$ is the limit

$$\lim_{|P|\to 0} \sum_{[u,v)\in P\cap[s,t)} \big(Df(X_u)X_{u,v} + D^2 f(X_u)\mathbb{X}_{u,v}\big),$$

and

$$D^2 f(X_u)\mathbb{X}_{u,v} = D^2 f(X_u)\mathrm{Sym}(\mathbb{X}_{u,v}).$$

The last equality is true because $(D^2 f(X_u))^{i,j} = [\partial_i\partial_j f](X_u) = (D^2 f(X_u))^{j,i}$ is a symmetric matrix, while $\mathrm{Anti}(\mathbb{X}_{u,v}) = \mathbb{X}_{u,v} - \mathrm{Sym}(\mathbb{X}_{u,v})$ is an antisymmetric one. For a symmetric matrix A and antisymmetric matrix B, the contraction $\sum_{i,j=1}^n A_{i,j}B_{i,j} = 0$. Each component of the vector $D^2 f(X_u)\mathrm{Anti}(\mathbb{X}_{u,v})$ is just such a contraction.

Therefore,

$$f(X_t) - f(X_s) - \int_s^t Df(X_r)\,\mathrm{d}\boldsymbol{X}_r = \lim_{|P|\to 0} \sum_{[u,v)\in P\cap[s,t)} \frac{1}{2} D^2 f(X_u) \underbrace{\big(X_{u,v}\otimes X_{u,v} - 2\mathrm{Sym}\mathbb{X}_{u,v}\big)}_{=[\boldsymbol{X}]_{u,v}}.$$

Since $[\boldsymbol{X}_{u,v}]$ is the increment of a $C^{2\alpha}$-function and $s \mapsto D^2 f(X_s)$ is in C^α, the claim now follows from $\alpha > 1/3$ and Youngs theorem (3.4). □

(4.44) Examples

a) Since $\boldsymbol{B}^{\mathrm{Strat}}$ is geometric, we have $[\boldsymbol{B}^{\mathrm{Strat}}]_{s,t} = 0$ for all s,t. Therefore

$$f(B_t) - f(B_s) = \int_s^t Df(B_r)\,\mathrm{d}\boldsymbol{B}_r^{\mathrm{Strat}},$$

i.e. the classical fundamental theorem of calculus holds. By (formally!) differentiating this with respect to t, one obtains the „chain rule“ $\partial_t f(B_t) = Df(B_t)\dot{\boldsymbol{B}}_t$, where of course the time derivative $\dot{\boldsymbol{B}}$ of $\boldsymbol{B}$ does not really make sense.

b) Since $[\boldsymbol{B}^{\mathrm{Itô}}]_{s,t} = \frac{1}{2}(t-s)\mathrm{id}$, Itô's formula yields the classical Itô-Stratonovich-correction also for rough integrals.

Our next aim is to define rough differential equations and prove existence of their solutions. We prepare this by two technical estimates that tell us what happens when we plug controlled rough paths into a smooth function φ.

(4.45) Proposition

Let $X \in C^\alpha([0,T],V)$, $Y \in \mathcal{D}_X^{2\alpha}([0,T],W)$ and $\varphi \in C_b^2(W,\bar{W})$. Define $Z_t := \varphi(Y_t)$ and $Z'_t = ([D\varphi](Y_t))Y'_t \in \mathcal{L}(V,W')$. Then

$$Z_X := (Z,Z') \in \mathcal{D}_X^{2\alpha}([0,T],\bar{W}),$$

and

$$\|Z_X\|_{X,2\alpha} \leqslant \|D\varphi\|_\infty \|Y_X\|_{X,2\alpha} + \|D^2\varphi\|_\infty\big(\|Y\|_\alpha^2 + \|Y\|_\alpha \|Y'\|_\infty\big).$$

Proof: Since

$$R^Z_{s,t} = Z_{s,t} - Z'_s X_{s,t} = \varphi(Y_t) - \varphi(Y_s) - D\varphi(Y_s) \underbrace{Y'_s X_{s,t}}_{=Y_{s,t}+R^Y_{s,t}},$$

and

$$\varphi(Y_t) - \varphi(Y_s) - D\varphi(Y_s)Y_{s,t} = \int_0^1 \mathrm{d}r \int_0^r \mathrm{d}u D^2\varphi(Y_s + uY_{s,t})Y_{s,t} \otimes Y_{s,t},$$

we obtain

$$\|R^Z_{s,t}\|_{2\alpha,\mathrm{diag}} \leqslant \|D^2\varphi\|_\infty \|Y\|_\alpha^2 + \|D\varphi\|_\infty \|R^Y\|_{2\alpha,\mathrm{diag}}.$$

Since

$$Z'_{s,t} = \big(D\varphi(Y_t) - D\varphi(Y_s)\big)Y'_t + D\varphi(Y_s)(Y'_t - Y'_s) = \int_0^1 D^2\varphi(Y_s + rY_{s,t})Y_{s,t} \otimes Y'_t \,\mathrm{d}r + D\varphi(Y_s)(Y'_t - Y'_s),$$

we get

$$\|Z'\|_\alpha \leqslant \|D^2\varphi\|_\infty \|Y\|_\alpha \|Y'\|_\infty + \|D\varphi\|_\infty \|Y'\|_\alpha.$$

Combining the two estimates yields the claim. □

(4.46) Proposition

Let $\boldsymbol{X}, \tilde{\boldsymbol{X}} \in \mathcal{C}^\alpha([0,T],V)$, $Y_X \in \mathcal{D}^{2\alpha}_X([0,T],W)$, $\tilde{Y}_{\tilde{X}} \in \mathcal{D}^{2\alpha}_{\tilde{X}}([0,T],W)$, and $\varphi \in \mathcal{C}^3_b(W,\bar{W})$. Define Z_X and $\tilde{Z}_{\tilde{X}}$ as in (4.45). Then there exist constants $\boldsymbol{C}$ and $\boldsymbol{C}'$ that depend on the quantities $\|X\|_\alpha$, $\|\tilde{X}\|_\alpha$, $\|Y_X\|_{X,2\alpha}$, $|Y'_0|$, $\|\tilde{Y}_{\tilde{X}}\|_{\tilde{X},2\alpha}$ and $|\tilde{Y}'_0|$, such that

$$d_{X,\tilde{X},2\alpha}(Z_X, \tilde{Z}_{\tilde{X}}) \leqslant \boldsymbol{C}\|\varphi\|_{C^3_b}\big(T^{3\alpha} \vee 1\big)\Big(\rho_\alpha(\boldsymbol{X}, \tilde{\boldsymbol{X}}) + |Y_0 - \tilde{Y}_0| + |Y'_0 - \tilde{Y}'_0| + d_{X,\tilde{X},2\alpha}(Y_X, \tilde{Y}_{\tilde{X}})\Big),$$

and

$$\|Z_X - \tilde{Z}_{\tilde{X}}\|_\alpha \leqslant \boldsymbol{C}'\|\varphi\|_{C^3_b}\big(T^{3\alpha} \vee 1\big)\Big(\rho_\alpha(\boldsymbol{X}, \tilde{\boldsymbol{X}}) + |Y_0 - \tilde{Y}_0| + |Y'_0 - \tilde{Y}'_0| + d_{X,\tilde{X},2\alpha}(Y_X, \tilde{Y}_{\tilde{X}})\Big).$$

This means that Z_X and $\tilde{Z}_X$ are similar (s controlled rough paths) if $\boldsymbol{X}$ and $\tilde{\boldsymbol{X}}$ are similar as rough paths, and at the same time Y_X and $\tilde{Y}_{\tilde{X}}$ are similar as controlled rough paths.

Proof: The proof is not very difficult, but long and tedious. To not lose track of what we want to do, let us start with some preparations. We define

$$\boldsymbol{\Delta} = \rho_\alpha(\boldsymbol{X}, \tilde{\boldsymbol{X}}) + |Y_0 - \tilde{Y}_0| + |Y'_0 - \tilde{Y}'_0| + \|Y' - \tilde{Y}'\|_\alpha + \|R^Y - R^{\tilde{Y}}\|_{2\alpha}$$

and note that all estimates below must produce a factor of $\boldsymbol{\Delta}$; the prefactor is less important as long as it is finite.

The proof will consist of splitting

$$d_{X,\tilde{X},2\alpha}(Z_X, \tilde{Z}_{\tilde{X}}) = \|Z' - \tilde{Z}'\|_\alpha + \|R^Z - R^{\tilde{Z}}\|_{2\alpha}$$

into many pieces and proving for each part separately that it is bounded by a constant times $\boldsymbol{\Delta}$. For this, we will frequently use a few inequalities that we now collect. The first one is

$$\|Y\|_\infty \leqslant |Y_0| + T^\alpha \|Y\|_\alpha \qquad (I1).$$

For the second one, we invoke Lemma (4.38) and obtain

$$\|Y - \tilde{Y}\|_\alpha \leqslant C_0(1 \vee T^\alpha)\boldsymbol{\Delta}, \qquad (I2)$$

where C_1 only depends on the norms that $\boldsymbol{C}$ depends on. We will also need

$$\|Y\|_\alpha \leqslant C(1+\|X\|_\alpha)\big(|Y_0'| + T^\alpha \|Y_X\|_{X,2\alpha}\big) \leqslant C_1(T^\alpha \vee 1)\|X\|_\alpha \|Y_X\|_{X,2\alpha,0}, \qquad (I3)$$

see Proposition (4.35). Finally, since we need to estimate Hölder norms, we will always try to add and subtract terms so that we retain differences of the same quantity at different time points. For this, the identity (for suitable f)

$$f(y) - f(x) = \int_0^1 \partial_u f(x+u(y-x))\,\mathrm{d}u = \int_0^1 [Df](x+u(y-x))\mathrm{d}u(y-x) \qquad (*)$$

will be needed. A direct consequence of it is the Taylor estimate

$$f(x) - f(y) \leqslant \|Df\|_\infty |x-y| \qquad (I4)$$

a) We start by investigating $\|Z' - \tilde{Z}'\|_\alpha = \|D\varphi(Y)Y' - D\varphi(\tilde{Y})\tilde{Y}'\|_\alpha$. For two functions f, g, we have

$$(fg)_{s,t} = f_t g_t - f_s g_s = f_{s,t} g_t + f_s g_{s,t}.$$

This means that

$$\begin{aligned}(fg)_{s,t} - (\tilde{f}\tilde{g})_{s,t} &= f_{s,t} g_t + f_s g_{s,t} - \tilde{f}_{s,t}\tilde{g}_t - \tilde{f}_s \tilde{g}_{s,t} \\ &= f_{s,t}(g_t - \tilde{g}_t) + (f_{s,t} - \tilde{f}_{s,t})\tilde{g}_t + f_s(g_{s,t} - \tilde{g}_{s,t}) + (f_s - \tilde{f}_s)\tilde{g}_{s,t}.\end{aligned}$$

We apply this to $f = D\varphi(Y)$, $g = Y'$, $\tilde{f} = D\varphi(\tilde{Y})$ and $\tilde{g} = \tilde{Y}'$ and estimate the four terms on the right hand side above.

(i) The first term is bounded by

$$\begin{aligned}|f_{s,t}(g_t - \tilde{g}_t)| = |D\varphi(Y_t) - D\varphi(Y_s)||Y_t' - \tilde{Y}_t'| &\overset{(I4)}{\leqslant} \|D^2\varphi\|_\infty |Y_{s,t}| \|Y' - \tilde{Y}'\|_\infty \\ &\overset{(I1)}{\leqslant} \|D^2\varphi\|_\infty |t-s|^\alpha \|Y\|_\alpha \big(|Y_0' - \tilde{Y}_0'| + T^\alpha \|Y' - \tilde{Y}'\|_\alpha\big) \\ &\overset{(I3)}{\leqslant} |t-s|^\alpha \|\varphi\|_{C_b^3} C_1(T^\alpha \vee 1)\|X\|_\alpha \|Y_X\|_{X,2\alpha,0}(T^\alpha \vee 1)\boldsymbol{\Delta}.\end{aligned}$$

Dividing by $|t-s|^\alpha$ and taking the supremum over $s \neq t$, we obtain an estimate of the correct form for the first term.

(ii): The second term is the most tricky one. It is given by

$$(f_{s,t} - \tilde{f}_{s,t})\tilde{g}_t = (D\varphi(Y) - D\varphi(\tilde{Y}))_{s,t}\tilde{Y}_t' = (D\varphi(Y_t) - D\varphi(Y_s) - D\varphi(\tilde{Y}_t) + D\varphi(\tilde{Y}_s))\tilde{Y}_t',$$

and the problem is that we want a product of a term that is a difference of evaluations of the same quantity at different times t, s, and another term that is the difference of a quantity with a tilde and one without, at the same time. The most elegant way seems to use $(*)$ and obtain

$$\begin{aligned}\big([D\varphi](Y) - [D\varphi](\tilde{Y})\big)_{s,t} = &\int_0^1 \mathrm{d}r\, \underbrace{\big([D^2\varphi](Y_s + rY_{s,t}) - [D^2\varphi](\tilde{Y}_s + r\tilde{Y}_{s,t})\big)}_{=:\delta(r,s,t)} Y_{s,t} \\ &+ \int_0^1 \mathrm{d}r\, [D^2\varphi](\tilde{Y}_s + r\tilde{Y}_{s,t})(Y_{s,t} - \tilde{Y}_{s,t}).\end{aligned}$$

We have

$$\begin{aligned}|\delta(r,s,t)| &\overset{(I4)}{\leqslant} \|D^3\varphi\|_\infty\big|Y_s + rY_{s,t} - \tilde Y_s - r\tilde Y_{s,t}\big| \leqslant \|D^3\varphi\|_\infty\Big(\|Y_s-\tilde Y_s\|_\infty + r|Y_{s,t}-\tilde Y_{s,t}|\Big)\\ &\overset{(I1)}{\leqslant} \|\varphi\|_{C_b^3}\Big(|Y_0-\tilde Y_0| + T^\alpha\|Y-\tilde Y\|_\alpha + r|t-s|^\alpha\|Y-\tilde Y\|_\alpha\Big)\\ &\overset{(I2)}{\leqslant} \|\varphi\|_{C_b^3}\Big(|Y_0-\tilde Y_0| + (1+r)T^\alpha C_0(1\vee T^\alpha)\mathbf{\Delta}\Big) \leqslant \|\varphi\|_{C_b^3}\Big(1+(1+r)C_0(T^\alpha\vee T^{2\alpha})\Big)\mathbf{\Delta}.\end{aligned}$$

Similarly,

$$\Big|\int_0^1 \mathrm{d}r\,[D^2\varphi](\tilde Y_s + r\tilde Y_{s,t})(Y_{s,t}-\tilde Y_{s,t})\Big| \leqslant \|D^2\varphi\|_\infty|t-s|^\alpha\|Y-\tilde Y\|_\alpha \overset{(I2)}{\leqslant} \|\varphi\|_{C_b^3}|t-s|^\alpha C_0(1\vee T^\alpha)\mathbf{\Delta}.$$

Together, this gives

$$\|[D\varphi](Y) - [D\varphi](\tilde Y)\|_\alpha \leqslant \|\varphi\|_{C_b^3}\mathbf{\Delta}\Big(\big(1+\frac{3}{2}C_0(T^\alpha\vee T^{2\alpha})\big)\|Y\|_\alpha + C_0(1\vee T^\alpha)\Big)$$

By using $(I3)$ on the expression $\|Y\|_\alpha$ (which produces another factor of T^α), we see that this too is of the desired form.

(iii): The third term is

$$|f_s(g_{s,t}-\tilde g_{st})| \leqslant |D\varphi(Y_s)||Y'_{s,t}-\tilde Y'_{s,t}| \leqslant \|\varphi\|_{C_b^3}\|Y'-\tilde Y'\|_\alpha|t-s|^\alpha,$$

which immediately leads to a bound of the correct form.

(iv): The fourth term is

$$|(f_s-\tilde f_s)\tilde g_{s,t}| \leqslant |D\varphi(Y_s) - D\varphi(\tilde Y_s)||\tilde Y'_{s,t}| \leqslant \|\varphi\|_{C_b^3}\|Y-\tilde Y\|_\infty\|\tilde Y'\|_\alpha|t-s|^\alpha,$$

and now we apply inequalities $(I1)$, $(I2)$ and $(I3)$ to obtain a suitable bound.

b) Next we investigate the term $\|R^Z - R^{\tilde Z}\|_{2\alpha}$. We have

$$R^Z_{s,t} = Z_{s,t} - D\varphi(Y_s)Y'_sX_{s,t} = Z_{s,t} - D\varphi(Y_s)Y_{s,t} + D\varphi(Y_s)R^Y_{s,t},$$

and so

$$R^Z_{s,t} - R^{\tilde Z}_{s,t} = \underbrace{Z_{s,t} - D\varphi(Y_s)Y_{s,t} - \tilde Z_{s,t} - D\varphi(\tilde Y_s)\tilde Y_{s,t}}_{=:A_{s,t}} + \underbrace{D\varphi(Y_s)R^Y_{s,t} - D\varphi(\tilde Y_s)R^{\tilde Y}_{s,t}}_{=:B_{s,t}}.$$

As above, we get

$$Z_{s,t} = \int_0^1 D\varphi(Y_s + rY_{s,t})\,\mathrm{d}rY_{s,t} = \int_0^1 \mathrm{d}r\Big(D\varphi(Y_s)Y_{s,t} + \int_0^r \mathrm{d}uD^2\varphi(Y_s+uY_{s,t})Y_{s,t}\otimes Y_{s,t}\Big),$$

we rearrange this and obtain

$$\begin{aligned}A_{s,t} = \int_0^1\mathrm{d}r\int_0^r\mathrm{d}u\Big(&\big(D^2\varphi(Y_s+uY_{s,t}) - D^2\varphi(\tilde Y_s + u\tilde Y_{s,t})\big)Y_{s,t}\otimes Y_{s,t}+\\ &+ D^2\varphi(\tilde Y_s+u\tilde Y_{s,t})\big(Y_{s,t}\otimes Y_{s,t} - \tilde Y_{s,t}\otimes\tilde Y_{s,t}\big)\Big)\end{aligned}$$

We use inequalities (I4) and (I1) on the first term under the integral, and use the equality $f \otimes f - \tilde{f} \otimes \tilde{f} = (f - \tilde{f}) \otimes f + \tilde{f} \otimes (f - \tilde{f})$ in the second term, and obtain

$$\|A\|_{2\alpha} \leqslant \|D^3\varphi\|_\infty \big(\|Y - \tilde{Y}\|_\infty + T^\alpha \|Y - \tilde{Y}\|_\alpha\big) \|Y_\alpha\|^2 + \|D^2\varphi\|_\infty \|Y - \tilde{Y}\|_\alpha (\|Y\|_\alpha + \|\tilde{Y}\|_\alpha).$$

With the same considerations as in the first part, we get a bound of the form $C_3 \|\varphi\|_{C_b^3} (T^3\alpha \vee 1)\boldsymbol{\Delta}$, with C_3 depending on the suitable norms only. The final term is again easier, we have

$$\|B\|_{2\alpha} \leqslant \|D^2\varphi\|_\infty \|Y - \tilde{Y}\|_\infty \|R^Y\|_{2\alpha} + \|D\varphi\|_\infty \|R^Y - R^{\tilde{Y}}\|_{2\alpha},$$

which again gives the correct bound with the same estimates we used above. We have now shown the first claimed inequality. The second one follows from Lemma (4.38). ☐

(4.47) Definition

Let $\boldsymbol{X} \in \mathcal{C}^\alpha([0,T], V)$ and $f \in C_b^2(W, \mathcal{L}(V, W))$. We say that $Y \in C^\alpha([0,T], W)$ **solves the rough differential equation (RDE)**

$$\dot{Y}_s = f(Y_s)\dot{X}_s, \qquad Y_0 = \xi \in W$$

if

(i): with $Y'_s := f(Y_s)$ we have $Y_X := (Y, Y') \in \mathcal{D}_X^{2\alpha}([0,T], W)$,

(ii): $Y_t = \xi + \int_0^t f(Y_s) \, \mathrm{d}\boldsymbol{X}_s$ for all $t \in [0,T]$.

(4.48) Remark

Stochastic differential equations are often of the form

$$\mathrm{d}Y_t = f(Y_t) \, \mathrm{d}t + \sigma(Y_t) \mathrm{d}B_t.$$

For RDE, this corresponds to

$$\dot{Y}_t = g(Y_t) + h(Y_t)\dot{X}_t.$$

This case is actually already included in the above definition by adding a „smooth component“ to the original rough path: we choose

$$\tilde{X}_t = \begin{pmatrix} X_t \\ t \end{pmatrix} \in C^\alpha([0,T], V \times \mathbb{R})$$

and

$$\tilde{\mathbb{X}}_{s,t} = \begin{pmatrix} \mathbb{X}_{s,t} & \int_s^t X_{s,r} \mathrm{d}r \\ \int_s^t (r - s) \mathrm{d}X_s & \frac{1}{2}(t-s)^2 \end{pmatrix}.$$

The integrals on the off-diagonal are simply Young integrals. Then $\tilde{\boldsymbol{X}}$ is an α-rough path if $\boldsymbol{X}$ is one, and the above RDE takes the form

$$\dot{Y}_t = F(Y_t)\dot{\tilde{X}}_t \qquad \text{with} \qquad F(Y_t) = (h(Y_t), g(Y_t)).$$

For the proof of the main result of this chapter, we need one more technical lemma.

(4.49) Lemma: stability of rough integration

Let $\boldsymbol{X} = (X, \mathbb{X}) \in \mathcal{C}^\alpha([0,T], V)$, $Y_X \in \mathcal{D}_X^{2\alpha}([0,T], \mathcal{L}(V,W))$,

$$Z_X := \int_0^\cdot Y_s \,\mathrm{d}\boldsymbol{X}_s, \qquad Z'_X = Y.$$

Assume the same for the quantities $\tilde{\boldsymbol{X}}$, $\tilde{Y}_{\tilde{X}}$ and $\tilde{Z}_{\tilde{X}}$. Then, with $c(\alpha) = 2^{3\alpha}\zeta(\alpha)$, we have

$$\begin{aligned} d_{X,\tilde{X},2\alpha}(Z_X, \tilde{Z}_{\tilde{X}}) &\leqslant \rho_\alpha(\boldsymbol{X}, \tilde{\boldsymbol{X}})\big(3c(\alpha)T^\alpha \|Y_X\|_{X,2\alpha} + 2|Y'_0|\big) \\ &+ \rho_\alpha(\mathbf{0}, \tilde{\boldsymbol{X}})\Big(T^\alpha \|Y' - \tilde{Y}'\|_\alpha + |Y'_0 - \tilde{Y}'_0| + T^\alpha(1 + \|X\|_\alpha + c(\alpha)) d_{X,\tilde{X},2\alpha}(Y_X, \tilde{Y}_{\tilde{X}})\Big) \end{aligned}$$

For the case $\boldsymbol{X} = \tilde{\boldsymbol{X}}$ and $Y'_0 = \tilde{Y}'_0$, the right hand side is thus bounded by a multiple of T^α.

Proof: We define

$$\Xi_{s,t} = Y_s X_{s,t} + Y'_s \mathbb{X}_{s,t}, \quad \tilde{\Xi}_{s,t} = \tilde{Y}_s \tilde{X}_{s,t} + \tilde{Y}'_s \tilde{\mathbb{X}}_{s,t}, \qquad \Delta_{s,t} = \Xi_{s,t} - \tilde{\Xi}_{s,t}.$$

Then as in (4.9 d) we obtain

$$\Delta_{u,w} - \Delta_{u,v} - \Delta_{v,w} = -R^Y_{u,v} X_{v,w} - Y'_{u,v}\mathbb{X}_{v,w} + R^{\tilde{Y}}_{u,v}\tilde{X}_{v,w} + \tilde{Y}'_{u,v}\tilde{\mathbb{X}}_{v,w}.$$

By the usual trick of writing $fg - \tilde{f}\tilde{g} = f(g - \tilde{g}) - (\tilde{f} - f)\tilde{g}$, we obtain

$$\begin{aligned} \|\delta\Delta\|_{3\alpha} &:= \sup_{u<v<w} \frac{1}{|w-u|^{3\alpha}} |\Delta_{u,w} - \Delta_{u,v} - \Delta_{v,v}| \\ &\leqslant \|R^Y\|_{2\alpha}\|X - \tilde{X}\|_\alpha + \|\tilde{X}\|_\alpha \|R^Y - R^{\tilde{Y}}\|_{2\alpha} + \|Y'\|_\alpha \|\mathbb{X} - \tilde{\mathbb{X}}\|_{2\alpha} + \|\tilde{\mathbb{X}}\|_{2\alpha}\|Y' - \tilde{Y}'\|_\alpha \\ &\leqslant \rho_\alpha(\boldsymbol{X}, \tilde{\boldsymbol{X}})\|Y_X\|_{X,2\alpha} + \rho_\alpha(\tilde{\boldsymbol{X}}, \mathbf{0}) d_{X,\tilde{X},2\alpha}(Y_X, \tilde{Y}_{\tilde{X}}). \end{aligned}$$

By the Sewing Lemma (4.11), the intergal $\mathcal{I}\Delta$ exists, and then it must equal $\mathcal{I}\Xi - \mathcal{I}\tilde{\Xi}$. Since

$$R^Z_{s,t} = \int_s^t Y_r \mathrm{d}\boldsymbol{X}_r - Y_s X_{s,t} = (\mathcal{I}\Xi)_{s,t} - \Xi_{s,t} + Y'\mathbb{X}_{s,t},$$

and analogously for $R^{\tilde{Z}}_{s,t}$, Lemma (4.11) gives

$$\begin{aligned} |R^Z_{s,t} - R^{\tilde{Z}}_{s,t}| &\leqslant |(\mathcal{I}\Delta)_{s,t} - \Delta_{s,t}| + |Y'_s||\mathbb{X}_{s,t} - \tilde{\mathbb{X}}_{s,t}| + |\tilde{\mathbb{X}}_{s,t}||Y'_s - \tilde{Y}'_s| \\ &\leqslant c(\alpha)\|\delta\Delta\|_{3\alpha}|t-s|^{3\alpha} + \|Y'\|_\infty \rho_\alpha(\boldsymbol{X}, \tilde{\boldsymbol{X}})|t-s|^{2\alpha} + \|\tilde{\mathbb{X}}\|_{2\alpha}\|Y' - \tilde{Y}'\|_\infty |t-s|^{2\alpha} \\ &\leqslant |t-s|^{2\alpha}\Big(c(\alpha)T^\alpha\big(\rho_\alpha(\boldsymbol{X}, \tilde{\boldsymbol{X}})\|Y_X\|_{X,2\alpha} + \rho_\alpha(\tilde{\boldsymbol{X}}, \mathbf{0}) d_{X,\tilde{X},2\alpha}(Y_X, \tilde{Y}_{\tilde{X}})\big) + \\ &\qquad + \rho_\alpha(\boldsymbol{X}, \tilde{\boldsymbol{X}})\big(|Y'_0| + T^\alpha\|Y'\|_\alpha\big) + \|\tilde{\mathbb{X}}\|_{2\alpha}\big(|Y'_0 - \tilde{Y}'_0| + T^\alpha\|Y' - \tilde{Y}'\|_\alpha\big)\Big). \end{aligned}$$

This gives the estimate on the $\|R^Z - R^{\tilde{Z}}\|_{2\alpha}$-part of the reqiured distance. The estimate on the part $\|Z' - \tilde{Z}'\|_\alpha$ follows from the fact that $Z' = Y$ and $\tilde{Z}' = \tilde{Y}$, and from Lemma (4.38). Combining all the terms and doing some cosmetic estimates give the claim. □

(4.50) Theorem: existence of solutions to RDE

Let $\xi \in W$, $f \in C^3_b(W, \mathcal{L}(V,W))$, $\boldsymbol{X} \in \mathcal{C}^\beta([0,T], W)$ for some $\beta \in (1/3, 1/2]$. Then the RDE

$$\dot{Y}_s = f(Y_s)\dot{\boldsymbol{X}}_s, \qquad Y_0 = \xi$$

has a unique solution $Y \in C^\beta([0,T],W)$.

Proof: We will first prove that there exists a solution of slightly lower regularity, namely a solution $Y \in C^\alpha([0,T],W)$ for some $\alpha \in (1/3,\beta)$. In a second step we will then show that this solution must even be in $C^\beta([0,T],W)$. The reason for this step is that for $\boldsymbol{X} \in \mathcal{C}^\beta$ and small enough $\delta > 0$, we have

$$\|\boldsymbol{X}\|_{\delta,\alpha} := \|X\|_{\delta,\alpha} + \sqrt{\|\mathbb{X}\|_{\delta,2\alpha,\mathrm{diag}}} \leqslant \delta^{\beta-\alpha} \|\boldsymbol{X}\|_{\delta,\beta}\,.$$

This means that for any given β-rough path $\boldsymbol{X}$, we can make its α-rough norm as small as we want by choosing δ small enough.

The strategy is then the same as for the existence proof for classical ordinary differential equations. Let

$$\mathcal{M}_\delta : \mathcal{D}_X^{2\alpha}([0,\delta],W) \to \mathcal{D}_X^{2\alpha}([0,\delta],W), \quad (Y,Y') \mapsto \Big(\Big(\xi + \int_0^u f(Y_s)\,\mathrm{d}\boldsymbol{X}_s\Big)_{0\,\leqslant\, u\,\leqslant\,\delta}, f(Y)\Big)$$

be the „right hand side operator“ of the rough differential equation on the time interval $[0,\delta]$. $\mathcal{M}_\delta$ really maps $\mathcal{D}_X^{2\alpha}([0,\delta],W)$ to itself: for $Y \in C^\alpha([0,T],W)$, we have $f(Y) \in C^\alpha([0,\delta],\mathcal{L}(V,W))$ by Proposition (4.45), and thus $\mathcal{M}_\delta(Y,Y') \in \mathcal{D}_X^{2\alpha}([0,\delta],W)$ by Theorem (4.36 c).

$Y \in \mathcal{D}_X^{2\alpha}([0,\delta],W)$ is a solution of the RDE (written in integral notation) if and only if for all $t \leqslant \delta$,

$$Y_t = \xi + \int_0^t f(Y_s)\,\mathrm{d}\boldsymbol{X}_s, \qquad Y'_t = f(Y_t).$$

In other words, a solution to the RDE is a fixed point of the map $\mathcal{M}_\delta$. To prove the existence of such a fixed point, we want to use Banach's fixed point theorem. To do this, we will find a closed subset

$$U_\delta \subset \mathcal{D}_X^{2\alpha}([0,\delta],W) \subset C^\alpha([0,\delta],W) \times C_2^{2\alpha}([0,\delta],W\otimes W)$$

of the Banach space $C^\alpha([0,\delta],W) \times C_2^{2\alpha}([0,\delta],W\otimes W)$ and show that this subset is invariant under the map $\mathcal{M}_\delta$. We will then show that $\mathcal{M}_\delta$ is a contraction on U_δ with respect to the metric $d_{X,2\alpha,0}$, where

$$d_{X,2\alpha,0}(Y_X,\tilde{Y}_X) := |Y_0 - \tilde{Y}_0| + |Y'_0 - \tilde{Y}'_0| + \|Y' - \tilde{Y}'\|_\alpha + \|R^Y - R^{\tilde{Y}}\|_{2\alpha}.$$

These are the ingredients needed to apply Banach's fixed point theorem, which will then give the existence of a unique fixed point, hence a unique solution.

Step 1: Definition and invariance of U_δ:

A subset of $\mathcal{D}_X^{2\alpha}([0,\delta],W)$ that could contain a solution must certainly contain at least (at least some) elements with the correct initial condition. Define therefore

$$\tilde{\mathcal{D}}_X^{2\alpha} \equiv \tilde{\mathcal{D}}_X^{2\alpha}([0,\delta],W) := \{Y_X \in \mathcal{D}_X^{2\alpha}([0,\delta],W) : Y_0 = \xi, Y'_0 = f(\xi)\}.$$

Clearly $\mathcal{M}_\delta(\tilde{\mathcal{D}}_X^{2\alpha}) = \tilde{\mathcal{D}}_X^{2\alpha}$, so this subset is already left invariant. But it is too large. What we need is a small ball (in the $d_{X,2\alpha,0}$-metric) around a simple element from $\tilde{\mathcal{D}}_X^{2\alpha}$. For (first order) ordinary differential equations the right choice for the center of the ball would simply be the constant function with the correct initial condition. This does not work here, since $Y \in \mathcal{D}_X^{2\alpha}$ means that $R^Y \in C_2^{2\alpha}$, but with $R_{s,t} = Y_{s,t} - Y'_s X_{s,t}$ the necessary cancellation of roughness does not take place if we choose $(Y_s,Y'_s) = (\xi, f(\xi))$; in this case $R_{s,t}$ will usually be only in C_2^α. This is why we choose as midpoint the controlled rough path $\bar{Y}_X = (\bar{Y}, \bar{Y}')$ with $\bar{Y}_s = \xi + f(\xi)X_s$ and $\bar{Y}'_s = f(\xi)$ for all s. Then $R^{\bar{Y}} \equiv 0$, and thus for any $Y_X \in \tilde{\mathcal{D}}_X^{2\alpha}$, we have

$$d_{X,2\alpha,0}(Y_X, \bar{Y}_X) = \|Y' - f(\xi)\|_\alpha + \|R^Y\|_{2\alpha} = \|Y'\|_\alpha + \|R^Y\|_{2\alpha} = \|Y_X\|_{X,2\alpha},$$

and we define

$$U_\delta := \{Y_X \in \tilde{\mathcal{D}}_X^{2\alpha} : d_{X,2\alpha,0}(Y_X, \bar{Y}_X) \leqslant 1\} = \{Y_X \in \tilde{\mathcal{D}}_X^{2\alpha} : \|Y_X\|_{X,2\alpha} \leqslant 1\}.$$

Since we already know that $\tilde{\mathcal{D}}_X^{2\alpha}$ is invariant under $\mathcal{M}_\delta$, what remains to show is that there exists $\delta_0 > 0$ so that for all $\delta < \delta_0$ and all $Y_Y \in U_\delta$, we have $\|\mathcal{M}_\delta Y_X\|_{X,2\alpha} \leqslant 1$. Let $\delta > 0$. We write

$$(Z, Z') := (f(Y), f(Y)') = (f(Y), Df(Y)Y')$$

and estimate

$$\begin{aligned}\|\mathcal{M}_\delta(Y_X)\|_{X,2\alpha} &= \Big\|\Big(\int_0^\cdot Z'_r \,\mathrm{d}\boldsymbol{X}_r, Z\Big)\Big\|_{X,2\alpha} \\ &\overset{(4.36c)}{\leqslant} \|Z\|_\alpha + \|Z'\|_\infty \|\mathbb{X}\|_{2\alpha} + \tilde{C}(\alpha)(\delta^\alpha \vee 1)\big(\|X\|_\alpha \|R^Z\|_{2\alpha} + \|\mathbb{X}\|_{2\alpha}\|Z'\|_\alpha\big) = (*)\end{aligned}$$

Since $Z_{s,t} = Z'_s X_{s,t} + R^Z_{s,t}$, we have $\|Z\|_\alpha \leqslant \|Z'\|_\infty \|X\|_\alpha + \delta^\alpha \|R^Z\|_{2\alpha}$, and we also have $\|Z'\|_\infty \leqslant |Z'_0| + \delta^\alpha \|Z'\|_\alpha = |Df(\xi)f(\xi)| + \delta^\alpha \|Z'\|_\alpha$. This gives

$$\begin{aligned}(*) &\leqslant (|Z'_0| + \delta^\alpha \|Z'\|_\alpha)\, \|\boldsymbol{X}\|_\alpha + \delta^\alpha \|R^Z\|_{2\alpha} + (|Z'_0| + \delta^\alpha \|Z'\|_\alpha)\, \|\boldsymbol{X}\|_\alpha^2 \\ &\quad + \tilde{C}(\alpha)(\delta^\alpha \vee 1)(\|R^Z\|_{2\alpha} + \|Z'\|_\alpha)(\|\boldsymbol{X}\|_\alpha + \|\boldsymbol{X}\|_\alpha^2),\end{aligned}$$

and assuming $\delta \leqslant 1$ we have

$$(*) \leqslant C_0(\alpha)(|Z'_0| + \|Z_X\|_{X,2\alpha})(\|\boldsymbol{X}\|_\alpha + \|\boldsymbol{X}\|_\alpha^2 + \delta^\alpha),$$

with the constant $C_0(\alpha)$ only depending on α. By Proposition (4.45), we have

$$\|Z_X\|_{X,2\alpha} \leqslant \|D\varphi\|_\infty \underbrace{\|Y_X\|_{X,2\alpha}}_{\leqslant 1} + \|D^2\varphi\|_\infty \big(\|Y\|_\alpha^2 + \|Y\|_\alpha \|Y'\|_\infty\big).$$

As above (and keeping in mind that $\delta \leqslant 1$), we find

$$\begin{aligned}\|Y\|_\alpha &\leqslant \|Y'\|_\infty \|X\|_\alpha + \delta^\alpha \|R^Y\|_{2\alpha} \leqslant (|Y'_0| + \delta^\alpha \|Y'\|_\alpha)\|X\|_\alpha + \delta^\alpha \|R^Y\|_{2\alpha} \\ &\leqslant (|f(\xi)| + \|Y_X\|_{X,2\alpha})(1 + \|X\|_\alpha) \leqslant (|f(\xi)| + 1)(1 + \|X\|_\alpha),\end{aligned}$$

with a similar estimate for $\|Y'\|_\infty$. We find that there is a constant $C(\alpha, f)$ only depending on α and $\|f\|_{C_b^2}$ such that

$$\|\mathcal{M}_\delta(Y_X)\|_{X,2\alpha} \leqslant C(\alpha, f)(\|\boldsymbol{X}\|_\alpha + \|\boldsymbol{X}\|_\alpha^2 + \delta^\alpha)$$

Since actually $\boldsymbol{X} \in \mathcal{C}^\beta$, we can use the remark at the beginning of the proof to choose $\delta_0 > 0$ so that for all $\delta < \delta_0$, the right hand side is bounded by 1. This shows the invariance. Note that δ_0 does *not* depend on the starting point ξ, only on $f(\xi)$, which is controlled by $\|f\|_{C_b^2}$.

Step 2: Contraction property.

We will show that there exists $\delta_0 > 0$ so that for all $\delta < \delta_0$, and all $Y_X, \tilde{Y}_X \in U_\delta$, we have

$$d_{X,2\alpha,0}(\mathcal{M}_\delta(Y_X), \mathcal{M}_\delta(\tilde{Y}_X)) \leqslant \frac{1}{2} d_{X,2\alpha,0}(Y_X, \tilde{Y}_X). \qquad (**)$$

As above, we set $(Z,Z')=(f(Y),Df(Y)Y')$, use Lemma (4.49) in the special case $\boldsymbol{X}=\tilde{\boldsymbol{X}}$, assume in addition that $\delta\leqslant 1$, and obtain

$$\begin{aligned} d_{X,2\alpha,0}(\mathcal{M}_\delta(Y_X),\mathcal{M}_\delta(\tilde{Y}_X)) &\leqslant \delta^\alpha\rho_\alpha(\mathbf{0},\boldsymbol{X})\Big(\|Z'-\tilde{Z}'\|_\alpha+\underbrace{\delta^{-\alpha}|Z'_0-\tilde{Z}'_0|}_{=0}\\ &\qquad +(1+\|X\|_\alpha+c(\alpha))d_{X,2\alpha}(Z_X,\tilde{Z}_X)\Big)\\ &\leqslant \delta^\alpha\big(2+\|X\|_\alpha+c(\alpha)\big)\rho_\alpha(\mathbf{0},\boldsymbol{X})d_{X,2\alpha}(Z_X,\tilde{Z}_X). \end{aligned}$$

By Proposition (4.46), we have

$$d_{X,2\alpha}(Z_X,\tilde{Z}_X)\leqslant \boldsymbol{C}d_{X,2\alpha}(Y_X,\tilde{Y}_X),$$

where the constant $\boldsymbol{C}$ only depends on $\|X\|_\alpha$, $\|Y_X\|_{X,2\alpha}$ and $\|\tilde{Y}_X\|_{X,2\alpha}$ which are both bounded by one, and $|Y'_0|$ and $|\tilde{Y}'_0|$, which are both equal to $f(\xi)$. This means that

$$d_{X,2\alpha,0}(\mathcal{M}_\delta(Y_X),\mathcal{M}_\delta(\tilde{Y}_X))\leqslant \delta^\alpha C_1 d_{X,2\alpha}(Y_X,\tilde{Y}_X)$$

with a constant that does not depend on ξ. Making δ small enough we obtain $(**)$.

Step 3: Wrapping up and going from C^α to C^β.

By the previous two steps, we can now apply Banach's fixed point theorem and conclude that there exists a solution $(Y,Y')\in\mathcal{D}_X^{2\alpha}([0,\delta],W)$ to the RDE for sufficiently small $\delta>0$. We can then extend this solution to the whole time interval $[0,T]$ by iteration: having obtained the solution in the interval $[0,\delta]$, we start at time δ with the initial condition $\xi_\delta=Y_\delta\in W$, and repeat the procedure. We have been careful to make sure that none of our estimates depends on the initial condition of the RDE, so we find an interval of the same length δ for this starting point. Now we have a solution on $[0,2\delta]$ by just putting the two together (check that the required equality still holds). Continuing in this way we can cover $[0,T]$.

What remains to show is that our solution is actually in $\mathcal{D}_X^{2\beta}([0,\delta],W)$ for the original β, i.e. we need to show that $\|Y'\|_\beta<\infty$ and $\|R^Y\|_{2\beta}<\infty$. We know that

$$\frac{1}{|t-s|^\beta}|Y_{s,t}|=\frac{1}{|t-s|^\beta}\Big|Y'_sX_{s,t}+R^Y_{s,t}\Big|\leqslant\|Y'\|_\infty\frac{1}{|t-s|^\beta}|X_{s,t}|+\frac{1}{|t-s|^\beta}|R^Y_{s,t}|.$$

We assumed $X\in C^\beta$, and we proved above that $R^Y\in C_2^{2\alpha}$. Since $2\alpha>2/3>1/2\geqslant\beta$, $R^Y\in C_2^\beta$, and taking suprema shows $\|Y\|_\beta<\infty$. Since $Y'=f(Y)$, this implies $\|Y'\|_\beta<\infty$. To estimate R^Y, we use that Y solves the RDE: we have

$$R^Y_{s,t}=Y_{s,t}-Y'_sX_{s,t}=\int_s^t f(Y_r)\,\mathrm{d}\boldsymbol{X}_r-f(Y_s)X_{s,t}.$$

With $Z=f(Y)$, $Z'=Df(Y)Y'$ as above we find

$$\begin{aligned} |R^Y_{s,t}| &\leqslant \Big|\int_s^t Z_r\,\mathrm{d}\boldsymbol{X}_r-Z_sX_{s,t}-Z'_s\mathbb{X}_{s,t}\Big|+|Z'_s\mathbb{X}_{s,t}|\\ &\overset{(4.36a)}{\leqslant} c(\alpha)(\|X\|_\alpha\|R^Z\|_{2\alpha,\mathrm{diag}}+\|Y'\|_\alpha\|\mathbb{X}\|_{2\alpha,\mathrm{diag}})|t-s|^{3\alpha}+\|Z'\|_\infty\|\mathbb{X}\|_{2\beta}|t-s|^{2\beta}. \end{aligned}$$

Since $3\alpha>1\geqslant 2\beta$, this shows $\|R^Y\|_{2\beta,\mathrm{diag}}<\infty$. The proof is finished. □

(4.51) Remark

If we only have $f \in C^3$ instead of C_b^3, the proof of the previous theorem can still be done. In several places we were using that

$$f(Y_t) - f(\tilde{Y}_t) = \int_0^1 Df(Y_t + r(Y_t - \tilde{Y}_t))\,\mathrm{d}r(Y_t - \tilde{Y}_t)$$

and similar expressions, and were then just estimating the integral by $\|Df\|_\infty$. If instead we estimate it by $\sup\{|f(Y_t + r\tilde{Y}_t)| : Y_X \in U_\delta(\xi)\}$, we get away with local boundedness of Df instead of uniform boundedness, because U_δ is a bounded set. However, now all the estimates *will* depend on ξ via the supremum above. We can still prove existence of the solution for some interval δ and then iterate, but the next interval δ_2 might be considerably shorter, and the sum of all the δ that we can get in this way might be finite. This is not unexpected, because it also happens with regular ODE: if f grows too quickly, there is *explosion*, i.e. solutions go to infinity (ore become otherwise ill-defined) in finite time. In this case, one can obtain only local solutions.

Our next big result will be about the continuity of the solution map to RDE: for two RDE $\dot{Y}_t = f(Y)\dot{\boldsymbol{X}}_t$ and $\dot{\tilde{Y}}_t = f(\tilde{Y})\dot{\tilde{\boldsymbol{X}}}_t$ where only the rough path „driving“ the RDE is different, the distance of the solutions can be bounded by the distance of the rough paths plus the distance of starting points. To make this theorem sufficiently powerful, we first need an a priori estimate that tells us how large a solution to a RDE can at most get.

(4.52) A priori estimates

Let $f \in C_b^3$, and let Y be a solution of the RDE $\dot{Y}_t = f(Y_t)\dot{\boldsymbol{X}}_t$, $Y_0 = \xi$. Then

a) $|f(Y)'_{s,t}| \leqslant 2\|f\|^2_{C_b^2}|Y_{s,t}|$.

b) $|R^{f(Y)}_{s,t}| \leqslant \|f\|_{C_b^2}(\frac{1}{2}|Y_{s,t}|^2 + |R^Y_{s,t}|)$.

c) Set $\|X\|_{\alpha,h} := \sup\{\frac{1}{|t-s|^\alpha}|X_{s,t}| : |t-s| \leqslant h\}$. Then there exists $h_0 > 0$, depending only on $\|\boldsymbol{X}\|_\alpha$ and $\|f\|_{C_b^2}$, such that for all $h \leqslant h_0$, we have

$$\|R^Y\|_{2\alpha,h} \leqslant 2\big(\tfrac{1}{2}\|Y\|_{\alpha,h} + \|f\|_{C_b^2}\|\mathbb{X}\|^{1/2}_{2\alpha,h,\mathrm{diag}}\big)^2.$$

d) With $c(\alpha) = 2^{3\alpha}\zeta(3\alpha)$,

$$\|Y\|_\alpha \leqslant (2c(\alpha))^{\frac{1-\alpha}{\alpha}} T^{1-\alpha}(\|f\|_{C_b^2}\,\|\boldsymbol{X}\|_\alpha)^{1/\alpha} + \|f\|_{C_b^2}\,\|\boldsymbol{X}\|_\alpha\,.$$

e) There exists a constant C' depending only on $\|f\|_{C_b^2}$, $\|\boldsymbol{X}\|_\alpha$, α and T such that $\|Y_X\|_{X,2\alpha} \leqslant C'$.

Proof:

a) We have

$$\begin{aligned} f(Y)'_{s,t} &= Df(Y_t)Y'_t - Df(Y_s)Y'_s = Df(Y_t)Y'_{s,t} + (Df(Y_t) - Df(Y_s))Y'_s \\ &= Df(Y_t)f(Y)_{s,t} + Df(Y)_{s,t}f(Y_s), \end{aligned}$$

and so

$$|f(Y)'_{s,t}| \leqslant \|Df\|^2_\infty|Y_{s,t}| + \|D^2f\|_\infty|Y_{s,t}|\|f\|_\infty,$$

which implies the claim.
b) We have

$$R^{f(Y)}_{s,t} = f(Y)_{s,t} - Df(Y_s)Y'_s X_{s,t} = f(Y)_{s,t} - Df(Y_s)Y_{s,t} + Df(Y_s)R^Y_{s,t}.$$

The first two terms on the right hand side are bounded by $\frac{1}{2}\|D^2 f\|_\infty |Y_{s,t}|^2$, and the last one is bounded by $\|Df\|_\infty |R^Y_{s,t}|$. This proves the claim.
c) We have

$$R^Y_{s,t} = Y_{s,t} - Y'_s X_{s,t} = \int_s^t f(Y_r)\,\mathrm{d}\boldsymbol{X}_r - \underbrace{f(Y_s)X_{s,t} - f(Y)'_s \mathbb{X}_{s,t}}_{=\Xi_{s,t}} + f(Y)'_s \mathbb{X}_{s,t}.$$

By (4.36 a), we find that for s,t with $|t-s| \leqslant h$, and with $c(\alpha) = 2^{3\alpha}\zeta(3\alpha)$,

$$|R^Y_{s,t}| \leqslant c(\alpha)\big(\|X\|_{\alpha,h}\|R^{f(Y)}\|_{2\alpha,h,\mathrm{diag}} + \|f(Y)'\|_{\alpha,h}\|\mathbb{X}\|_{2\alpha,h,\mathrm{diag}}\big)|t-s|^{3\alpha} + \|f\|^2_{C^2_b}\|\mathbb{X}\|_{2\alpha,h,\mathrm{diag}}|t-s|^{2\alpha}.$$

By parts a) and b), we find

$$\|R^Y\|_{2\alpha,h,\mathrm{diag}} \leqslant c(\alpha)h^\alpha\Big(\|X\|_{\alpha,h}\|f\|_{C^2_b}(\tfrac{1}{2}\|Y\|^2_{\alpha,h} + \|R^Y\|_{2\alpha,h,\mathrm{diag}}) + 2\|f\|^2_{C^2_b}\|\mathbb{X}\|_{2\alpha,h,\mathrm{diag}}\|Y\|_{\alpha,h}\Big) + \|f\|^2_{C^2_b}\|\mathbb{X}\|_{2\alpha,h,\mathrm{diag}}.$$

Now set $h_0 = (2c(\alpha)\|f\|_{C^2_b}\,\|\!|\boldsymbol{X}|\!\|_\alpha)^{-1/\alpha}$, then for all $h \leqslant h_0$ we have

$$c(\alpha)h^\alpha\|X\|_{\alpha,h}\|f\|_{C^2_b} \leqslant \frac{\|X\|_{\alpha,h}}{2\,\|\!|\boldsymbol{X}|\!\|_\alpha} \leqslant \frac{1}{2}, \qquad \text{and} \qquad c(\alpha)h^\alpha\|f\|_{C^2_b}\|\mathbb{X}\|^{1/2}_{2\alpha,h,\mathrm{diag}} \leqslant \frac{\sqrt{\|\mathbb{X}\|_{\alpha,h}}}{2\,\|\!|\boldsymbol{X}|\!\|_\alpha} \leqslant \frac{1}{2}.$$

Thus,

$$\|R^Y\|_{2\alpha,h,\mathrm{diag}} \leqslant \frac{1}{4}\|Y\|^2_{\alpha,h} + \frac{1}{2}\|R^Y\|_{2\alpha,h,\mathrm{diag}} + \|f\|_{C^2_b}\|\mathbb{X}\|^{1/2}_{2\alpha,h,\mathrm{diag}}\|Y\|_{\alpha,h} + \|f\|^2_{C^2_b}\|\mathbb{X}\|_{2\alpha,h,\mathrm{diag}}.$$

Subtracting $\frac{1}{2}\|R^Y\|_{2\alpha,h,\mathrm{diag}}$ on both sides and recognizing the perfect square gives the result.
d) Set $D = \|f\|_{C^2_b}\,\|\!|\boldsymbol{X}|\!\|_\alpha$. We have $Y_{s,t} = f(Y_s)X_{s,t} + R^Y_{s,t}$, and thus

$$\|Y\|_{\alpha,h} \leqslant \|f\|_\infty\|X\|_{\alpha,h} + \|R^Y\|_{2\alpha,h,\mathrm{diag}}h^\alpha \overset{c)}{\leqslant} D + 2h^\alpha\big(\tfrac{1}{2}\|Y\|_{\alpha,h} + D\big)^2, \qquad (*)$$

for $h \leqslant h_0$ with h_0 from c). We define $J := \{h \in (0,h_0] : \|Y\|_{\alpha,h} \leqslant 2D\}$ and note that by $(*)$, $J \neq \emptyset$. We claim that $h_1 := \sup J = h_0$. To see this, first note that $h \mapsto g(h) := \|Y\|_{\alpha,h}$ is continuous on $(0,T]$. Indeed, g is clearly monotone increasing, and for all $\delta > 0$,

$$\begin{aligned} g(h+\delta) &\leqslant \max\Big\{g(h), \sup\Big\{\frac{|Y_{s,t}|}{|t-s|^\alpha} : h \leqslant |s-t| \leqslant h+\delta\Big\}\Big\} \\ &\leqslant \max\Big\{g(h), \frac{1}{h^\alpha}\sup\{|Y_{s,s+h}| : 0 \leqslant s \leqslant T-h-\delta\} \\ &\qquad + \frac{1}{h^\alpha}\sup\{|Y_{s+h,t}| : 0 \leqslant s < t \leqslant T : 0 \leqslant t-(s+h) \leqslant \delta\}\Big\} \\ &\leqslant g(h) + \Big(\frac{\delta}{h}\Big)^\alpha g(\delta). \end{aligned}$$

This shows continuity. Assume now that $h_1 < h_0$. Then $\|Y\|_{\alpha,h_1} = g(h_1) = 2D$. Inserting this into $(*)$, we obtain the inequality

$$2D \leqslant D + 8h_1^\alpha D^2 \qquad \Longrightarrow \qquad h_1^\alpha \geqslant \frac{1}{8D}.$$

This means that $h_1 \geqslant (8D)^{-1/\alpha}$. On the other hand, we have $h_0 = (2c(\alpha)D)^{-1/\alpha}$. It is not hard to check (e.g. on the computer) that $c(\alpha) = 2^{3\alpha}\zeta^{3/\alpha} \geqslant 8$ for all α (we only would need $1 < \alpha < 3/2$), which means that $h_0 < (8D)^{-1/\alpha}$. This shows that $h_1 = h_0$.

Let now $0 \leqslant s < t \leqslant T$. We set $s_0 = s$ and choose $n = \left\lfloor \frac{t-s}{h_0} \right\rfloor \geqslant 0$ points $s_1, \dots, s_n$ so that $s_i - s_{i-1} = h_0$ for all $i \geqslant 1$. By the triangle inequality, we then have

$$\begin{aligned} |Y_{s,t}| &\leqslant \sum_{j=1}^n |Y_{s_i,s_{i-1}}| + |Y_{s_n,t}| \leqslant n\|Y\|_{\alpha,h_0} h_0^\alpha + \|Y\|_{\alpha,h_0}|t-s|^\alpha \\ &\leqslant 2D\Big(\frac{t-s}{h_0} h_0^\alpha + |t-s|^\alpha\Big) \leqslant 2D(t-s)^\alpha \Big(\big(\tfrac{T}{h_0}\big)^{1-\alpha} + 1\Big). \end{aligned}$$

Since $h_0^{\alpha-1} = (2c(\alpha)D)^{-1+1/\alpha}$, we arrive at

$$\frac{|Y_{s,t}|}{|t-s|^\alpha} \leqslant (2c(\alpha))^{(1-\alpha)/\alpha} T^{1-\alpha} D^{1/\alpha} + D$$

The claim follows by taking the supremum over $s \neq t$.

e) For h_0 as in d), we have

$$\begin{aligned} \|R^Y\|_{2\alpha,\text{diag}} &\leqslant \|R^Y\|_{2\alpha,h_0,\text{diag}} + \sup\Big\{\frac{R^Y_{s,t}}{|t-s|^{2\alpha}} : |t-s| \geqslant h_0\Big\} \leqslant \|R^Y\|_{2\alpha,h,\text{diag}} + h_0^{-2\alpha}\|R^Y\|_\infty \\ &\leqslant \|R^Y\|_{2\alpha,h,\text{diag}} + \frac{T^\alpha}{h_0^{2\alpha}}(\|Y\|_\alpha + \|f\|_\infty \|X\|_\alpha). \end{aligned}$$

We now use part c) to estimate $\|R^Y\|_{2\alpha,h,\text{diag}}$, producing a term $\|Y\|_{\alpha,h}$ which we estimate by $\|Y\|_\alpha$. Now we apply part d) to both occurrances of $\|Y\|_\alpha$ and obtain the result. ☐

(4.53) Theorem

Let $f \in C_b^3(W, \mathcal{L}(V,W))$, $1/3 < \alpha \leqslant 1/2$, and set

$$\mathcal{B}_M^\alpha := \{X \in \mathcal{C}^\alpha([0,T],V) : |\!|\!|\boldsymbol{X}|\!|\!|_\alpha \leqslant M\}.$$

Then there is a constant C depending only on α, $\|f\|_{C_b^3}$, M and T such that for all $\boldsymbol{X}, \tilde{\boldsymbol{X}} \in \mathcal{B}_M^\alpha$, the solutions Y, $\tilde{Y}$ of the RDEs

$$\dot{Y}_t = f(Y_t)\dot{\boldsymbol{X}}_t, \qquad Y_0 = \xi.$$

and

$$\dot{\tilde{Y}}_t = f(\tilde{Y}_t)\dot{\tilde{\boldsymbol{X}}}_t, \qquad \tilde{Y}_0 = \tilde{\xi},$$

satisfy the following two estimates:

a) $d_{X,\tilde{X},2\alpha}(Y_X, \tilde{Y}_{\tilde{X}}) \leqslant C\big(|\xi - \tilde{\xi}| + \rho_\alpha(\boldsymbol{X}, \tilde{\boldsymbol{X}})\big)$,

b) $\|Y - \tilde{Y}\|_\alpha \leqslant C\big(|\xi - \tilde{\xi}| + \rho_\alpha(\boldsymbol{X}, \tilde{\boldsymbol{X}})\big)$.

Proof: We set

$$Z_t = f(Y_t), \qquad Z'_t = Df(Y_t)Y'_t, \qquad \text{and} \qquad W_t = \int_0^t Z_s \,\mathrm{d}\boldsymbol{X}_s.$$

Since Y solves the RDE, we have $W_t = Y_t - \xi$, and $Y'_t = f(Y_t)$. We choose $f(Y_t)$ as a Gubinelli derivative for $Y - \xi$ as well. Then $R^{Y-\xi}_{s,t} = R^Y_{s,t}$.

We do the same for the quantities with the tilde. Then

$$d_{X,\tilde{X},2\alpha}(Y_X, \tilde{Y}_{\tilde{X}}) = d_{X,\tilde{X},2\alpha}(Y_X - \xi, \tilde{Y}_{\tilde{X}} - \tilde{\xi}).$$

then follows from the definition of rough path distance and from the equality $R^{Y-\xi}_{s,t} = R^Y_{s,t}$.

Let $\delta > 0$, we will later choose it small enough. Lemma (4.49) with $T = \delta$ now gives

$$\begin{aligned} d_{X,\tilde{X},2\alpha}(Y_X, \tilde{Y}_{\tilde{X}}) &= d_{X,\tilde{X},2\alpha}(W_X, \tilde{W}_{\tilde{X}}) \leqslant \rho_\alpha(\boldsymbol{X}, \tilde{\boldsymbol{X}})\big(3c(\alpha)\delta^\alpha \|Z_X\|_{X,2\alpha} + 2|Z'_0|\big) \\ &\quad + \rho_\alpha(\boldsymbol{0}, \boldsymbol{X})\Big(\delta^\alpha \|Z' - \tilde{Z}'\|_\alpha + |Z'_0 - \tilde{Z}'_0| + \delta^\alpha(1 + \|X\|_\alpha + c(\alpha))d_{X,\tilde{X},2\alpha}(Z_X, \tilde{Z}_{\tilde{X}})\Big) \end{aligned}$$

By Proposition (4.45), $\|Z_X\|_{X,2\alpha}$ is bounded by a constant only depending on $\|f\|_{C^2_b}$ and $\|Y_X\|_{X,2\alpha}$, and $|Z'_0| \leqslant |Df(Y_0)||Y'_0| \leqslant \|f\|_{C^2_b}|f(Y_0)| \leqslant \|f\|^2_{C^2_b}$. By Theorem (4.52), $\|Y_X\|_{X,2\alpha}$ is itself bounded by a constant only depending on $\|f\|_{C^2_b}$, $\|\!|\boldsymbol{X}|\!\|_\alpha$, α and T, and thus the factor multiplying $\rho_\alpha(\boldsymbol{X}, \tilde{\boldsymbol{X}})$ only depends on these quantities. Turning to the quantities in the second line, we have

$$Z'_0 - \tilde{Z}'_0 = Df(Y_0)Y'_0 - Df(\tilde{Y}_0)\tilde{Y}'_0 = Df(\xi)f(\xi) - Df(\tilde{\xi})f(\tilde{\xi}),$$

and by the usual $gh - \tilde{g}\tilde{h} = g(h - \tilde{h}) + (g - \tilde{g})h$ trick, we find that

$$|Z'_0 - \tilde{Z}'_0| \leqslant \|Df\|_\infty |f(\xi) - f(\tilde{\xi})| + \|f\|_\infty |Df(\xi) - Df(\tilde{\xi})| \leqslant \|f\|^2_{C^2_b}|\xi - \tilde{\xi}|.$$

By definition, $\|Z' - \tilde{Z}'\|_\alpha \leqslant d_{X,\tilde{X},2\alpha}(Z_X, \tilde{Z}_{\tilde{X}})$, and by Proposition (4.46),

$$d_{X,\tilde{X},2\alpha}(Z_X, \tilde{Z}_{\tilde{X}}) \leqslant \boldsymbol{C}\|f\|_{C^3_b}\big(\delta^{3\alpha} \vee 1\big)\Big(\rho_\alpha(\boldsymbol{X}, \tilde{\boldsymbol{X}}) + |Y_0 - \tilde{Y}_0| + |Y'_0 - \tilde{Y}'_0| + d_{X,\tilde{X},2\alpha}(Y_X, \tilde{Y}_{\tilde{X}})\Big).$$

We have $|Y_0 - \tilde{Y}_0| = |\xi - \tilde{\xi}|$, and estimate $|Y'_0 - \tilde{Y}'_0| = |f(\xi) - f(\tilde{\xi})| \leqslant \|f\|_{C^1_b}|\xi - \tilde{\xi}|$. In summary, we find two constants $\bar{C}_1$, $\bar{C}_2$ and $\bar{C}_3$ only depending on $\|f\|_{C^2_b}$, $\|\!|\boldsymbol{X}|\!\|_\alpha$, α and T such that

$$d_{X,\tilde{X},2\alpha}(Y_X, \tilde{Y}_{\tilde{X}}) \leqslant \bar{C}_1\rho_\alpha(\boldsymbol{X}, \tilde{\boldsymbol{X}}) + \bar{C}_2|\xi - \tilde{\xi}| + \bar{C}_3\delta^\alpha d_{X,\tilde{X},2\alpha}(Y_X, \tilde{Y}_{\tilde{X}}).$$

It remains to choose δ such that $\bar{C}_3\delta^\alpha = 1/2$ and rearrange, to find that on $[0, \delta]$ we have

$$d_{X,\tilde{X},2\alpha}(Y_X, \tilde{Y}_{\tilde{X}}) \leqslant 2(\bar{C}_1\rho_\alpha(\boldsymbol{X}, \tilde{\boldsymbol{X}}) + \bar{C}_2|\xi - \tilde{\xi}|). \qquad (*)$$

It remains to cover the whole interval $[0, T]$ instead of just $[0, \delta]$. For this, first note that the size of δ only depends on $\bar{C}_3$ and thus only on the „allowed" quantities. We have $|Y_\delta - \tilde{Y}_\delta| \leqslant \delta^\alpha \|Y - \tilde{Y}\|_{\alpha,\delta}$, and by Lemma (4.38) and similar estimates as above, we have

$$\|Y - \tilde{Y}\|_{\alpha,\delta} \leqslant \|f\|_\infty \rho_\alpha(\boldsymbol{X}, \tilde{\boldsymbol{X}}) + M\|Df\|_\infty |\xi - \tilde{\xi}| + \delta^\alpha(1 + M)d_{X,\tilde{X},2\alpha}(Y_X, \tilde{Y}_{\tilde{X}}),$$

where the controlled rough path distance is over the interval $[0, \delta]$ and we therefore can bound it by $(*)$. We thus find that

$$|Y_\delta - \tilde{Y}_\delta| \leqslant \bar{C}_4\rho_\alpha(\boldsymbol{X}, \tilde{\boldsymbol{X}}) + \bar{C}_5|\xi - \tilde{\xi}|,$$

with constants $\bar{C}_4$ and $\bar{C}_5$ only depending on allowed constants. This enables us to choose Y_δ and $\tilde{Y}_\delta$ as initial conditions for the RDEs on the interval $[\delta, 2\delta]$, repeat the above estimates (with the same constants as in the previous interval), and find that on $[0, 2\delta]$, we have

$$d_{X,\tilde{X},2\alpha}(Y_X, \tilde{Y}_{\tilde{X}}) \leqslant 2(\bar{C}_1 \rho_\alpha(\boldsymbol{X}, \tilde{\boldsymbol{X}}) + \bar{C}_2|\xi - \tilde{\xi}|) + 2(\bar{C}_1 \rho_\alpha(\boldsymbol{X}, \tilde{\boldsymbol{X}}) + \bar{C}_4 \rho_\alpha(\boldsymbol{X}, \tilde{\boldsymbol{X}}) + \bar{C}_5|\xi - \tilde{\xi}|).$$

We only need to repeat this procedure a number of times that only depends on the allowed constants, and thus obtain the desired estimate.

For part b), we use Lemma (4.38) and once more the estimate $|\tilde{Y}_0' - Y_0'| \leqslant \|Df\|_\infty |\tilde{\xi} - \xi|$. □

(4.54) Remark

Theorem (4.53) is probably the most important result in the theory of rough differential equations (and needs almost all of the theory we have developed so far). Why is it so important?

In many applications, a RDE of the type

$$\dot{Y}_t = f(Y_t) + g(Y_t)\dot{\boldsymbol{X}}_t$$

is interpreted as a physical system that is perturbed by „random“ noise: the idealized, noiseless system would fulfil the ODE $\dot{Y}_t = f(Y_t)$, but the presence of the noise $(\dot{\boldsymbol{X}}_t)$ changes the behaviour of the system - the factor $g(Y_t)$ just tells us how important the noise is at different values of Y_t.

The statement of Theorem (4.53) now says that two realizations of the „random noise“ (really: two rough paths, that need not be and usually are not random at all) lead to very similar solutions if the are similar *in rough path distance*. The last bit is really crucial: if we would not introduce the second rough path component, there is no chance to obtain such a result e.g. for Brownian rough paths. A corresponding negative result is cited in Friz/Hairer as Proposition 1.1. Apart from the intuitive appeal of the statement that „similar perturbations lead to similar solutions“, Theorem (4.53) will also be useful for rather quick proofs of important results - see below.

It seems that all of the above results are limited by the requirement that f is uniformly bounded. However, in many cases this can be circumvented: Assume that (as will be the case below) we have a sequence $\boldsymbol{X}^{(n)}$ of rough paths that converges to $\boldsymbol{X}$ in rough path distance. Consider the RDE $\dot{Y}_t = f(Y_t)\dot{\boldsymbol{X}}_t$, similar for $\dot{Y}^{(n)}$. Assume that $f \in C^3$, but not bounded. Then the solution Y will exist, possibly only on a (small) interval $[0, T]$. Since the solution is continuous, $M := \sup_{t \leqslant T} |Y_t|$ is finite. We can thus replace f by a bounded C^3-function $\tilde{f}$ so that $f(\xi) = \tilde{f}(\xi)$ for all $|\xi| \leqslant 2M$. For this function, now all our theorems hold. In particular, the solution $\tilde{Y}^{(n)}$ of the RDE with $\tilde{f}$ will be close to Y when n is large, and thus for all large enough n will fulfil $\sup_{t \leqslant T} |Y_t^{(n)}| \leqslant 3/2M$. This means that it is equal to the solution for the original f, and we have extended our approximation result to f.

(4.55) Theorem

Let $f \in C_b^3(V, \mathcal{L}(V, W))$, B a Brownian motion, $\boldsymbol{B}^{\text{Itô}}$ as in (4.20). a) For all $\omega \in \Omega$, let $Y(\omega)$ be the solution to the RDE

$$\dot{Y}_t(\omega) = f(Y_t(\omega))\dot{\boldsymbol{B}}_t^{\text{Itô}}(\omega).$$

Then the random variable $\omega \mapsto (Y_t(\omega))_{t\in[0,T]}$ is a strong solution of the SDE

$$\mathrm{d}Y_t = f(Y_t)\,\mathrm{d}B_t$$

in the sense of Definition (3.49). b) If $\boldsymbol{B}^{\text{Itô}}$ is replaced by $\boldsymbol{B}^{\text{Strat}}$ above, then the solution Y of the RDE is a solution of the stochastic Stratonovich differential equation

$$Y_t = Y_0 + \int_0^t f(Y_s) \circ \mathrm{d}B_s.$$

Proof: The only thing we need to show is that Y is progressively measurable. Then (4.39) and (4.40) will ensure that the SDE hold in integral sense. We start with $\boldsymbol{B}^{\text{Strat}}$. Let $\boldsymbol{B}^{(n)}$ be the piecewise dyadic approximations of Brownian motion as given in (4.30), then Theorem (4.31) guarantees that $\rho(\boldsymbol{B}^{(n)}, \boldsymbol{B}^{\text{Strat}}) \to 0$ as $n \to \infty$, $\mathbb{P}$-almost surely. We immediately restrict our probability space to the case where the convergence holds. By Theorem (4.53), the map

$$\mathcal{C}_{\mathrm{g}}^{\alpha}([0,T],V) \to C^{\alpha}([0,T],\mathcal{L}(V,W)), \qquad \boldsymbol{X} \mapsto \text{ the solution of the RDE } \dot{Y}_t = f(Y_t)\dot{\boldsymbol{X}}_t$$

is locally Lipschitz, in particular continuous. Therefore the solutions $Y^{(n)}(\omega)$ of the approximate RDE $\dot{Y}_t^{(n)} = f(Y_t^{(n)})\dot{\boldsymbol{X}}_t^{(n)}$ converge to Y in C^{α}. It is not hard to check that each $Y^{(n)}$ (which is really the solution of an ordinary differential equation) is progressively measurable, and thus the limit is progressively measurable too. The claim thus follows in the case of $\boldsymbol{B}^{\text{Strat}}$. For $\boldsymbol{B}^{\text{Itô}}$, just use the formula $\boldsymbol{B}^{\text{Itô}} = \boldsymbol{B}^{\text{Strat}} + \frac{1}{2}(t-s)I$ and proceed as above. □

(4.56) Theorem (Wong-Zakai; Clark; Stroock-Varadhan)

Let B be a Brownian motion, $f \in C_b^3$, Y the solution of the Stratonovich SDE

$$\mathrm{d}Y_t = g(Y_t)\,\mathrm{d}t + f(Y_t) \circ \mathrm{d}B_t.$$

For each $n \in \mathbb{N}$ and each $\omega \in \Omega$, let $Y^{(n)}$ be the solution of the random ODE

$$\partial_t Y_t^{(n)}(\omega) = g(Y_t^{(n)}(\omega)) + f(Y_t^{(n)}(\omega))\partial_t B_t^{(n)}(\omega),$$

where $B^{(n)}$ is the dyadic linear approximation of B, and we set $\partial_t B_t^{(n)} = 0$ on the separating points of the dyadic partition. Then for all $\alpha < 1/2$,

$$\lim_{n\to\infty} \|Y(\omega) - Y^{(n)}(\omega)\|_{\alpha,0} = 0 \qquad \mathbb{P}\text{-almost surely.}$$

Proof: By Theorem (4.31), $\rho_\alpha(\boldsymbol{B}^{(n)}, \boldsymbol{B}^{\text{Strat}}) \to 0$ as $n \to \infty$, $\mathbb{P}$-almost surely. By Theorem (4.53), $\|Y - Y^{(n)}\|_\alpha \leqslant C\rho_\alpha(\boldsymbol{B}^{(n)}, \boldsymbol{B}^{\text{Strat}})$; the initial condition is the same for all n. This shows the claim. □