Stochastics III: Stochastic Analysis*

by Lucio Galeati

February 10, 2025

[1]

Table of contents

*. This article has been written using GNU TeXmacs [11].

1. These lecture notes are based on the ones given by Prof. Nicolas Perkowski in the WS 2022-23 and I'm very thankful to him for sharing them with me. If you spot any typo, please email me (lgaleati@zedat.fu-berlin.de); thanks to all who send comments.

Introduction and motivation

Stochastic analysis is the analysis of continuous time stochastic processes.

In Stochastics II, you already encountered discrete time processes as models for random phenomena that evolve in discrete time steps. Such processes can be constructed mathematically (Kolmogorov's criterion) and you have likely already seen some of the most important classes of processes (martingales, Markov chains), as well as some of the most fundamental results on their behavior (martingale convergence, martingale inequalities, fundamental theorem of Markov chains, Donsker's invariance principle, etc.).

But of course it is also be interesting to study continuous time phenomena, which do not evolve in clearly separated time steps. This is more complex: in discrete time, one time is following after the next one, and to describe stochastic processes we only have to understand how they transition from one step to the next. In continuous time there is no "next step", we can go from t to $t+1$, but also to $t+\frac{1}{2}$ or $t+\frac{\pi}{100}$. Therefore, we will need many new mathematical tools to describe and analyze continuous time stochastic processes.

As a motivation, let us look at some examples where continuous time stochastic processes arise:

Example. (Stock price) The pictures of stock price trajectories typically look very irregular, bouncing up and down constantly. As toy model for the evolution of a stock price, we can consider a Brownian motion (which already appeared in Stochastik II). *Brownian motion* is a continuous time stochastic process $(B_t)_{t\geqslant 0}$ with continuous trajectories, such that $B_t \sim \mathcal{N}(0,t)$ for all $t\geqslant 0$, where $\mathcal{N}(0,t)$ denotes the normal distribution with mean 0 and variance t, and such that $B_{t+s} - B_t$ is independent of $(B_r)_{0\leqslant r\leqslant t}$.

The intuition behind this model is that there are many small traders, who all independently of each other try to buy or sell the stock. Each time a small trader buys, the stock price moves up a bit. Each time they sell, it moves down a bit. The fact that the increments are normally distributed thus follows from the central limit theorem. Moreover, if we assume that the decisions of each trader is independent of all previous decisions by all the other traders, then we get the independence of $B_{t+s} - B_t$ and $(B_r)_{0\leqslant r\leqslant t}$.

We will see later in the lecture how to construct the Brownian motion and that the description above characterizes it uniquely. And we will study some of its path properties to see that it indeed behaves quite wildly and it resembles the familiar pictures of stock price trajectories.

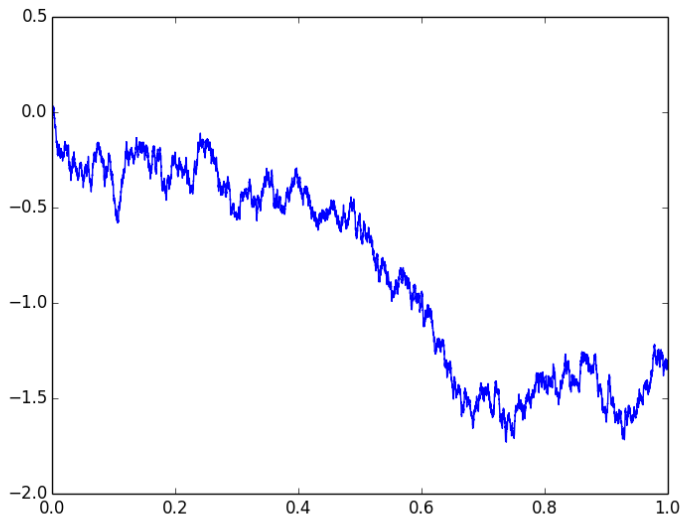


Figure 1. A typical realization of a Brownian motion.

For example, the Brownian motion has no isolated zeros, meaning that if $B_t = 0$ for some t, then in any small interval $[t-\varepsilon, t+\varepsilon]$ there are infinitely many s with $B_s = 0$. We will also see that B is nowhere differentiable and behaves roughly speaking like

$$|B_{t+\mathrm{d}t} - B_t| \simeq \sqrt{\mathrm{d}\, t}.$$

Of course, this is not a mathematical statement and part of the work will be to find a suitable mathematical statement that we can actually prove.

Exercise. Throughout the text there will be small exercises like this one, marked in blue. You should think about the exercises and try to solve them before the next lecture. Some of them will be elementary and only require you to revisit the definitions and make sure that you understood everything. Some others might be a bit more involved and might need some inspiration. It is no problem if you cannot solve an exercise, but you should always at least try.

Look at the y-axis of Figure 1. Do these numbers make you think of a stock price? If not, can you think of a simple transformation that we could do in order to obtain a more reasonable candidate for the price process?

Example. (Stochastic differential equations from Donker's invariance principle) The following difference equation is a prototypical example of a random discrete time evolution:

$$X_{n+1} = X_n + b(X_n) + \sigma(X_n)Y_n, \tag{1}$$

where Y_n is random influence, "noise". More concretely, one could consider for example a (stochastic) Malthusian population growth model, where X_n is the size of a population and

$$X_{n+1} = X_n + bX_n + X_nY_n,$$

where $b \in \mathbb{R}$ is the deterministic growth rate and $(Y_n)_{n\in\mathbb{N}}$ is a centered family of independent and identically distributed (i.i.d.) random variables that models randomly occurring deviations from the deterministic growth rate.

If we assume that the $(Y_n)_{n\in\mathbb{N}}$ are a centered family of i.i.d. random variables with finite variance, then Donsker's invariance principle (which you might have seen in Stochastik II) asserts that $S_n = \sum_{k=1}^n Y_k$ can be rescaled so that it converges to a Brownian motion. S is evolving in discrete time steps, but under the rescaling for Donsker's theorem the transition times between the steps become infinitely small, and in the limit one finds a continuous time variable.

It then seems reasonable to expect (and under suitable assumptions it can be proven) that X can be rescaled in such a way that it converges to a process $(Z_t)_{t\geqslant 0}$ satisfying for $t \geqslant 0$ and $h > 0$

$$Z_{t+h} = Z_t + b(Z_t)\, h + \sigma(Z_t)(B_{t+h} - B_t),$$

where B is a Brownian motion. Bringing Z_t to the left hand side, dividing by h and letting $h \to 0$, we formally obtain

$$\partial_t Z_t = b(Z_t) + \sigma(Z_t)\partial_t B_t.$$

But B is not differentiable in time, so it is not clear how to interpret this equation! To make sense of such "stochastic differential equations", and to this end first of "stochastic integrals", will be one of the main goals of the lecture.

Example. (Noise to model unresolved influences) Another situation where stochastic differential equations appear is the following: applied scientists often model time-evolving systems by ordinary differential equations (ODEs)

$$\dot{X}_t = b(X_t).$$

However, in reality the modelled system might not be isolated from its environment. To model influences of the environment, we have two choices:

i. We increase the dimension of our system by attempting to also model the environment; however this ultimately leads to an infinite-dimensional system, and often is unfeasible because nature is just too complex.

ii. We try to find a "random" model for the influence of the environment. Under suitable assumptions we should be able to invoke the central limit theorem, so that these random influences should be centered Gaussians.

In the second scenario, in many situations it is also reasonable to assume that the random influences are stationary in time, and independent for different times. So formally we end up with the equation

$$\dot{X}_t = b(X_t) + \xi_t,$$

where $(\xi_t)_{t\geqslant 0}$ is an i.i.d. family of centered Gaussian variables. It turns out that this equation does not make sense, because it is not possible to construct "a version" of ξ that has measurable trajectories and thus it is not clear how to interpret the equation. The solution to this problem is to formally assume that ξ_t has infinite variance for fixed times. We will see how to make this rigorous and how to model an ODE forced by "white noise" (which turns out to be, intuitively, the "time derivative" of Brownian motion).

Example. (A toy model for Earth's climate) A more concrete version of this example is a particle in a double well potential: Consider $b(x) = -U'(x)$ for $U(x) = \frac{1}{4}x^4 - \frac{1}{2}x^2$. One can easily verify that there are three fixed points for the dynamics $\dot{X}_t = -U'(X_t)$ (imagine a ball rolling down the potential U, except damping at the bottom): two stable fixed points $\{-1, 1\}$ and one unstable fixed point $\{0\}$.

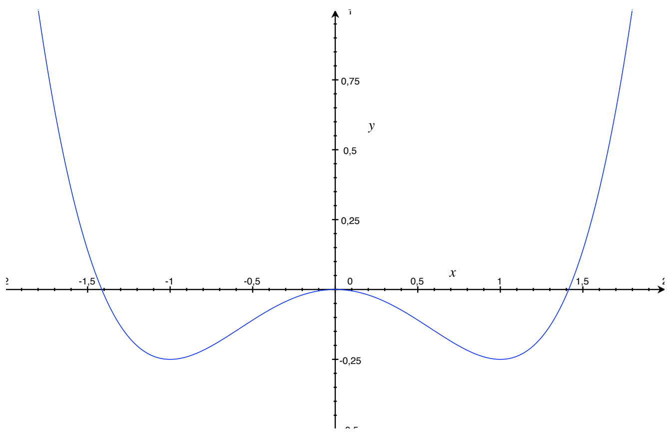


Figure 2. Double well potential U.

So if we start in $x<0$ the solution will converge to -1 for $t\to\infty$, and if we start in $x>0$ it will converge to 1.

The ODE $\dot{X}_t = -U'(X_t)$ could serve as a qualitative toy model for the earth's climate: Assume -1 represents an ice age and $+1$ a warm period. These two states are quite stable for the climate, after all we are not constantly switching between ice ages and warm periods. But from time to time there are transitions, and in the ODE model we never see them. But if we add a very small random forcing of white noise type, as described above, then the forcing can "kick" (rarely) the solution over the hill into the domain of attraction of the other stable fixed point. It might then be interesting to calculate how long this will typically take.

Example. (Stochastic gradient descent, we didn't cover it in the lectures) In many applied problems, e.g. statistical estimation or training of artificial neural networks in machine learning, we are interested in finding the minimum of a function of the form

$$F(x) = \frac{1}{n}\sum_{i=1}^{n} f_i(x),$$

where typically n is very large. A *gradient descent* would solve the ODE

$$\dot{X}_t = -\nabla F(X_t)$$

and for $t \to \infty$ the solution would converge to a local minimum of F (note that $\partial_t F(X_t) = -|\nabla F(X_t)|^2 < 0$; compare also with the double well example from above). We can implement a gradient descent on the computer with a simple Euler scheme:

$$X_{k+1} = X_k - \eta \nabla F(X_k) = X_k - \eta \frac{1}{n} \sum_{i=1}^{n} \nabla f_i(X_k),$$

where $\eta > 0$ is a small parameter, called the learning rate. However, in practice one usually uses the following *stochastic gradient descent* instead: At each step pick $i \in \{1, ..., n\}$ uniformly at random and set

$$X_{k+1} = X_k - \eta \nabla f_i(X_k).$$

This has two advantages: if n is large, the stochastic gradient descent is much cheaper to compute. And by introducing randomness into the algorithm it is no longer a pure descent and transitions to X_{k+1} with $F(X_{k+1}) > F(X_k)$ are possible. This means that we may exit the domain of attraction of a local minimum, and by carefully tuning the algorithm we might hope to converge to a global minimum. Sometimes it is argued that the stochastic gradient descent is similar to

$$X_{k+1} = X_k - \eta \nabla F(X_k) + \text{“noise”},$$

which by similar arguments as above is a discretization of the differential equation

$$\partial_t X_t = -\nabla F(X_t) + \partial_t B_t.$$

So if we understand the behavior of this SDE for $t \to \infty$, then we may learn something about stochastic gradient descents.

Example. (Brownian motion and PDEs) If B is a Brownian motion, then for $t > 0$ the random variable B_t has the density

$$p(t, x) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{x^2}{2t}\right).$$

It is a simple exercise to verify that p solves the *heat equation*:

$$\partial_t p(t, x) = \frac{1}{2} \partial_{xx}^2 p(t, x)$$

for all $t > 0$ and $x \in \mathbb{R}$. As a consequence, we get for any “nice” φ (i.e. nice enough so that the following manipulations are admissible) that the function

$$u(t, x) := \mathbb{E}[\varphi(x + B_t)]$$

solves

$$\begin{aligned} \partial_t u(t, x) &= \partial_t \left(\int_{\mathbb{R}} \varphi(x + y)\, p(t, y) dy \right) = \int_{\mathbb{R}} \varphi(x + y)\, \frac{1}{2} \partial_{yy}^2 p(t, y) dy \\ &= \int_{\mathbb{R}} \frac{1}{2} \partial_{yy}^2 \varphi(x + y) p(t, y) dy = \int_{\mathbb{R}} \frac{1}{2} \partial_{xx}^2 \varphi(x + y) p(t, y) dy = \frac{1}{2} \partial_{xx}^2 u(t, x), \end{aligned}$$

where we applied integration by parts to shift ∂_{yy}^2 from p to φ. Moreover, u obviously has the initial condition $u(0, x) = \varphi(x)$, so that we found a solution to the equation

$$\partial_t u = \partial_{xx}^2 u, \qquad u(0) = \varphi.$$

This suggests a link between stochastic processes and partial differential equations (PDEs), and in fact this link is quite deep and powerful. For example, if for $x \in \mathbb{R}$ the process X^x solves the stochastic differential equation

$$\partial_t X_t^x = b(X_t^x) + \sigma(X_t^x)\,\partial_t B_t, \qquad X_0 = x,$$

then $u(t,x) = \mathbb{E}[\varphi(X_t^x)]$ solves the (one-dimensional) PDE

$$\partial_t u(t,x) = b(x)\partial_x u(t,x) + \frac{1}{2}\sigma^2(x)\partial_{xx}^2 u(t,x), \qquad u(0) = \varphi,$$

and conversely the PDE can be used to characterize the law of X^x.

Example. (Diffusion Monte Carlo) In many applications we have to sample from a measure

$$\mu(\mathrm{d}x) = \frac{1}{Z}\exp(-V(x))\mathrm{d}x$$

on $\mathbb{R}^d$, where $V: \mathbb{R}^d \to \mathbb{R}$ is a differentiable function with $\int_{\mathbb{R}^d}\exp(-V(x))\mathrm{d}x < \infty$ and $Z = \int_{\mathbb{R}^d}\exp(-V(x))\mathrm{d}x$ is chosen so that μ is a probability measure. This is a very difficult problem, especially if d is large or V is complicated. One way of obtaining approximate samples from μ is to find a stochastic process with invariant measure μ. The most famous such process is the *Langevin diffusion*, which solves the stochastic differential equation

$$\partial_t X_t = -\nabla V(X_t) + \sqrt{2}\partial_t B_t.$$

Hopefully these examples show that there are many interesting questions to be asked and problems to be studied. We will now start to develop the basic tools and methods of stochastic analysis.

Literature

Large parts of the lecture are inspired by or directly taken from Le Gall's beautiful notes [16]. There is much more material in Le Gall's notes than we can cover in the lecture and they are a useful resource for further details. Further good references are the lecture notes by Jacod [12], the classic monographs [14, 23, 20, 13, 7], or the great "almost sure" blog https://almostsure.wordpress.com/. The monograph [18] is nearly entirely devoted to the Brownian motion, and it provides a much more detailed picture of its fascinating path properties than we can obtain in the lecture. In the beginning of the notes we repeat some material from Stochastics I & II, and good additional references are [15, 6, 26] and some chapters of [7].

Notation and conventions

- Unless explicitly mentioned otherwise, we always assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a given probability space.
- $\mathbb{N} = \{1, 2, \ldots\}$, $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, $\mathbb{R}_+ = [0, \infty)$, $\mathbb{Q}_+ = \mathbb{R}_+ \cap \mathbb{Q}$.
- $x \cdot y = \sum_{j=1}^d x_j y_j$, A^T is the transpose of the matrix A.
- $x^+ = \max\{x, 0\}$ and $x^- = \max\{-x, 0\}$.

- $\lim_{s\downarrow t} f(s) = \lim_{\varepsilon\to 0^+} f(t+\varepsilon)$ and $\lim_{s\uparrow t} f(s) = \lim_{\varepsilon\to 0^+} f(t-\varepsilon)$.
- $X_{s,t} := X_t - X_s$.
- The indicator function of a set A is denoted by $\mathbb{1}_A$.
- $\mathcal{B}(S)$ is the Borel σ-algebra of the topological space S. 2^Ω are the subsets of Ω.
- If we do not specify it, we always assume an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ as given.
- $a \lesssim b$ means there exists some $C > 0$, independent of the relevant variables under consideration, such that $a \leqslant C b$. For example, $(x+y)^2 \leqslant 2(y^2+y^2)$, so we would write $(x+y)^2 \lesssim x^2 + y^2$.

1 Gaussian processes, pre-Brownian motion and white noise

Some background with probability theory from a measure theoretic perspective will be assumed throughout the course; ideally all you need are things learned from Stochastik I and II. For convenience, some material is recalled in Appendices A.1 and A.2; this material will not be examinable, but it will be at the basis of many of the results we will develop, so please have a proper look at it to understand if you are already familiar enough with it.

1.1 Gaussian processes

The star of this lecture is the Brownian motion, which is a particular *Gaussian process*. Recall that for $d \in \mathbb{N}$ a random variable X with values in $\mathbb{R}^d$ is called *(centered) Gaussian* or *(centered) normal* if for any $u \in \mathbb{R}^d$ the linear combination

$$u \cdot X = \sum_{j=1}^{d} u_j X_j$$

of the entries of X has a one-dimensional (centered) Gaussian distribution. We also call $(X_1, \ldots, X_d)$ *jointly Gaussian*.

Equivalently, there exist $m \in \mathbb{R}^d$ and a symmetric positive semi-definite matrix $C \in \mathbb{R}^{d\times d}$ such that X has the characteristic function

$$\mathbb{E}[e^{iu\cdot X}] = e^{iu\cdot m - (u^T C u)/2}, \qquad u \in \mathbb{R}^d.$$

Moreover,

$$\mathbb{E}[u\cdot X] = u\cdot m, \qquad \mathrm{var}(u\cdot X) = u^T C u.$$

We write $X \sim \mathcal{N}(m, C)$.

Definition 1.1. *Let $\mathbb{T} \neq \emptyset$ be an index set. A real-valued stochastic process $X = (X_t)_{t\in\mathbb{T}}$ is called a* (centered) Gaussian process *if for every finite subset $I \subset \mathbb{T}$ and for all $(\alpha_t)_{t\in I} \in \mathbb{R}^I$ the random variable $\sum_{t\in I} \alpha_t X_t$ is a real-valued (centered) Gaussian random variable.*

Exercise. Show that $X = (X_1, \ldots, X_d)$ is a d-dimensional Gaussian random variable if and only if $(X_t)_{t\in\mathbb{T}}$ with $\mathbb{T} = \{1, \ldots, d\}$ is a d-dimensional Gaussian process indexed by $\mathbb{T}$.

Equivalently, X is (centered) Gaussian if for any finite $I \subset \mathbb{T}$ the vector $(X_t)_{t\in I}$ is an $|I|$-dimensional Gaussian random variable. There exist two functions $m: \mathbb{T} \to \mathbb{R}$ and $\Gamma: \mathbb{T} \times \mathbb{T} \to \mathbb{R}$ such that $\mathbb{E}[X_t] = m(t)$ for all $t \in \mathbb{T}$ and $\mathrm{cov}(X_s, X_t) = \Gamma(s,t)$ for all s, $t \in \mathbb{T}$. Moreover, the finite-dimensional distributions (and thus the law) of X are uniquely determined by m and Γ, and Γ is *symmetric* (i.e. $\Gamma(s,t) = \Gamma(t,s)$ for all $s, t \in \mathbb{T}$) and *positive semi-definite*, i.e. for any finite $I \subset \mathbb{T}$ and any $(\alpha_t)_{t\in I} \in \mathbb{R}^I$ we have

$$\sum_{(s,t)\in I\times I} \alpha_s \alpha_t \Gamma(s,t) = \mathrm{var}\left(\sum_{t\in I} \alpha_t X_t\right) \geqslant 0.$$

We say that X has *mean* m and *covariance* Γ.

We can construct Gaussian random variables on $\mathbb{R}^d$ with a given mean m and covariance C by transforming a "standard" d-dimensional random variable $Y \sim \mathcal{N}(0, \mathbb{I})$ via $X = m + \sqrt{C}Y$. It is not clear how to adapt this construction to infinite $\mathbb{T}$, so given m and Γ we need more sophisticated tools to construct a Gaussian process with mean m and covariance Γ. We achieve this with Kolmogorov's extension theorem:

Proposition 1.2. *Let $\mathbb{T} \neq \emptyset$ be an index set, let $m: \mathbb{T} \to \mathbb{R}$, and let $\Gamma: \mathbb{T} \times \mathbb{T} \to \mathbb{R}$ be a symmetric and positive semi-definite function. Then there exists a Gaussian process X with mean m and covariance Γ, and the law of X is uniquely determined by m and Γ.*

Proof. (Skipped in class, because of possible overlap with Stochastics II)

This follows from Kolmogorov's extension theorem. For any finite subset $I \subset \mathbb{T}$ let $\mathbb{P}_I$ be the law of a $\mathcal{N}((m(t))_{t\in I}, (\Gamma(s,t))_{s,t\in I})$ random variable. For $J \supset I$ let $\pi_{JI}: \mathbb{R}^J \to \mathbb{R}^I$ be the projection $\pi_{JI}((x_t)_{t\in J}) = (x_t)_{t\in I}$. The existence of X follows from Kolmogorov's extension criterion once we show the consistency condition $\mathbb{P}_J \circ \pi_{JI}^{-1} = \mathbb{P}_I$. For $u \in \mathbb{R}^I$ we have

$$\begin{aligned}\int_{\mathbb{R}^I} e^{iu\cdot x} \mathbb{P}_J \circ \pi_{JI}^{-1}(\mathrm{d}x) &= \int_{\mathbb{R}^J} e^{i\sum_{t\in I} u(t)x(t)} \mathbb{P}_J(\mathrm{d}x) = e^{i\sum_{t\in I} u(t)m(t) - \frac{1}{2}\sum_{s,t\in I} u(s)\Gamma(s,t)u(t)} \\ &= \int_{\mathbb{R}^I} e^{i\sum_{t\in I} u(t)x(t)} \mathbb{P}_I(\mathrm{d}x).\end{aligned}$$

So $\mathbb{P}_J \circ \pi_{JI}^{-1}$ and $\mathbb{P}_I$ have the same characteristic function, and thus the two measures agree and the proof is complete. □

——— End of the lecture on October 16 ———

Example 1.3. (Pre-Brownian motion) Let $\mathbb{T} = \mathbb{R}_+ = [0, \infty)$, $m(t) = 0$ and $\Gamma(s,t) = s \wedge t := \min\{s,t\}$. Then Γ is obviously symmetric. It is also positive semi-definite: For $n \in \mathbb{N}$ and $t_1, \ldots, t_n \geqslant 0$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ we have

$$\sum_{i,j=1}^n \alpha_i \alpha_j \Gamma(t_i, t_j) = \sum_{i,j=1}^n \alpha_i \alpha_j \int_0^\infty \mathbb{1}_{[0,t_i]}(s) \mathbb{1}_{[0,t_j]}(s) \mathrm{d}s = \int_0^\infty \left(\sum_{i=1}^n \alpha_i \mathbb{1}_{[0,t_i]}(s)\right)^2 \mathrm{d}s \geqslant 0.$$

The Gaussian process B with mean m and covariance Γ is called a *pre-Brownian motion*.

Later, we will define a Brownian motion as a pre-Brownian motion with continuous trajectories.

Lemma 1.4. (Alternative characterization of the pre-Brownian motion) *Let $(B_t)_{t\geqslant 0}$ be a real-valued stochastic process. Then B is a pre-Brownian motion if and only if the following conditions are satisfied:*

i. $B_0 = 0$ *almost surely;*

ii. for all $0 \leqslant s < t$ *the random variable* $B_t - B_s$ *is independent of the variables* $(B_r)_{0\leqslant r\leqslant s}$*;*

iii. for all $0 \leqslant s < t$ *we have* $B_t - B_s \sim \mathcal{N}(0, t-s)$.

Exercise. Prove the above lemma.

Example 1.5. (pre-Fractional Brownian motion, or pre-fBm for short) Let $H \in (0, 1)$, let $\mathbb{T} = \mathbb{R}_+$ and $m(t) = 0$ and

$$\Gamma^H(s,t) = \frac{1}{2}(t^{2H} + s^{2H} - |t-s|^{2H}).$$

Γ is obviously symmetric, but it is not so easy to see that it is positive semi-definite. One way of showing it is similar to our argument for pre-Brownian motion: one can show that

$$\Gamma^H(s,t) = \int_{\mathbb{R}} \Phi(s,r)\Phi(t,r)\mathrm{d}r$$

for

$$\Phi(s,r) := \frac{1}{\gamma(H+1/2)}\big((s-r)_+^{H-1/2} - (-r)_+^{H-1/2}\big),$$

where γ is the Gamma function and (just for this time) we set $x_+ := x^+ = \max\{x, 0\}$. Therefore

$$\sum_{i,j=1}^n \alpha_i\alpha_j\Gamma^H(t_i,t_j) = \sum_{i,j=1}^n \alpha_i\alpha_j\int_{\mathbb{R}} \Phi(t_i,r)\Phi(t_j,r)\mathrm{d}r = \int_{\mathbb{R}}\left(\sum_{i=1}^n \alpha_i\Phi(t_i,r)\right)^2 \mathrm{d}r \geqslant 0.$$

The Gaussian process B^H with mean $m \equiv 0$ and covariance Γ^H is called the *fractional pre-Brownian motion* with *Hurst index* H. In other words, pre-fBm of parameter $H \in (0,1)$ is characterized by being a Gaussian process $(B_t^H)_{t\geqslant 0}$ with

$$\mathbb{E}[B_t^H] = 0, \qquad \mathbb{E}[B_t^H B_s^H] = \frac{1}{2}(t^{2H} + s^{2H} - |t-s|^{2H}).$$

Exercise. Show that for $H = 1/2$ the process B^H is a pre-Brownian motion.

One can show that B^H becomes more and more irregular if we decrease H.

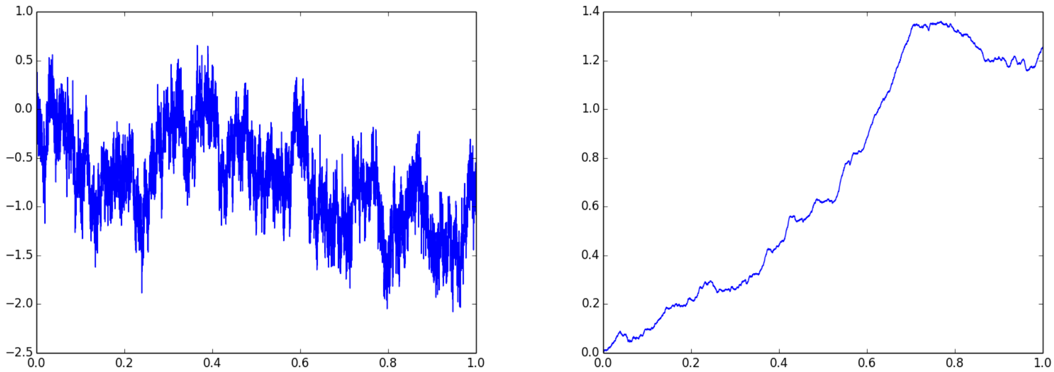


Figure 1.1. Fractional Brownian motion with Hurst index $H = 0.2$ and $H = 0.8$, respectively.

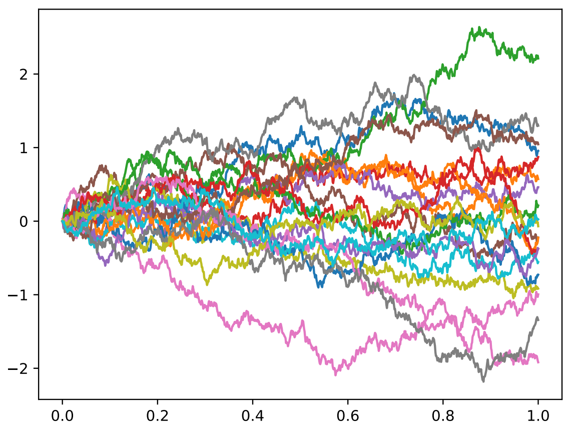
Example 1.6. (pre-Brownian bridge) Let $\mathbb{T}=[0,1]$ and $m(t)=0$ and $\Gamma(s,t)=s\wedge t-st$. This Γ is symmetric and in the exercise below you show that it is positive semi-definite. The centered Gaussian process with covariance function Γ is called the (pre-)*Brownian bridge*, and it looks like a Brownian motion on $[0,1]$, except that at time 1 it ends up in 0 instead of being "free" like the end-point of the Brownian motion on $[0,1]$. To understand this, we plot 20 samples of the Brownian motion and 20 samples of the Brownian bridge and compare the results.

Exercise. Let $(B_t)_{t\geqslant 0}$ be a pre-Brownian motion and let $X_t=B_t-tB_1$, $t\in[0,1]$. Show that X is a pre-Brownian bridge. In particular, Γ is positive semi-definite because it is the covariance function of X.

20 Samples of Brownian motion:

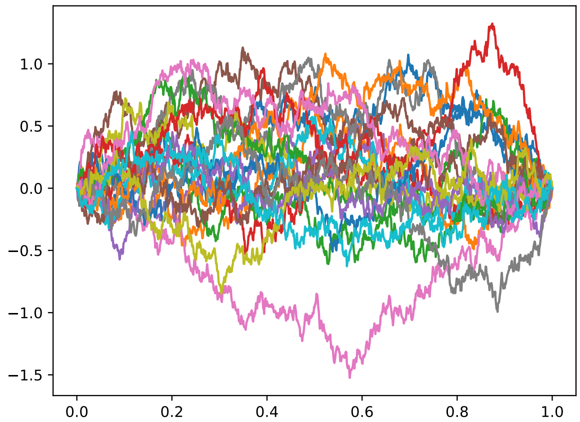
```
Python 3.7.4 [/opt/anaconda3/bin/python3]
Python plugin for TeXmacs.
Please see the documentation in Help -> Plugins -> Python
```

```
>>> import numpy as np
    import matplotlib.pyplot as plt

    T, h = 1, 1e-3
    n = int(T/h)
    k = 20

    time = np.arange(0,T+h,h)
    dB = np.sqrt(h)*(np.random.randn(k,n))
    BM = np.zeros((k,n+1))
    BM[:,1:] = np.cumsum(dB, axis=1)

    plt.clf()

    for i in range(k):
        plt.plot(time,BM[i,:])

    pdf_out(plt.gcf())
```



Next, we plot 20 samples of the Brownian bridge:

```
>>> BB = np.zeros((k,n+1))
    BB[:,1:] = np.cumsum(dB, axis=1) - np.outer(BM[:,n], time[1:])

    plt.clf()

    for i in range(k):
        plt.plot(time,BB[i,:])

    pdf_out(plt.gcf())
```




```
>>>
```

1.2 White noise and Brownian motion

Example 1.7. (Naive white noise) In the introduction we discussed examples where we want to add noise to ordinary differential equations (ODEs). Intuitively, the most natural noise seems to be an i.i.d. family of standard normal variables $(\xi_t)_{t\in\mathbb{R}_+}$, i.e. ξ is a centered Gaussian process with covariance $\Gamma(s,t)=\mathbb{1}_{s=t}$. This would mean that the noise is stationary (it has the same distribution at each time) and what happens at time t is independent of what happens at any other time $s\neq t$. We call this process a *naive white noise*. To understand where the name "white noise" comes from, note that if we plot i.i.d. random variables in the plane rather than on $\mathbb{R}_+$, then the image resembles the static "white noise" that you might remember from old analog televisions; see the next figure.

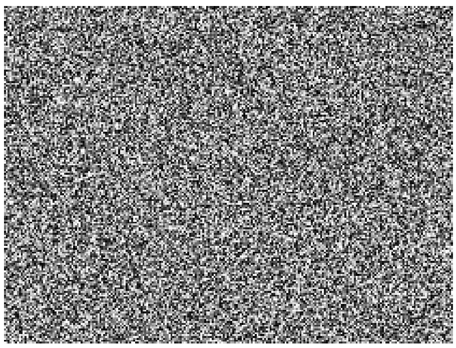


Figure 1.2. 2d naive white noise

Of course, then the question is why static noise should be called white noise. We will discuss this on Sheet 1.

We introduced the naive white noise because we want to consider ODEs perturbed by noise, say

$$\dot{X}_t = b(X_t) + \xi_t, \qquad X_0 = x_0,$$

which in integral form reads as

$$X_t = x_0 + \int_0^t b(X_s)\mathrm{d}s + \int_0^t \xi_s\,\mathrm{d}s.$$

Unfortunately, if ξ is a naive white noise, the map $(\omega, s) \mapsto \xi_s(\omega)$ cannot be jointly measurable and thus the formal expression $\int_0^t \xi_s\,\mathrm{d}s$ might be problematic:

Lemma 1.8. *Let $(\xi_t)_{t\geqslant 0}$ be a naive white noise and let $t>0$. Then the map*

$$\Omega \times [0,t] \ni (\omega, s) \mapsto \xi_s(\omega) \in \mathbb{R}$$

is <u>not</u> measurable with respect to the product σ-algebra $\mathcal{F}\otimes\mathcal{B}([0,t])$, and in particular $\omega \mapsto \int_0^t \xi_s(\omega)\mathrm{d}s$ might not be defined or even if it is defined it might not be a random variable.

Proof. Assume to the contrary that $\xi|_{\Omega\times[0,t]}$ is measurable with respect to $\mathcal{F}\otimes\mathcal{B}([0,t])$. Then also

$$\Omega \times [0,t] \times [0,t] \ni (\omega, s_1, s_2) \mapsto \xi_{s_1}(\omega)\xi_{s_2}(\omega) \in \mathbb{R}$$

is measurable with respect to $\mathcal{F}\otimes\mathcal{B}([0,t])\otimes\mathcal{B}([0,t])$, and for each $r\in[0,t]$:

$$\int_0^r \int_0^r \mathbb{E}[|\xi_{s_1}\xi_{s_2}|]\mathrm{d}s_1\mathrm{d}s_2 < \infty,$$

as $\mathbb{E}[|\xi_{s_1}\xi_{s_2}|]\leqslant 1$ by the Cauchy-Schwarz inequality. Thus, the Fubini-Tonelli theorem shows that for all $r\in[0,t]$

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^r \xi_s \mathrm{d}s\right)^2\right] &= \mathbb{E}\left[\int_0^r \xi_{s_1}\mathrm{d}s_1 \int_0^r \xi_{s_2}\mathrm{d}s_2\right]\\ &= \mathbb{E}\left[\int_0^r\int_0^r \xi_{s_1}\xi_{s_2}\,\mathrm{d}s_1\mathrm{d}s_2\right]\\ &= \int_0^r\int_0^r \mathbb{E}[\xi_{s_1}\xi_{s_2}]\,\mathrm{d}s_1\mathrm{d}s_2\\ &= \int_0^r\int_0^r \mathbb{1}_{s_1=s_2}\,\mathrm{d}s_1\mathrm{d}s_2 = 0.\end{aligned}$$

Therefore, $\int_0^r \xi_s\mathrm{d}s = 0$ almost surely, and since the countable union of null sets is a null set we deduce that almost surely $\int_0^r \xi_s \mathrm{d}s = 0$ for all $r\in[0,t]\cap\mathbb{Q}$.

On the other hand, again by Fubini-Tonelli

$$\mathbb{E}\left[\int_0^t |\xi_s|\,\mathrm{d}s\right] = \int_0^t \mathbb{E}[|\xi_s|]\mathrm{d}s = \sqrt{\frac{2}{\pi}}\,t < \infty,$$

where we used the fact that $\mathbb{E}[|\xi_s|] = \sqrt{2/\pi}$ since $\xi_s \sim \mathcal{N}(0,1)$ (and we don't need to know this precise formula but only that $\mathbb{E}[|X|]\in(0,\infty)$ for a standard normal distribution X). Therefore a.s. $s\mapsto\xi_s(\omega)$ is Lebesgue integrable on $[0,t]$ and by the dominated convergence theorem the map $r\mapsto\int_0^r \xi_s(\omega)\mathrm{d}s$ is continuous on $[0,t]$.

Combined with the above, this implies that almost surely $\int_0^r \xi_s \mathrm{d}s = 0$ for all $r \in [0,t]$. But then a.s. $\xi_s = 0$ for Lebesgue-almost all $s \in [0,t]$, thus $\mathbb{E}[\int_0^t |\xi_s| \mathrm{d}s] = 0$; but this is absurd because we just saw that $\mathbb{E}[\int_0^t |\xi_s| \mathrm{d}s] = t\sqrt{2/\pi}$. Therefore, the assumption must have been incorrect. □

Exercise. Justify the last step of the proof: Show that if $f \in L^1([0,t])$ is such that $\int_0^r f(s)\mathrm{d}s = 0$ for all $r \in [0,t]$, then $f(s) = 0$ for Lebesgue-almost all $s \in [0,t]$.
Hint: Recall Dynkin's $\pi - \lambda$ theorem, cf. Theorem A.8 in Appendix A.2.

The way out of this dilemma is to formally assume that $(\xi_t)_{t \geqslant 0}$ is an i.i.d. family of $\mathcal{N}(0,\infty)$ variables rather than $\mathcal{N}(0,1)$ variables, i.e. that $\mathbb{E}[\xi_t^2] = \infty$. Of course, this makes no sense. But let us abandon mathematical rigor for a moment and argue as physicists. Then we can consider the physicist's Dirac delta function $\delta \colon \mathbb{R} \to \{0,\infty\}$, which satisfies

$$\delta(x) = \begin{cases} \infty, & x = 0, \\ 0, & x \neq 0, \end{cases} \qquad \text{and} \qquad \int_{\mathbb{R}} f(x)\delta(x)\mathrm{d}x = f(0),$$

and we assume that $(\xi_t)_{t \geqslant 0}$ is a centered Gaussian process with covariance function

$$\mathbb{E}[\xi_s \xi_t] = \delta(t - s).$$

This still does not make any sense, but if we assume that we can integrate $\int_0^\infty \xi_t f(t)\mathrm{d}t$ for $f \in L^2(\mathbb{R}_+)$, then we obtain formally

$$\begin{aligned} \mathbb{E}\left[\int_0^\infty \xi_t f(t)\mathrm{d}t \int_0^\infty \xi_s g(s)\mathrm{d}s\right] &= \int_0^\infty \int_0^\infty \mathbb{E}[\xi_t \xi_s] f(t) g(s)\mathrm{d}s\mathrm{d}t \\ &= \int_0^\infty \int_0^\infty \delta(t-s) f(t) g(s)\mathrm{d}s\mathrm{d}t \\ &= \int_0^\infty \left(\int_{-\infty}^t \delta(s) f(t-s)\mathrm{d}s\right) g(t)\mathrm{d}t \\ &= \int_0^\infty f(t) g(t)\mathrm{d}t. \end{aligned}$$

These were only formal manipulations. But now we can take this formal identity and take it as the definition of a (non-naive) white noise. We interpret

$$\xi(f) = \int_0^\infty \xi_t f(t)\mathrm{d}t,$$

where the right hand side is formal notation assuming that ξ has a density, and the left hand side is the "action of ξ on f". This is conceptually similar to formally writing

$$\int_0^\infty f(t)\mu(t)\mathrm{d}t := \mu(f) := \int_0^\infty f(t)\mu(\mathrm{d}t)$$

for a measure μ on $\mathcal{B}(\mathbb{R}_+)$, even if μ does not have a density with respect to Lebesgue measure. Note however that the white noise is not a measure, so even this interpretation is dubious. Instead we abandon the connection with densities and measures, and we define the white noise rigorously as follows:

Definition 1.9. (White noise) *Let* $\mathbb{T} = L^2(\mathbb{R}_+)$ *and*

$$\Gamma(f,g) = \int_0^\infty f(t) g(t)\mathrm{d}t = \langle f, g \rangle_{L^2(\mathbb{R}_+)}.$$

Then Γ *is symmetric and positive semi-definite, and the centered Gaussian process* $(\xi(f))_{f\in L^2(\mathbb{R}_+)}$ *with covariance* Γ *is called* white noise.

Exercise. Show that Γ is indeed positive semi-definite.

Definition 1.10. *Let* (X, d_X) *and* (Y, d_Y) *be metric spaces. A map* $T: X \to Y$ *is called an* isometry *if* $d_Y(T(x), T(x')) = d_X(x, x')$ *for all* $x, x' \in X$.

Lemma 1.11. *If* $(\xi(f))_{f\in L^2(\mathbb{R}_+)}$ *is a white noise, then*

$$L^2(\mathbb{R}_+) \ni f \mapsto \xi(f) \in L^2(\Omega)$$

is a linear isometry.

Proof. Let us write $\langle \cdot, \cdot \rangle = \langle \cdot, \cdot \rangle_{L^2(\mathbb{R}_+)}$. We have for $f, g, h \in L^2(\mathbb{R}_+)$ and $\alpha, \beta \in \mathbb{R}$

$$\mathbb{E}[(\alpha\xi(f) + \beta\xi(g) - \xi(\alpha f + \beta g))\xi(h)] = \alpha\langle f, h\rangle + \beta\langle g, h\rangle - \langle \alpha f + \beta g, h\rangle = 0,$$

so in particular

$$\mathbb{E}[(\alpha\xi(f) + \beta\xi(g) - \xi(\alpha f + \beta g))^2] = 0,$$

i.e. $\alpha\xi(f) + \beta\xi(g) = \xi(\alpha f + \beta g)$ and ξ is linear. Since $\|\xi(f)\|^2_{L^2(\Omega)} = \mathbb{E}[\xi(f)^2] = \|f\|^2_{L^2(\mathbb{R}_+)}$, ξ is an isometry. □

To recap: Our motivation for introducing the naive white noise was that it seems to be natural noise to add to an ODE. But we saw that for a naive white noise ξ we cannot make sense of the ODE

$$X_t = x_0 + \int_0^t b(X_s)\mathrm{d}s + \int_0^t \xi_s \mathrm{d}s,$$

because the integral on the right hand side is not defined or not a random variable. But now we can take a white noise and use formal notation to write

$$\int_0^t \xi_s \mathrm{d}s = \int_0^\infty \mathbb{1}_{[0,t]}(s)\xi_s \mathrm{d}s = \xi(\mathbb{1}_{[0,t]}),$$

and since $\mathbb{1}_{[0,t]} \in L^2(\mathbb{R}_+)$ the right hand side is perfectly well defined.

Exercise. Let ξ be a white noise and define $B_t = \xi(\mathbb{1}_{[0,t]})$, for $t \geqslant 0$. Show that B is a pre-Brownian motion.

——— End of the lecture on October 17 ———

Above we performed the construction of white noise for $(\xi(f))_{f\in H}$ where $H = L^2(\mathbb{R}_+)$ us a Hilbert space. More generally, given *any* Hilbert space H, one can construct a centered Gaussian process $(X(h))_{h\in H}$ with $\mathbb{E}[X(f)X(g)] = \langle f, g\rangle_H$, and also in that case $h \mapsto X(h)$ is a linear isometry; see Exercise Sheet 1 for such a construction. In this case, ξ is called a *white noise on* H.

Recap: having now interpreted $\int_0^t \xi_s \mathrm{d}s$ as $\xi(\mathbb{1}_{[0,t]}) =: B_t$ and having established that the latter is a pre-Brownian motion, we finally end up with the stochastic differential equation (SDE)

$$X_t = x_0 + \int_0^t b(X_s)\mathrm{d}s + B_t.$$

This is still problematic because of bad path properties of the pre-Brownian motion (the map $t \mapsto B_t(\omega)$ might not be measurable), but now we just have to turn the pre-Brownian motion into an actual Brownian motion with continuous trajectories and then we can solve the SDE. We will do this later in the course, for now we discuss the relation between white noise and Brownian motion further.

The previous exercise shows that formally the (pre-)Brownian motion is the integral of the white noise. Conversely, we formally have $\xi_t = \partial_t B_t$, i.e. white noise is the derivative of the (pre-)Brownian motion:

Lemma 1.12. (Wiener integral) *Let $(B_t)_{t \geqslant 0}$ be a pre-Brownian motion and let*

$$\mathcal{E} = \left\{ f \in L^2(\mathbb{R}_+) \colon f(t) = \sum_{k=0}^{n-1} x_k \mathbb{1}_{(t_k, t_{k+1}]}(t), n \in \mathbb{N}, x_0, \ldots, x_{n-1} \in \mathbb{R}, 0 \leqslant t_0 < t_1 < \cdots < t_n \right\}.$$

For such f we define

$$\xi(f) := \int_0^\infty f(s) \mathrm{d}B_s := \sum_{k=0}^{n-1} x_k (B_{t_{k+1}} - B_{t_k}).$$

This definition does not depend on the specific representation of f, and we have

$$\|\xi(f)\|_{L^2(\Omega)}^2 = \mathbb{E}[\xi(f)^2] = \|f\|_{L^2(\mathbb{R}_+)}^2.$$

Therefore, ξ has a unique continuous extension to $L^2(\mathbb{R}_+)$, also denoted by ξ, and we also write

$$\int_0^\infty f(s) \mathrm{d}B_s := \xi(f).$$

The process $(\xi(f))_{f \in L^2(\mathbb{R}_+)}$ is a white noise.

The integral $\int_0^\infty f(s) \mathrm{d}B_s$ is called *Wiener integral* and it is a precursor of the *Itô integral*, which we will construct later in the course and allows random integrands, not just deterministic f like the Wiener integral.

Exercise. Why is $\int_0^\infty f(s) \mathrm{d}B_s := \sum_{k=0}^{n-1} x_k (B_{t_{k+1}} - B_{t_k})$ a sensible definition?

Proof. We leave it as an exercise to check that the definition of $\xi(f)$ does not depend on the representation of f, i.e. that if $\sum_{k=0}^{n-1} x_k \mathbb{1}_{(t_k, t_{k+1}]} = \sum_{\ell=0}^{m-1} y_\ell \mathbb{1}_{(s_k, s_{k+1}]}$, then

$$\sum_{k=0}^{n-1} x_k (B_{t_{k+1}} - B_{t_k}) = \sum_{\ell=0}^{m-1} y_\ell (B_{t_{\ell+1}} - B_{t_\ell}).$$

It is clear from the definition that the map $\mathcal{E} \ni f \mapsto \xi(f)$ is linear; let us show the isometry property. We have

$$\begin{aligned} \|\xi(f)\|_{L^2(\Omega)}^2 &= \mathbb{E}\left[\left(\sum_{k=0}^{n-1} x_k (B_{t_{k+1}} - B_{t_k})\right)^2\right] \\ &= \sum_{k,\ell=0}^{n-1} x_k x_\ell \mathbb{E}[(B_{t_{k+1}} - B_{t_k})(B_{t_{\ell+1}} - B_{t_\ell})]. \end{aligned}$$

If (say) $k < \ell$, then $B_{t_{\ell+1}} - B_{t_\ell}$ is independent of $B_{t_{k+1}} - B_{t_k}$ and the expectation vanishes. Therefore, we remain with the diagonal terms and obtain

$$\|\xi(f)\|_{L^2(\Omega)}^2 = \sum_{k=0}^{n-1} x_k^2 \mathbb{E}[(B_{t_{k+1}} - B_{t_k})^2] = \sum_{k=0}^{n-1} x_k^2 (t_{k+1} - t_k) = \int_0^\infty |f(t)|^2 \mathrm{d}t = \|f\|_{L^2(\mathbb{R}_+)}^2.$$

Therefore,

$$\xi: (\mathcal{E}, \|\cdot\|_{L^2(\mathbb{R}_+)}) \to (L^2(\Omega), \|\cdot\|_{L^2(\Omega)})$$

is a linear isometry, and in particular it is uniformly continuous. As $\mathcal{E}$ is dense in $L^2(\mathbb{R}_+)$, the map ξ has a unique continuous extension to all of $L^2(\mathbb{R}_+)$, which is still a linear isometry and which we still denote by ξ. It remains to show that ξ is a white noise.

By centered Gaussianity of B, the process $(\xi(f))_{f\in\mathcal{E}}$ is centered Gaussian, and by a limiting argument (cf. Lemma A.3 in Appendix A.1) also $(\xi(f))_{f\in L^2(\mathbb{R}_+)}$ is centered Gaussian. By polarization we have

$$\mathbb{E}[\xi(f)\xi(g)] = \langle f, g\rangle_{L^2(\mathbb{R}_+)}, \qquad f, g \in L^2(\mathbb{R}_+),$$

and thus ξ is a white noise.

Polarization: Let X be an $\mathbb{R}$-vector space and let $[\cdot,\cdot]_1, [\cdot,\cdot]_2: X\times X \to \mathbb{R}$ be two symmetric bilinear forms such that $[x,x]_1 = [x,x]_2$ for all $x\in X$. Then $[x,y]_1 = [x,y]_2$ for all $x, y\in X$:

$$\begin{aligned}[x,y]_1 &= \frac{1}{4}([x+y,x+y]_1 - [x-y,x-y]_1)\\ &= \frac{1}{4}([x+y,x+y]_2 - [x-y,x-y]_2)\\ &= [x,y]_2.\end{aligned}$$

□

With formal notation we have

$$\int_0^\infty f(s)\mathrm{d}B_s = \int_0^\infty f(s)\partial_s B_s\,\mathrm{d}s, \qquad \xi(f) = \int_0^\infty f(s)\xi_s\,\mathrm{d}s,$$

and since $\xi(f) = \int_0^\infty f(s)\mathrm{d}B_s$ for all $f\in L^2(\mathbb{R}_+)$ we formally get $\partial_t B = \xi$. One can make this link rigorous with the help of Schwartz's theory of generalized functions, see Exercise Sheet 2 for some details; but we will not need this in the main lectures.

2 Brownian motion and Poisson process

2.1 Continuity of stochastic processes

The pre-Brownian motion is not very useful yet. To turn it into an interesting and useful process, we need to add one more property to its definition:

Definition 2.1. (Continuous stochastic process, (fractional) Brownian motion)

i. *We say that a stochastic process $X = (X_t)_{t\geqslant 0}$ with values in $\mathbb{R}^d$ is* continuous *if all of its trajectories are continuous, i.e. $t\mapsto X_t(\omega)$ is continuous for all $\omega\in\Omega$.*

ii. *A continuous pre-Brownian motion such that $B_0(\omega) = 0$ for all $\omega\in\Omega$ (rather than $B_0 = 0$ a.s.) is called a* Brownian motion *or* Wiener process.

iii. *A continuous fractional pre-Brownian motion such that $B_0(\omega) = 0$ for all $\omega\in\Omega$ (rather than $B_0 = 0$ a.s.) is called a* fractional Brownian motion.

It is natural to ask under which conditions a process X is continuous and if a Brownian motion exists. This turns out to be quite subtle, because it is not possible to find conditions on the finite-dimensional distributions of X which guarantee the continuity of its trajectories:

Example 2.2. Let $(X_t)_{t\geqslant 0}$ be a continuous stochastic process with values in $\mathbb{R}$, and let τ be a random variable which is uniformly distributed on $[0,1]$. Then

$$\tilde{X}_t(\omega) = X_t(\omega) + \mathbb{1}_{\{\tau(\omega)\}}(t), \qquad t \geqslant 0,$$

is discontinuous for all ω and satisfies $\mathbb{P}\left(\tilde{X}_t = X_t\right) = 1$ for all $t \geqslant 0$. In particular, $\tilde{X}$ and X have the same finite-dimensional distributions.

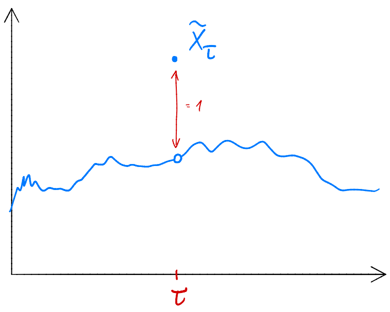


Figure 2.1.

Recall from Stochastics II that the *law* of $(X_t)_{t\geqslant 0}$ is defined as the measure $\mathbb{P}_X = \mathbb{P}\circ X^{-1}$ on $(\mathbb{R}^{\mathbb{R}_+}, \mathcal{B}(\mathbb{R})^{\otimes \mathbb{R}_+})$, and that $\mathbb{P}_X$ is uniquely determined by the finite-dimensional distributions of X, i.e. the family of measures $(\mathbb{P}\circ(X_t)_{t\in I}^{-1})_{I\subset\mathbb{T},|I|<\infty}$. Since X and $\tilde{X}$ have the same finite-dimensional distributions, they also have the same law on $(\mathbb{R}^{\mathbb{R}_+}, \mathcal{B}(\mathbb{R})^{\otimes \mathbb{R}_+})$: $\mathbb{P}_X = \mathbb{P}_{\tilde{X}}$. So we have two processes with the same law, but $X(\omega)$ is continuous for all $\omega \in \Omega$, while $\tilde{X}(\omega)$ is discontinuous for all $\omega \in \Omega$. Consequently, we cannot determine from the law of a process whether it is continuous.

In particular, the set $C(\mathbb{R}_+, \mathbb{R})$ is not in $\mathcal{B}(\mathbb{R})^{\otimes \mathbb{R}_+}$: Otherwise $\mathbb{P}_X(C(\mathbb{R}_+,\mathbb{R})) = \mathbb{P}_{\tilde{X}}(C(\mathbb{R}_+,\mathbb{R}))$ would be defined and this would lead to the contradiction

$$\begin{aligned} 1 &= \mathbb{P}(\Omega) \\ &= \mathbb{P}(X \in C(\mathbb{R}_+,\mathbb{R})) \\ &= \mathbb{P}_X(C(\mathbb{R}_+,\mathbb{R})) \\ &= \mathbb{P}_{\tilde{X}}(C(\mathbb{R}_+,\mathbb{R})) \\ &= \mathbb{P}(\tilde{X} \in C(\mathbb{R}_+,\mathbb{R})) \\ &= \mathbb{P}(\emptyset) = 0. \end{aligned}$$

Therefore, our assumption $C(\mathbb{R}_+,\mathbb{R}) \in \mathcal{B}(\mathbb{R})^{\otimes \mathbb{R}_+}$ must have been wrong. Even worse, a variation of the same argument shows that not even the point set $\{0\}$ is in $\mathcal{B}(\mathbb{R})^{\otimes \mathbb{R}_+}$ (where we write 0 for the function which maps every t to 0).

The problem is that the law of X is defined on $\mathcal{B}(\mathbb{R})^{\otimes \mathbb{R}_+}$, and roughly speaking sets from this σ-algebra only depend on countably many $(X_{t_1}, X_{t_2}, \ldots)$. But to determine whether X is continuous we need to evaluate it at all $t \in \mathbb{R}_+$.

Structure of $\boldsymbol{\mathcal{B}(\mathbb{R})^{\otimes \mathbb{R}_+}}$: The following discussion is irrelevant for our lecture. A subset $A \subset \mathbb{R}^{\mathbb{R}_+}$ is in $\mathcal{B}(\mathbb{R})^{\otimes \mathbb{R}_+}$ if and only if there exists $t_1, t_2, \ldots \in \mathbb{R}_+$ and $B \in \mathcal{B}(\mathbb{R})^{\otimes \mathbb{N}}$ such that

$$A = \{\omega \in \mathbb{R}^{\mathbb{R}_+} : (\omega(t_1), \omega(t_2), \ldots) \in B\}.$$

Proving this amounts to showing that the family of sets of the claimed form is a Dynkin system and also stable by intersection, and then to apply the $\pi - \lambda$ theorem.

Definition 2.3. (Modification, indistinguishable) *Let $X=(X_t)_{t\in\mathbb{T}}$ and $\tilde{X}=(\tilde{X}_t)_{t\in\mathbb{T}}$ be two stochastic processes with values in a measurable state space S. We say that*

i. *$\tilde{X}$ is a* modification *of X if $\mathbb{P}(X_t=\tilde{X}_t)=1$ for all $t\in\mathbb{T}$;*

ii. *X and $\tilde{X}$ are* indistinguishable *if there exists a measurable set $A\in\mathcal{F}$ with $\mathbb{P}(A)=1$ and such that $X_t(\omega)=\tilde{X}_t(\omega)$ for all $\omega\in A$ and all $t\in\mathbb{T}$. Formally, we also write $\mathbb{P}(X_t=\tilde{X}_t \text{ for all } t\in\mathbb{T})=1$.*

Note that $\mathbb{P}(X_t=\tilde{X}_t$ for all $t\in\mathbb{T})$ might in general not be defined, because

$$\{X_t=\tilde{X}_t \text{ for all } t\in\mathbb{T}\}=\bigcap_{t\in\mathbb{T}}\{X_t=\tilde{X}_t\}$$

is an intersection of uncountably many events. Therefore, we require the existence of the measurable set $A\in\mathcal{F}$ in ii.

The second property is much stronger than the first one. For example, if X is a continuous process and X and $\tilde{X}$ are indistinguishable, then $\tilde{X}$ is almost surely continuous. While Example 2.2 shows that a continuous process can have a discontinuous modification.

——— End of the lecture on October 23 ———

Exercise. Let $(\tilde{X}_t)_{t\geqslant 0}$ be a modification of $(X_t)_{t\geqslant 0}$. Show that:

i. X and $\tilde{X}$ have the same finite dimensional distributions and therefore the same law.

ii. If X and $\tilde{X}$ take values in a metric space and are both continuous, then they are indistinguishable.

There are essentially two ways to solve these problems and to construct continuous processes:

– Either we construct the process of interest X on a different probability space than $(\mathbb{R}^{\mathbb{R}_+},\mathcal{B}(\mathbb{R})^{\otimes\mathbb{R}_+})$, for example on $(C(\mathbb{R}_+,\mathbb{R}),\mathcal{B}(C(\mathbb{R}_+,\mathbb{R})))$ (say via Donsker's invariance principle for the Brownian motion).

– Or we use the Kolmogorov extension problem to construct a process $\tilde{X}$ on $(\mathbb{R}^{\mathbb{R}_+},\mathcal{B}(\mathbb{R})^{\otimes\mathbb{R}_+})$ which has all the prescribed finite dimensional distributions that we want, and then try to construct a continuous modification X of $\tilde{X}$ (so in particular X has the required law).

Of course, there are also probability laws on $(\mathbb{R}^{\mathbb{R}_+},\mathcal{B}(\mathbb{R})^{\otimes\mathbb{R}_+})$ for which the associated process can never be continuous, for example the (deterministic) process $X_t=\mathbb{1}_{[1,\infty)}(t)$, or the Poisson process that we will encounter later. But in many cases of interest, most notably for the pre-Brownian motion, one or both of these approaches can be used to construct a continuous process with the given law. Here we will follow the second approach.

Definition 2.4. (Hölder continuity) *For $\alpha\in(0,1]$ and $T\in(0,+\infty)$, the space of α-*Hölder continuous *functions on $[0,T]$ is defined as*

$$C^\alpha([0,T],\mathbb{R})=\{f\colon[0,T]\to\mathbb{R}, \|f\|_\alpha<\infty\},\qquad \textit{where } \|f\|_\alpha:=\sup_{0\leq s<t\leq T}\frac{|f(t)-f(s)|}{|t-s|^\alpha}.$$

In case of ambiguity of the time interval, we also write more explicitly

$$\|f\|_{C^\alpha([0,T])}:=\sup_{0\leq s<t\leq T}\frac{|f(t)-f(s)|}{|t-s|^\alpha}$$

Exercise.

i. Do you know another name for 1-Hölder continuous functions?

ii. Show that for $\beta \leqslant \alpha$ we have $C^\alpha([0,T],\mathbb{R}) \subset C^\beta([0,T],\mathbb{R})$.

One of the most important tools for constructing continuous stochastic processes is Kolmogorov's continuity criterion.

Theorem 2.5. (Kolmogorov's continuity criterion) *Let $T \in (0,+\infty)$ and let $(X_t)_{t\in[0,T]}$ be a real-valued stochastic process such that there exist $p \in [1,\infty)$, $\alpha > \frac{1}{p}$ and $K \geqslant 0$ with*

$$\mathbb{E}[|X_t - X_s|^p]^{1/p} \leqslant K|t-s|^\alpha. \tag{2.1}$$

Then there exists a continuous modification $\tilde{X}$ of X. Moreover, for all $\beta \in (0, \alpha - \frac{1}{p})$ there exists a constant $C = C(\alpha,\beta,p,T) > 0$ such that

$$\mathbb{E}[\|\tilde{X}\|_\beta^p]^{1/p} \leqslant CK. \tag{2.2}$$

In particular, $\tilde{X}$ is a.s. β-Hölder continuous.

Let us postpone the proof of Theorem 2.5 and first present some applications, to show its power. Armed with it, we can finally construct the Brownian motion.

Corollary 2.6. *The Brownian motion $B = (B_t)_{t\geq 0}$ exists and $(B_t)_{t\in[0,T]}$ is almost surely in $C^\alpha([0,T],\mathbb{R})$ whenever $T \in (0,\infty)$ and $\alpha < 1/2$. We have*

$$\mathbb{E}[\|B\|_{C^\alpha([0,T])}^p] < \infty \qquad \forall p \in [1,\infty).$$

Proof. Let $(\tilde{B}_t)_{t\geqslant 0}$ be a pre-Brownian motion. Since B is a centered Gaussian, all its p-moments scale in the same way and so for $p > 0$ we have

$$\mathbb{E}[|\tilde{B}_t - \tilde{B}_s|^p]^{1/p} = \left(c_p\,\mathbb{E}[|\tilde{B}_t - \tilde{B}_s|^2]^{\frac{p}{2}}\right)^{1/p} = \left(c_p\,|t-s|^{\frac{p}{2}}\right)^{1/p} = c_p^{1/p}|t-s|^{\frac{1}{2}}.$$

It follows from Corollary 2.7 below that $(\tilde{B}_t)_{t\geqslant 0}$ has a continuous modification $(B_t)_{t\geqslant 0}$; so here let us focus on proving the statement (2.6). Applying Kolmogorov's continuity criterion to $(B_t)_{t\in[0,T]}$ for $\alpha < \frac{1}{2}$ and $\frac{1}{p} < \frac{1}{2} - \alpha$, we find

$$\mathbb{E}[\|B\|_{C^\alpha([0,T])}^p] < \infty.$$

The claim was that this is true for all $p \geqslant 1$, and the above shows it for p large enough, i.e. $p > \left(\frac{1}{2} - \alpha\right)^{-1}$; instead for $p \in [1, \left(\frac{1}{2} - \alpha\right)^{-1}]$ we can find $n \in \mathbb{N}$ such that $n > \left(\frac{1}{2} - \alpha\right)^{-1} \geqslant p$ and then bound it using Jensen's inequality:

$$\mathbb{E}[\|B\|_{C^\alpha([0,T])}^p]^{\frac{1}{p}} \leqslant \mathbb{E}[\|B\|_{C^\alpha([0,T])}^n]^{\frac{1}{n}} < \infty. \qquad \square$$

Exercise. How big do we need to choose p in the previous argument to at least be able to apply Kolmogorov's continuity criterion? Does $p = 2$ work? Which Hölder continuity would we get with $p = 2 + \varepsilon$?

Corollary 2.7. *Let $(X_t)_{t\geqslant 0}$ be a real-valued stochastic process; suppose that there exist $p \in [1,\infty)$, $\alpha > \frac{1}{p}$ and $K \geqslant 0$ such that*

$$\mathbb{E}[|X_t - X_s|^p]^{1/p} \leqslant K|t-s|^\alpha \quad \forall s \leqslant t.$$

Then there exists a continuous modification $\tilde{X}$ of X.

Proof. The difference to the formulation of Theorem 2.5 is that now X is indexed by $\mathbb{R}_+$, and not by a compact interval $[0,T]$. But we can apply for each $n \in \mathbb{N}$ Kolmogorov's continuity criterion to obtain a continuous modification $(\tilde{X}_t^{(n)})_{t\in[0,n]}$ of $(X_t)_{t\in[0,n]}$. We would like to define $\tilde{X}_t = \tilde{X}_t^{(n)}$ for $n \geqslant t$. The problem is that there are infinitely many possible choices for $n \geqslant t$, so we have to justify that this definition does not depend on n and that it leads to a continuous process.

If $m \geqslant n$, then $(\tilde{X}_t^{(n)})_{t\in[0,n]}$ and $(\tilde{X}_t^{(m)})_{t\in[0,n]}$ are both modifications of $(X_t)_{t\in[0,n]}$ and they are both continuous, so they are indistinguishable. Since the countable union of null sets is a null set, we obtain $\mathbb{P}(N) = 0$ for

$$N = \left\{\omega \in \Omega \colon \exists m, n \in \mathbb{N} \text{ s.t. } n \leqslant m \text{ and } \tilde{X}_t^{(n)}(\omega) \neq \tilde{X}_t^{(m)} \text{ for some } t \leqslant n\right\}.$$

We then define for $t \leqslant n$:

$$\tilde{X}_t(\omega) = \begin{cases} \tilde{X}_t^{(n)}(\omega), & \omega \in N^c, \\ 0, & \omega \in N. \end{cases}$$

Since N is a null set, $\tilde{X}$ is a modification of X, and it is trivially continuous for $\omega \in N$. And since for $\omega \in N^c$ we have

$$\tilde{X}_t(\omega) = \tilde{X}_t^{(n)}(\omega) = \tilde{X}_t^{(m)}(\omega)$$

for all $n, m \geqslant t$ and all the $\left(\tilde{X}_t^{(k)}\right)_t$ are continuous, we get that $\tilde{X}(\omega)$ is continuous. □

We can finally present the

Proof of Theorem 2.5. We use the notation

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

1. By rescaling time $t \to T \cdot t$, we may assume without loss of generality that $T = 1$ (convince yourself of this! See the blue exercise later).
2. Assume that we already showed for a dense subset $D \subset [0,1]$ that:

$$\mathbb{E}\left[\left(\sup_{s\neq t\in D} \frac{|X_{s,t}|}{|t-s|^\beta}\right)^p\right]^{1/p} \leqslant CK. \tag{2.3}$$

   Then in particular $\sup_{s\neq t\in D} \frac{|X_{s,t}(\omega)|}{|t-s|^\beta} < \infty$ for almost all ω, and for such ω the function $X(\omega)$ is uniformly continuous on the dense subset $D \subset [0,1]$. Therefore, it has a unique continuous extension to a function $\tilde{X}(\omega)$ on $[0,1]$, which satisfies

$$\sup_{s\neq t\in[0,1]} \frac{|\tilde{X}_{s,t}(\omega)|}{|t-s|^\beta} = \sup_{s\neq t\in D} \frac{|X_{s,t}(\omega)|}{|t-s|^\beta}.$$

   If ω is in the null set for which $\sup_{s\neq t\in D} \frac{|X_{s,t}(\omega)|}{|t-s|^\beta} = \infty$, we simply define $\tilde{X}_t(\omega) = 0$ for all $t \in [0,1]$. Then $\tilde{X}$ is continuous and it satisfies (2.2), but we still have to show that it is a modification of X.

   For $t \in D$ we have a.s. $X_t = \tilde{X}_t$ by construction. For $t \notin D$ consider a sequence $(t_n) \subset D$ with $t_n \to t$. Then $\tilde{X}_{t_n}(\omega) \to \tilde{X}_t(\omega)$ for all ω by continuity of $\tilde{X}$, and $X_{t_n} \to X_t$ in L^p because $\mathbb{E}[|X_{t_n} - X_t|^p]^{1/p} \leqslant K|t - t_n|^\alpha$. Therefore, the sequence $(\tilde{X}_{t_n} = X_{t_n})_n$ converges a.s. to $\tilde{X}_t$ and it converges in L^p to X_t, thus $\tilde{X}_t = X_t$ a.s. and $\tilde{X}$ is indeed a modification of X.

3. It remains to show (2.3). For $n \in \mathbb{N}_0$, consider the *dyadic times*

$$D_n := \{t_k^n := k2^{-n}, 0 \leqslant k \leqslant 2^n\}, \qquad D := \bigcup_{n=0}^{\infty} D_n,$$

and let

$$\Delta_n = \{(t_k^n, t_{k+1}^n) \colon 0 \leqslant k \leqslant 2^n - 1\}$$

be the nearest neighbors in D_n. Then D is dense and it suffices to show (2.3) for this D. Let for $n \in \mathbb{N}_0$:

$$M_n := \max_{k=0,\ldots,2^n-1} |X_{t_k^n, t_{k+1}^n}| = \max_{(s,t) \in \Delta_n} |X_{s,t}|.$$

Then

$$\begin{aligned} \mathbb{E}[M_n^p]^{1/p} &= \mathbb{E}\Big[\max_{k=0,\ldots,2^n-1} |X_{t_k^n, t_{k+1}^n}|^p\Big]^{1/p} \\ &\leqslant \mathbb{E}\Bigg[\sum_{k=0}^{2^n-1} |X_{t_k^n, t_{k+1}^n}|^p\Bigg]^{1/p} = \Bigg(\sum_{k=0}^{2^n-1} \mathbb{E}[|X_{t_k^n, t_{k+1}^n}|^p]\Bigg)^{1/p} \\ &\leqslant 2^{n/p} K 2^{-n\alpha} = K 2^{-n\left(\alpha - \frac{1}{p}\right)}. \end{aligned} \tag{2.4}$$

This bound would suffice if we only wanted to compare $s, t \in \bigcup_n \Delta_n$. But we also have to treat the case $s = t_k^n$ and $t = t_\ell^m$ for arbitrary m, n and k, ℓ. We claim that

$$\sup_{s \neq t \in D} \frac{|X_{s,t}|}{|t - s|^\beta} \leqslant 2^{\beta+1} M, \tag{2.5}$$

where $M := \sum_{n=0}^{\infty} 2^{n\beta} M_n \in [0, \infty]$. If this is the case, then by the triangle inequality for the L^p-norm ("Minkowski's inequality"), we have:

$$\begin{aligned} \mathbb{E}[M^p]^{1/p} &= \|M\|_{L^p} \leqslant \sum_{n=0}^{\infty} \|2^{n\beta} M_n\|_{L^p} \\ &\overset{(2.4)}{\leqslant} \sum_{n=0}^{\infty} 2^{n\beta} K 2^{-n\left(\alpha - \frac{1}{p}\right)} = K \sum_{n=0}^{\infty} 2^{n\left(\beta - \left(\alpha - \frac{1}{p}\right)\right)} = K\tilde{C}, \end{aligned}$$

where $\tilde{C} = \sum_{n=0}^{\infty} 2^{n\left(\beta - \left(\alpha - \frac{1}{p}\right)\right)} < \infty$ because $\beta < \alpha - \frac{1}{p}$, so that the geometric series converges. Combining the above bound on $\|M\|_{L^p(\Omega)}$ with claim (2.5) thus yields the conclusion (2.2).

——— End of the lecture on October 24 ———

4. To prove the claim (2.5), we use a *chaining argument*: Define for $s < t \in D$:

$$s_n := \min\{r \in D_n \colon r \geqslant s\}, \qquad t_n := \max\{r \in D_n \colon r \leqslant t\}.$$

Since $s, t \in D = \bigcup_m D_m$ we have $s_n = s$ and $t_n = t$ from some n on.

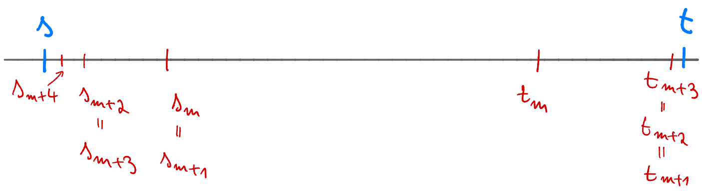


Figure 2.2. Illustration of the s_n and t_n.

Consider now $m \in \mathbb{N}_0$ such that $2^{-m-1} < t - s \leqslant 2^{-m}$. Then

$$\begin{aligned} |X_t - X_s| &\leqslant |X_t - X_{t_m}| + |X_{t_m} - X_{s_m}| + |X_{s_m} - X_s| \\ &\leqslant \sum_{n=m}^{\infty} |X_{t_{n+1}} - X_{t_n}| + |X_{t_m} - X_{s_m}| + \sum_{n=m}^{\infty} |X_{s_n} - X_{s_{n+1}}|, \end{aligned}$$

where the two series on the right hand side are actually finite sums. Since $2^{-m} \geqslant t - s$ we know that either $s_m = t_m$ or $(s_m, t_m) \in \Delta_m$. Moreover,

$$s_n - s_{n+1} \leqslant s_n - s < 2^{-n}$$

and $s_n, s_{n+1} \in D_{n+1}$, so either $s_n = s_{n+1}$ or $(s_{n+1}, s_n) \in \Delta_{n+1}$. Similarly for (t_n, t_{n+1}). Therefore, we can estimate

$$\begin{aligned} \frac{|X_{s,t}|}{|t-s|^{\beta}} &\leqslant |t-s|^{-\beta} \left(2 \sum_{n=m}^{\infty} \max_{(u,v) \in \Delta_{n+1}} |X_{u,v}| + \max_{(u,v) \in \Delta_m} |X_{u,v}| \right) \\ &\leqslant (2^{-m-1})^{-\beta} 2 \sum_{n=m}^{\infty} M_n \leqslant 2^{\beta+1} \sum_{n=m}^{\infty} 2^{n\beta} M_n \leqslant 2^{\beta+1} M, \end{aligned}$$

where $M = \sum_{n=0}^{\infty} 2^{n\beta} M_n$. This concludes the proof. □

Exercise.

i. Deduce that: If $(X_t)_{t \in [0,T]}$ is such that $\mathbb{E}[|X_t - X_s|^p] \leqslant K |t-s|^{\gamma}$ for some $p \geqslant 1$ and $\gamma > 1$, then X has a continuous modification. On Sheet 2 we will see that this is false for $\gamma = 1$.

ii. Convince yourself that the same proof works if X takes values in a complete metric space (important special case: Banach space). But completeness is important: where did we use the fact that $\mathbb{R}$ is complete?

iii. Justify Step 1 of the proof. In fact, apply scaling to get a more precise statement: if X satisfies (2.1) on $[0,T]$, then for all $\beta \in (0, \alpha - \frac{1}{p})$ there exists a constant $\tilde{C} = \tilde{C}(\alpha, \beta, p) > 0$ such that

$$\mathbb{E}[\|\tilde{X}\|_{\beta}^{p}]^{1/p} \leqslant \tilde{C} K T^{\alpha - \beta}$$

where now the dependence on T is explicit. Note that the r.h.s. explodes as $T \to \infty$. Can you reconstruct the exact expression of $\tilde{C}(\alpha, \beta, p)$ from the proof?

Lemma 1.4 is analogously true for Brownian motion: a *continuous* stochastic process is a Brownian motion if and only if the conditions i.-iii. in the lemma are satisfied.

Remark 2.8. The Brownian motion is only Hölder-continuous on compact intervals, but not on $\mathbb{R}_+$, i.e. we a.s. have $\sup_{0 \leqslant s < t < \infty} \frac{|B_t - B_s|}{|t-s|^{\alpha}} = \infty$ for all $\alpha \in \mathbb{R}$. You will show this on Sheet 2.

Remark 2.9.

i. Recall from Probability Theory II that the Borel σ-algebra on $C(\mathbb{R}_+, \mathbb{R}) = C(\mathbb{R}_+)$, equipped with the topology of locally uniform convergence, is given by

$$\mathcal{B}(C(\mathbb{R}_+)) = \mathcal{B}(\mathbb{R})^{\otimes \mathbb{R}_+} \cap C(\mathbb{R}_+) = \{A \cap C(\mathbb{R}_+) : A \in \mathcal{B}(\mathbb{R})^{\otimes \mathbb{R}_+}\}.$$

Therefore, a map $X:\Omega\to C(\mathbb{R}_+)$ is $\mathcal{F}-\mathcal{B}(C(\mathbb{R}_+,\mathbb{R}))$-measurable if and only if X_t: $\Omega\to\mathbb{R}$ is $\mathcal{F}-\mathcal{B}(\mathbb{R})$-measurable for all $t\geqslant 0$. In other words, X is a random variable taking values in $C(\mathbb{R}_+)$ if and only if $(X_t)_{t\geqslant 0}$ is a continuous stochastic process. Therefore, any continuous stochastic process induces a probability measure $\mathbb{P}_X$ on $(C(\mathbb{R}_+),\mathcal{B}(C(\mathbb{R}_+)))$ via $\mathbb{P}_X(A)=\mathbb{P}(X\in A)$ (i.e. $\mathbb{P}_X$ is the distribution of X).

ii. The distribution $\mathbb{P}_B$ of the Brownian motion is often called the *Wiener measure* on $(C(\mathbb{R}_+),\mathcal{B}(C(\mathbb{R}_+)))$; the probability space $(C(\mathbb{R}_+),\mathcal{B}(C(\mathbb{R}_+)),\mathbb{P}_B)$ is called the *Wiener space*.

In fact, we can extend the discussion in the previous remark to show that any continuous stochastic process can be realized on the space $(C(\mathbb{R}_+),\mathcal{B}(C(\mathbb{R}_+)))$ in a *canonical* way. To this end, we need to introduce some notation. Given $\omega\in C(\mathbb{R}_+)$ and $t\geqslant 0$, we define the *evaluation map* $e_t: C(\mathbb{R}_+)\to\mathbb{R}$ as $e_t(f)=f(t)$.

(In the lecture, the next statement was presented in a slightly more informal way as part of the discussion in the previous remark)

Lemma 2.10. *Let X be a continuous real valued stochastic process, $\mathbb{P}_X$ be its law. On the probability space $(C(\mathbb{R}_+),\mathcal{B}(C(\mathbb{R}_+)),\mathbb{P}_X)$, consider the collection $(e_t)_{t\geqslant 0}$. Then $(e_t)_{t\geqslant 0}$ is a continuous stochastic process, with law $\mathbb{P}_X$.*

Proof. It's easy to see that, for each t, e_t is a linear, continuous map; thus in particular is is measurable from $(C(\mathbb{R}_+),\mathcal{B}(C(\mathbb{R}_+)))$ to $(\mathbb{R},\mathcal{B}(\mathbb{R}))$, namely it is a random variable, so that $(e_t)_{t\geqslant 0}$ is a stochastic process. For any $f\in C(\mathbb{R}_+)$, the map $t\mapsto e_t(f)=f(t)$ is none other than f itself, which is continuous by definition, so $(e_t)_{t\geqslant 0}$ is a continuous stochastic process. Now consider $n\in\mathbb{N}$, $(t_1,\ldots,t_n)\in\mathbb{R}_+^n$, $A_i\in\mathcal{B}(\mathbb{R})$ for $i=1,\ldots n$, and the subset of $\mathbb{R}^{\mathbb{R}_+}$ given by $\Gamma=\{f\in\mathbb{R}^{\mathbb{R}_+}: f(t_i)\in A_i\}$. Then by construction

$$\begin{aligned}\mathbb{P}_X(f\in C(\mathbb{R}_+): e_{t_i}(f)\in A_i \text{ for } i=1,\ldots,n) &=\mathbb{P}_X((e_t)_{t\geqslant 0}\in\Gamma\cap C(\mathbb{R}_+))\\ &=\mathbb{P}(X\in\Gamma\cap C(\mathbb{R}_+))\\ &=\mathbb{P}(X_{t_i}\in A_i \text{ for } i=1,\ldots n)\end{aligned}$$

which shows that the finite dimensional distributions of $(e_t)_{t\geqslant 0}$ under $\mathbb{P}_X$ coincide with the finite dimensional distributions of X under $\mathbb{P}$. In particular, $(e_t)_{t\geqslant 0}$ and X (which are possibly defined on different probability spaces!) have the same law. ☐

In probability, quite often we do not really care about the underlying probability space $(\Omega,\mathcal{F},\mathbb{P})$, which is treated as an abstract object. The above result however tells us that, in the case of continuous stochastic processes, if needed we can make it very explicit: we can take $\Omega=C(\mathbb{R}_+)$, with the associated Borel σ-algebra, and $\mathbb{P}$ to be the law of the process itself. This is sometimes referred to as the *canonical representation* of the process X. Notice that in this case Ω becomes a complete separable metric space and a vector space, so it has a very nice structure.

For a Brownian motion B and $n\in\mathbb{N}$, we can consider the piecewise linear dyadic approximation $B^{(n)}$, which interpolates B linearly between the points $t_k^n:=k2^{-n}$, for $k\in\mathbb{N}_0$; namely

$$B_t^{(n)}:=\sum_{k=0}^{\infty}\mathbb{1}_{[t_k^n,t_{k+1}^n)}(t)\,(B_{t_k^n}+2^n(t-t_k^n)B_{t_k^n,t_{k+1}^n}).$$

By continuity of $B^{(n)}$ it is clear that $\sup_{t\leqslant T}|B_t^{(n)}(\omega)-B_t(\omega)|$ converges to 0 for all $\omega\in\Omega$ and all $T\in(0,+\infty)$. It is also not difficult to show that for $f\in L^2(\mathbb{R}_+)$ we have

$$\int_0^\infty f(t)\mathrm{d}B_t=\lim_{n\to\infty}\int_0^\infty f(t)\mathrm{d}B_t^{(n)}=\lim_{n\to\infty}\int_0^\infty f(t)\partial_t B_t^{(n)}\mathrm{d}t=:\lim_{n\to\infty}\int_0^\infty f(t)\xi_t^{(n)}\mathrm{d}t,$$

where the left hand side is the Wiener integral and $\xi_t^{(n)} := \partial_t B_t^{(n)}$ for all t (this is well defined for all $t \notin \{t_k^n : k \in \mathbb{N}_0\}$, and in t_k^n we could for example take the right derivative), and the convergence is in $L^2(\Omega)$. Since $\int_0^\infty f(t) \mathrm{d}B_t = \xi(f)$ for a white noise ξ, we get that $\xi(f) = \lim_{n\to\infty} \int_0^\infty f(t) \xi_t^{(n)} \mathrm{d}t$, so in a sense we can intepret ξ as limit of $\xi^{(n)}$. Moreover, by construction

$$\xi_t^{(n)} = \sum_{k=0}^{\infty} \mathbb{1}_{[t_k^n, t_{k+1}^n)}(t) 2^n B_{t_k^n, t_{k+1}^n}.$$

By independence of the Brownian increments we get that $\xi^{(n)}|_{[t_k^n, t_{k+1}^n)}$ and $\xi^{(n)}|_{[t_\ell^n, t_{\ell+1}^n)}$ are independent for $k \neq \ell$. Moreover, $\mathbb{E}[\xi_t^{(n)}] = 0$ and $\mathrm{Var}(\xi_t^{(n)}) = 2^n$ for all $t \geqslant 0$. So by letting $n \to \infty$ we formally obtain indeed that $(\xi_t)_{t\geqslant 0}$ are independent and identically distributed Gaussian random variables with infinite variance. Of course, this is again purely formal and just intended to guide your intuition.

2.2 Some path properties of the Brownian motion

So far we showed that the Brownian motion exists and it is almost surely α-Hölder continuous on compact subintervals of $\mathbb{R}_+$ whenever $\alpha < 1/2$. Our next aim is to understand its trajectories better. For example a priori it is not clear whether the Brownian motion can be more regular than α-Hölder continuous. Although in simulations it looks very rough, so we would not expect it to be a C^1 function, and indeed we will show that it is not.

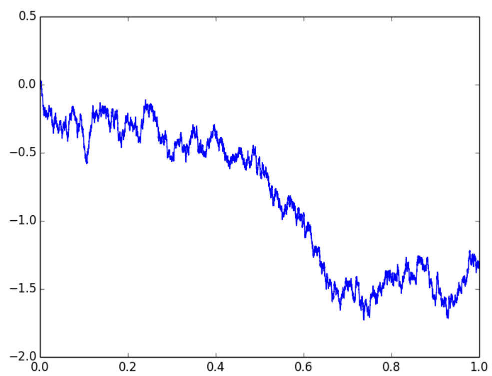


Figure 2.3. Sample path of a Brownian motion.

Throughout this section we fix a Brownian motion B. Let us start by showing the invariance of the law of B under certain path transformations:

Proposition 2.11.

i. $(-B_t)_{t\geqslant 0}$ *is a Brownian motion;*

ii. more generally, for any $\lambda \in \mathbb{R} \setminus \{0\}$, $(\lambda^{-1} B_{\lambda^2 t})_{t\geqslant 0}$ *is a Brownian motion;*

iii. $(B_{t+s} - B_s)_{t\geqslant 0}$ *for* $s \geqslant 0$ *is a Brownian motion, and is independent of* $(B_r)_{r\in[0,s]}$ *(*"Markov property"*);*

iv. $(t \cdot B_{1/t})_{t\geqslant 0}$, *where we set* $0 \cdot B_{1/0} := 0$, *is indistinguishable from a Brownian motion.*

Proof. *i.*, *ii.*, *iii.* were shown on Sheet 1.

iv.: The process $(t\cdot B_{1/t})_{t\geqslant 0}$ is Gaussian, continuous everywhere except possibly at 0, and for $0<s<t$ we have

$$\mathbb{E}\left[(s\cdot B_{1/s})(t\cdot B_{1/t})\right]=s\cdot t(1/t)=s=s\wedge t.$$

It remains to show the (almost sure) continuity at 0. For that purpose, note that by continuity of $t\mapsto tB_{1/t}$ on $(0,1]$, it holds

$$\left\{\omega\colon \lim_{t\to 0} t\cdot B_{1/t}(\omega)=0\right\}=\bigcap_{m\in\mathbb{N}}\bigcup_{n\in\mathbb{N}}\{\omega\colon |t\cdot B_{1/t}(\omega)|\leqslant 1/m \text{ for all } t\in\mathbb{Q}\cap(0,1/n]\}.$$

The event on the right hand side depends only on $(t\cdot B_{1/t})_{t\in\mathbb{Q}\cap(0,1]}$, and this process has the same law as $(B_t)_{t\in\mathbb{Q}\cap(0,1]}$ (both are centered Gaussian processes with the same covariance). Therefore

$$\begin{aligned}\mathbb{P}\left(\lim_{t\to 0} t\cdot B_{1/t}=0\right)&=\mathbb{P}\left(\bigcap_{m\in\mathbb{N}}\bigcup_{n\in\mathbb{N}}\{|t\cdot B_{1/t}|\leqslant 1/m \text{ for all } t\in\mathbb{Q}\cap(0,1/n]\}\right)\\&=\mathbb{P}\left(\bigcap_{m\in\mathbb{N}}\bigcup_{n\in\mathbb{N}}\{|B_t|\leqslant 1/m \text{ for all } t\in\mathbb{Q}\cap(0,1/n]\}\right)\\&=\mathbb{P}\left(\lim_{t\to 0} B_t=0\right)=1,\end{aligned}$$

and the proof is complete. □

Theorem 2.12. *With probability* 1 *there exists no* $t\in[0,+\infty)$ *at which* B *is differentiable.*

Proof. The countable union of null sets is a null set and $(B_{n+t}-B_n)_{t\in[0,1]}$ is a Brownian motion restricted to $[0,1]$, so it suffices to show that almost surely $(B_t)_{t\in[0,1]}$ is nowhere differentiable. If B is differentiable at $t\in[0,1]$, then there exists a constant $\tilde{C}>0$ such that

$$|B_{t+h}-B_t|\leqslant \tilde{C}h,\qquad \forall h\in[0,1]. \tag{2.6}$$

A priori, $\tilde{C}$ could be random; however, if we show that

$$\mathbb{P}(\omega\in\Omega\colon \exists t\in[0,1]\colon |B_{t+h}-B_t|\leqslant Ch\quad \forall h\in[0,1])=0 \tag{2.7}$$

for all deterministic, arbitrarily large constants $C>0$, then the same must hold for random but finite $\tilde{C}>0$ as well. In particular, to conclude it suffices to show (2.7) and from now on we can assume C to be deterministic and fixed. Define the event

$$\Gamma=\{\omega\in\Omega\colon \exists t\in[0,1]\colon |B_{t+h}-B_t|\leqslant Ch\quad \forall h\in[0,1]\}.$$

Notice that, by the order of quantifiers, t here is allowed to be random, i.e. depend on the fixed realization ω we are looking at; so it's hard to manipulate Γ directly, and we want to compare it to “simpler” events.

Suppose $t\in[0,1]$ is such that (2.6) hold; let $n\in\mathbb{N}$, and let $0\leqslant k<2^n$ be such that $t\in[k2^{-n},(k+1)2^{-n}]$. Then for all $1\leqslant j<2^n$ we have

$$\begin{aligned}|B_{(k+j+1)2^{-n}}-B_{(k+j)2^{-n}}|&\leqslant|B_{(k+j+1)2^{-n}}-B_t|+|B_t-B_{(k+j)2^{-n}}|\\&\leqslant C((k+j+1)2^{-n}-k2^{-n})+C((k+j)2^{-n}-k2^{-n})\\&\leqslant C(2j+1)2^{-n}.\end{aligned}$$

Let us define the events

$$\Omega_{n,k}=\bigcap_{j=1,2,3}\{|B_{(k+j+1)2^{-n}}-B_{(k+j)2^{-n}}|\leqslant C(2j+1)2^{-n}\};$$

then the previous argument in fact shows that, for any $n \in \mathbb{N}$,

$$\Gamma \subset \bigcup_{k=0}^{2^n-1} \Omega_{n,k}.$$

Therefore to show that $\mathbb{P}(\Gamma) = 0$, it suffices to estimate the probability of the set on the r.h.s., and show that it becomes infinitesimal as $n \to \infty$.

——— End of the lecture on October 30 ———

By the independence and scaling properties of B

$$\begin{aligned}\mathbb{P}(\Omega_{n,k}) &= \prod_{j=1}^{3} \mathbb{P}\left(|B_{(k+j+1)2^{-n}} - B_{(k+j)2^{-n}}| \leqslant C(2j+1)\, 2^{-n}\right)\\ &= \prod_{j=1}^{3} \mathbb{P}\left(|B_1| \leqslant C(2j+1) 2^{-n/2}\right)\\ &\leqslant \mathbb{P}(|B_1| \leqslant C \cdot 7 \cdot 2^{-n/2})^3\\ &\leqslant (C \cdot 7 \cdot 2^{-n/2})^3,\end{aligned}$$

where in the last step we used that the density of the standard normal distribution is bounded by $1/\sqrt{2\pi} \leqslant 1/2$. Thus,

$$\begin{aligned}\mathbb{P}(\Gamma) &\leqslant \mathbb{P}\left(\bigcup_{k=0}^{2^n-1} \Omega_{n,k}\right)\\ &\leqslant \sum_{k=0}^{2^n-1} \mathbb{P}(\Omega_{n,k}) \lesssim 2^n\, 2^{-3n/2} = 2^{-n/2}\end{aligned}$$

and our claim follows by sending $n \to \infty$. □

In fact, one can slightly improve the previous proof to obtain a stronger statement:

Proposition 2.13. *Let $\alpha > \frac{1}{2}$. With probability 1 there exists no $t \in [0, +\infty)$ at which B is α-Hölder continuous (i.e. such that $|B_s - B_t| \leqslant C|t - s|^\alpha$ for all $s \geqslant 0$).*

The proof is left as part of Exercise Sheet 3.

Remark 2.14. It is not true that the Brownian motion is nowhere 1/2-Hölder continuous: there are so-called "slow points" where it shows an exceptional behavior. This is beyond the scope of our lecture. But it is not very difficult to see that if $t \geqslant 0$ is fixed, then almost surely

$$\limsup_{s \to t} \frac{|B_s - B_t|}{|s - t|^{1/2}} = \infty. \tag{2.8}$$

One can combine our results on regularity of B at 0, with the fact that $(t B_{1/t})_{t \geqslant 0}$ is a Brownian motion (time invertion), to learn something about the long time behavior of B:

Corollary 2.15. *For any $\alpha > 1/2$, with probability 1 we have*

$$0 = \lim_{t \to \infty} \frac{|B_t|}{t^\alpha} < \limsup_{t \to \infty} \frac{|B_t|}{t^{1/2}} = \infty.$$

Proof. We have

$$\limsup_{t \to \infty} \frac{|B_t|}{t^\alpha} \overset{s = \frac{1}{t}}{=} \limsup_{s \to 0} \frac{|s B_{1/s}|}{s \cdot s^{-\alpha}} = \limsup_{s \to 0} \frac{|\tilde{B}_s|}{s^{1-\alpha}},$$

where $\tilde{B}_s = s B_{1/s}$ is another Brownian motion by Proposition 2.11-*iv.*; since $1-\alpha<1/2$, we can find $\varepsilon>0$ small enough such that a.s. $\tilde{B}\in C^{1-\alpha+\varepsilon}([0,1])$, so that

$$\limsup_{s\to 0}\frac{|\tilde{B}_s|}{s^{1-\alpha}}\leqslant \limsup_{s\to 0}\|\tilde{B}\|_{C^{1-\alpha+\varepsilon}}\frac{s^{1-\alpha+\varepsilon}}{s^{1-\alpha}}=0$$

which implies that the limsup is a limit and equals 0. For $\alpha=1/2$, we similarly get

$$\limsup_{t\to\infty}\frac{|B_t|}{t^{1/2}}=\limsup_{s\to 0}\frac{|\tilde{B}_s|}{s^{1/2}}=\infty$$

where the last equality comes from (2.8). □

So far we showed: Brownian motion is almost surely $(1/2-\varepsilon)$-Hölder continuous on every compact interval, and it is almost surely nowhere $(1/2+\varepsilon)$-Hölder continuous. With some more work, it can be shown that at $\alpha=\frac{1}{2}$ there are some logarithmic corrections.

The next statement is not examinable, but it is included here for completeness so that you have seen it at least once, as it's a fairly celebrated result.

Theorem 2.16. (Lévy's modulus of continuity & law of the iterated logarithm)

i. Lévy's modulus of continuity*: Almost surely, for any* $T\in(0,+\infty)$*:*

$$\lim_{r\to 0}\sup_{\substack{s,t\in[0,T]:\\|t-s|\leqslant r}}\frac{|B_t-B_s|}{\sqrt{2r\log(1/r)}}=1.$$

ii. Law of the iterated logarithm*: For any* $t>0$ *we have almost surely*

$$\limsup_{r\to 0}\frac{B_{t+r}-B_t}{\sqrt{2r\log\log(1/r)}}=1,\qquad \liminf_{r\to 0}\frac{B_{t+r}-B_t}{\sqrt{2r\log\log(1/r)}}=-1.$$

Proof. See Revuz-Yor [23], Theorem I.2.7 and Theorem II.1.9. □

Of course, we would suspect to have

$$-\infty=\liminf_{t\to\infty}\frac{B_t}{t^{1/2}}<\limsup_{t\to\infty}\frac{B_t}{t^{1/2}}=\infty,$$

and indeed one can use Theorem 2.16 and time invertion to prove this (and more). We will not do so, but instead obtain this later as a simple consequence of Blumenthal's 0-1-law.

Exercise. Use the law of the iterated logarithm to deduce a stronger statement about the long time behavior of B.

2.3 The Poisson process

We can interpret the Brownian motion as a continuous time random walk, because just like a random walk it has independent and stationary increments (i.e. B_t-B_s is independent of what happened until time s, and it has the same distribution as $B_{r+t}-B_{r+s}$ for any r). It is natural to ask whether there are other processes of this type. This leads to the following definition:

Definition 2.17. (Lévy process) *A real-valued stochastic process* $(X_t)_{t\geqslant 0}$ *is called a* Lévy process *if*

i. $X_0=0$*;*

ii. *for all* $0 \leqslant t_0 < t_1 < \cdots < t_n$ *the random variables* $(X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}})$ *are independent (*independent increments*);*

iii. *for all* $0 \leqslant s < t$ *the random variable* $X_t - X_s$ *has the same distribution as* X_{t-s} *(*stationary increments*);*

iv. *for all* $\varepsilon > 0$ *and* $t \geqslant 0$ *we have* $\lim_{h \to 0} \mathbb{P}(|X_{t+h} - X_t| > \varepsilon) = 0$ *(*continuity in probability*).*

Exercise.

a) Convince yourself of the following: parts *i.-ii.* imply that $X_t - X_s$ is independent of $(X_r)_{r \leqslant s}$; part *iv.* is equivalent to X_s converging in probability to X_t as $s \to t$.

b) Show that the Brownian motion is a Lévy process.

c) Show that for all $a \in \mathbb{R}$ the linear function $X_t = a \cdot t$ is a Lévy process.

If X is a Lévy process, then we can write

$$X_1 = X_{1/n} + (X_{2/n} - X_{1/n}) + \cdots + (X_1 - X_{(n-1)/n}).$$

Let us write μ_t for the law of X_t. Then the left hand side has law μ_1, and the right hand side has law given by the n*-fold convolution*

$$\mu_{1/n}^{*n} := \underbrace{\mu_{1/n} * \ldots * \mu_{1/n}}_{n\,\text{times}};$$

this is because, if U and V are independent random variables with $\mathrm{law}(U) = \pi$ and $\mathrm{law}(V) = \nu$, then $\mathrm{law}(U + V) = \pi * \nu$, where the convolution $\pi * \nu$ is the measure defined by

$$\pi * \nu(A) := \int \mathbb{1}_A(x + y) \pi(\mathrm{d}x) \nu(\mathrm{d}y).$$

Thus, for all $n \in \mathbb{N}$ there exists a measure $\mu_{1/n}$ such that $\mu_1 = \mu_{1/n}^{*n}$. Any μ which has this property is called *infinitely divisible*. So, if X is a Lévy process, then X_1 is infinitely divisible. Conversely, one can show that for any infinitely divisible distribution μ there exists a unique (in law) Lévy process X with $\mathrm{law}(X_1) = \mu$. Therefore, Lévy processes are in one-to-one correspondence with infinitely divisible distributions.

Exercise. (Difficult) Show that if X is a centered Lévy process such that X_1 is bounded (i.e. there exists $C > 0$ such that a.s. $|X_1| \leqslant C$), then $X_1 = 0$ a.s.

Hint: Consider $\mathrm{var}(X_1)$.

∘

The infinite divisibility imposes strong structural constraints on Lévy processes, and every Lévy process can be characterized in terms of its *Lévy-Khintchine representation*:

Theorem 2.18. (Lévy-Khintchine representation) *If* X *is a Lévy process, then the characteristic function of* X *satisfies*

$$\mathbb{E}[e^{iuX_t}] = e^{t\psi(u)}, \qquad t \geqslant 0, u \in \mathbb{R},$$

where ψ *is of the form*

$$\psi(u) = iau - \frac{1}{2}\sigma^2 u^2 + \int_{\mathbb{R} \setminus \{0\}} (e^{iux} - 1 - iux\mathbb{1}_{\{|x| < 1\}}) \nu(\mathrm{d}x),$$

for $a \in \mathbb{R}$, $\sigma^2 \geqslant 0$, *and for a measure* ν *on* $\mathbb{R} \setminus \{0\}$ *such that* $\int_{\mathbb{R} \setminus \{0\}} (1 \wedge |x|^2) \nu(\mathrm{d}x) < \infty$ *(a so called* Lévy measure*). We call* (a, σ^2, ν) *the* Lévy triple *of* X.

Proof. See Klenke [15], Theorem 16.17. □

Remark 2.19. Without further explanation, this result is not very interesting. But it has a neat probabilistic interpretation: Let X be a Lévy process with characteristic function $\mathbb{E}[e^{iuX_t}] = e^{t\psi(u)}$ for

$$\psi(u) = iau - \frac{1}{2}\sigma^2 u^2 + \int_{\mathbb{R}\setminus\{0\}} (e^{iux} - 1 - iux\mathbb{1}_{\{|x|<1\}})\nu(\mathrm{d}x).$$

Then we can decompose X into a sum of three independent processes, $X = X^{(1)} + X^{(2)} + X^{(3)}$, where

- $X_t^{(1)} = at$;
- $X_t^{(2)} = \sigma B_t$, for a Brownian motion B;
- $X_t^{(3)}$ is a "jump process", with jumps determined by the Lévy measure ν.

Exercise. Show that if $\nu = 0$ and thus $X^{(3)} \equiv 0$, then for $X_t^{(1)} = at$ and $X_t^{(2)} = \sigma B_t$ we get the claimed form of the characteristic function.

Since ν describes jumps (namely, points where jump discontinuities arise in the map $t \mapsto X_t$), this implies that the Brownian motion is the only centered and continuous Lévy process (up to constant multiples).

Example 2.20. (Pre-Poisson process) Let X be a Lévy process such that for $\lambda > 0$:

$$\psi(u) = \lambda(e^{iu} - 1),$$

i.e. $a = \sigma^2 = 0$ and $\nu = \lambda\delta_1$ is a multiple of the Dirac measure in $x = 1$. Then

$$\mathbb{E}[e^{iuX_t}] = e^{\lambda t(e^{iu}-1)},$$

which is the characteristic function of a Poisson distribution with parameter λt. We call this process the *pre-Poisson process (with intensity* λ*)*.

Poisson distribution: Recall that a random variable Y with values in $\mathbb{N}_0$ has a Poisson distribution with parameter $\lambda \geqslant 0$ if $\mathbb{P}(Y = k) = \frac{\lambda^k}{k!}e^{-\lambda}$, $k \in \mathbb{N}_0$. We write $Y \sim \mathrm{Poi}(\lambda)$.

Exercise. Show that the Poisson distribution with parameter $\lambda \geqslant 0$ has the characteristic function $\mathbb{E}[e^{iuX}] = \exp(\lambda(e^{iu} - 1))$.

Remark 2.21. Alternatively, we can describe the pre-Poisson process as follows: A stochastic process $(N_t)_{t\geqslant 0}$ is a pre-Poisson process with intensity $\lambda > 0$ if and only if the following conditions are satisfied:

i. $N_0 = 0$ almost surely;

ii. for all $0 \leqslant s < t$ the random variable $N_t - N_s$ is independent of $(N_r)_{0\leqslant r\leqslant s}$;

iii. for all $0 \leqslant s < t$ we have $N_t - N_s \sim \mathrm{Poi}(\lambda(t - s))$.

Exercise. Convince yourself of this!

Hopefully it will not be a surprise to you at this point that the Poisson process (without "pre-") will be a pre-Poisson process with nice trajectories. But note that the Poisson distribution takes values in $\mathbb{N}_0$, so the Poisson process cannot be continuous. To formulate its path properties, we introduce the following notation for $f: \mathbb{R}_+ \to \mathbb{R}$:

$$f(t+) := \lim_{s\downarrow t} f(s) := \lim_{\substack{s\to t\\ s>t}} f(s), \qquad f(t-) := \lim_{s\uparrow t} f(s) := \lim_{\substack{s\to t\\ s<t}} f(s),$$

and similarly $\limsup_{s\downarrow t} f(s)$, $\liminf_{s\uparrow t} f(s)$, etc. Recall that a function f is called *right-continuous* (resp. *left-continuous*) if $f(t+) = f(t)$ for all $t \geqslant 0$ (resp. $f(t-) = f(t)$ for all $t > 0$).

Exercise. Which of these functions is left- and/or right-continuous?

i. $f = \mathbb{1}_{[1,\infty)}$;
ii. $f = \mathbb{1}_{(1,\infty)}$;
iii. $f = \mathbb{1}_{\{1\}}$;
iv. $f(t) = \sin\left(\frac{1}{1-t}\right)$, $t \in [0,1)$, and $f(t) = 0$ for $t \geqslant 1$.

Definition 2.22. (Càdlàg) *A function $f: \mathbb{R}_+ \to \mathbb{R}$ is called* càdlàg *if it is right-continuous and at every $t > 0$ the limit $f(t-)$ exists (but might not be equal to $f(t)$). A càdlàg function $t \mapsto f(t)$ has a* jump *at t if*

$$\Delta f(t) := f(t) - f(t-) \neq 0.$$

The acronym càdlàg comes from French and stands for "continue à droite, limite à gauche", that is "continuous from the right, limits from the left".

Lemma 2.23. *Let $f: \mathbb{R}_+ \to \mathbb{R}$ be càdlàg; then f has at most countably many jumps, i.e. there exist at most countably many $\{t_n\}_{n\in\mathbb{N}}$ such that $\Delta f(t_n) > 0$.*

Proof. Exercise Sheet 3. □

Definition 2.24. (Càdlàg process, Poisson process)

i. We say that a stochastic process $X = (X_t)_{t\geqslant 0}$ with values in $\mathbb{R}^d$ is càdlàg *if all of its trajectories are* càdlàg*, i.e. $t \mapsto X_t(\omega)$ is* càdlàg *for all $\omega \in \Omega$.*

ii. A càdlàg pre-Poisson process is called a Poisson process.

To get a more intuitive understanding of the Poisson process, we use the following explicit construction:

Theorem 2.25. *Let $(T_n)_{n\in\mathbb{N}}$ be an i.i.d. sequence of exponentially distributed random variables with parameter $\lambda > 0$. We define $S_n := T_1 + \cdots + T_n$ and*

$$N_t := \max\{n: S_n \leqslant t\}, \qquad t \geqslant 0.$$

Then $(N_t)_{t\geqslant 0}$ is (indistinguishable from) a Poisson process with intensity λ.

Recall that the **exponential distribution** with parameter $\lambda > 0$ has density $\mathbb{1}_{\mathbb{R}_+}(x)\lambda e^{-\lambda x}$.

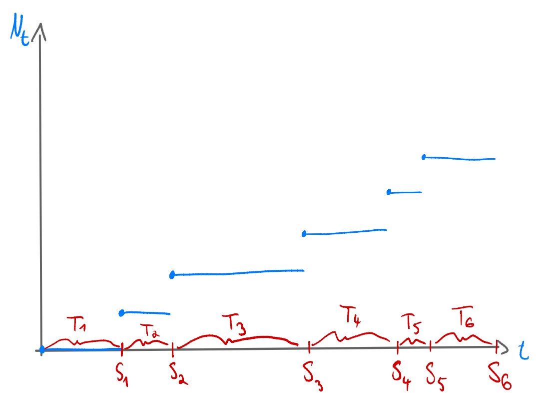


Figure 2.4. Poisson process constructed from $(T_n)_{n\geqslant 0}$.

Let us momentarily postpone the proof of Theorem 2.25 and discuss the insight it provides on the nature of the trajectories of $(N_t)_{t\geqslant 0}$ first.

The Poisson process is a “counting process” which is piecewise constant and which jumps up by 1 at random times $(S_n)_{n\in\mathbb{N}}$. The times between the jumps are independent, and they follow an exponential distribution with parameter λ.

The Poisson process is used to model the (cumulative) number of customers arriving at a store, or the damage claims at an insurance, the number of clicks of a Geiger counter (which corresponds to the number of decaying atoms), or the number of meteorites hitting earth.

The Poisson process is in some sense “the most elementary jump process” and almost all other pure jump processes can be constructed from it. On Sheet 3 you will construct a Lévy process with a general Lévy measure ν, provided that ν has finite mass ($\nu(\mathbb{R}\setminus\{0\})<\infty$). Such a process is also called *compound Poisson process*, because we can represent it as

$$X_t=\sum_{k=1}^{N_t} Y_k,$$

where $N=(N_t)_{t\geqslant 0}$ is a Poisson process with intensity $\lambda=\nu(\mathbb{R}\setminus\{0\})$, and $(Y_k)_{k\in\mathbb{N}}$ is independent of N and an i.i.d. sequence of random variables with distribution $Y_k\sim\frac{\nu}{\nu(\mathbb{R}\setminus\{0\})}$. It follows from a computation that the Lévy triple of X is $a=\int_{|x|<1} x\nu(\mathrm{d}x)$, $\sigma^2=0$, ν.

——— End of the lecture on October 31st ———

One could give an alternative derivation of the Poisson process, where we divide $\mathbb{R}_+$ in intervals of length $\frac{1}{n}$ and consider a discrete time process which on each of these intervals jumps up by 1 with a small probability of order $\frac{\lambda}{n}$, independently of what happened before. For $n\to\infty$, the finite-dimensional distributions of this process converge to the finite-dimensional distributions of a Poisson process with intensity λ (this is sometimes referred to as the *law of small numbers*); more details on this construction might appear later in the Exercise Sheets.

Hopefully, this makes it plausible why we can model the phenomena mentioned above with a Poisson process: For example, each second there is a small probability that a customer enters our store, and at first approximation we can consider the different second-long intervals as independent.

We can now finally conclude this section with the

Proof of Theorem 2.25. We show that N satisfies the properties of Remark 2.21, and that it is almost surely càdlàg.

- N is almost surely càdlàg: By definition, N is càdlàg on the set $\{\sup_n S_n=\infty\}$. But by σ-continuity we have

$$\begin{aligned}\mathbb{P}\Big(\sup_n S_n<\infty\Big) &= \lim_{m\to\infty}\mathbb{P}\Big(\sup_n S_n<m\Big)\leqslant \lim_{m\to\infty}\lim_{n\to\infty}\mathbb{P}(S_n<m)\\ &\leqslant \lim_{m\to\infty}\lim_{n\to\infty}\mathbb{P}\Big(\max_{k=1,\ldots,n} T_n<m\Big)\\ &= \lim_{m\to\infty}\lim_{n\to\infty}\mathbb{P}(T_1<m)^n=0,\end{aligned}$$

  because $\mathbb{P}(T_1<m)<1$.

- We have $N_0 = 0$ by definition. So to verify the conditions of Remark 2.21 it suffices to show that for all $n \in \mathbb{N}$ and $0 = t_0 < t_1 < \cdots < t_n$ the random variables $(N_{t_1} - N_{t_0}, \ldots, N_{t_n} - N_{t_{n-1}})$ are independent, and that $N_{t_{k+1}} - N_{t_k} \sim \mathrm{Poi}(\lambda(t_{k+1} - t_k))$. For simplicity we restrict our attention to the case $n = 2$. The general case can be handled by similar arguments, but the notation becomes much more tedious. We will show that for all $0 \leqslant s < t$ and all $k, \ell \in \mathbb{N}_0$

$$\mathbb{P}(N_s = k, N_t - N_s = \ell) = \frac{(\lambda s)^k}{k!} e^{-\lambda s} \frac{(\lambda(t-s))^\ell}{\ell!} e^{-\lambda(t-s)}. \tag{2.10}$$

By summing over $\ell \in \mathbb{N}_0$ respectively $k \in \mathbb{N}_0$ we see that $N_s \sim \mathrm{Poi}(\lambda s)$ and $N_t - N_s \sim \mathrm{Poi}(\lambda(t-s))$, and then that N_s and $N_t - N_s$ are independent.

To handle the computation, we need a bit of notation. Note that $(T_1, \ldots, T_{k+\ell+1})$ has the density $\mathbb{1}_{\mathbb{R}_+^{k+\ell+1}}(x) \lambda^{k+\ell+1} e^{-\lambda \Sigma_{k+\ell+1}(x)}$ with respect to Lebesgue measure on $\mathbb{R}^{k+\ell+1}$, where for $n \leqslant m$ we define $\Sigma_n: \mathbb{R}^m \to \mathbb{R}$ by

$$\Sigma_n(x) := x_1 + \cdots + x_n.$$

It will be also useful to exploit the following fact: it holds

$$\int_{\mathbb{R}_+^n} \mathbb{1}_{\{\Sigma_n(x) \leqslant r\}} \mathrm{d}x = \int_{\mathbb{R}_+^n} \mathbb{1}_{\{x_1 + \cdots + x_n \leqslant r\}} \mathrm{d}x = \frac{r^n}{n!} \quad \forall n \geqslant 1, r \in (0, +\infty). \tag{2.11}$$

We postpone the verification of (2.11) to the end of the proof.

- We start by considering the case $\ell = 0$. In this case

$$\begin{aligned}
\mathbb{P}(N_s = k, N_t - N_s = 0) &= \mathbb{P}(N_s = k, N_t = N_s) \\
&= \mathbb{P}(S_k \leqslant s < S_{k+1}, S_k \leqslant t < S_{k+1}) \\
&= \mathbb{P}(S_k \leqslant s, S_{k+1} > t) \\
&= \mathbb{P}(S_k \leqslant s, T_{k+1} > t - S_k).
\end{aligned}$$

Since T_{k+1} is independent of $(T_1, \ldots, T_k)$, thus S_k, we can compute the above probability by first conditioning on T_{k+1} as

$$\mathbb{P}(S_k \leqslant s, T_{k+1} > t) = \int_{\mathbb{R}_+^k} \mathbb{1}_{\{\Sigma_k(x) \leqslant s\}} \lambda^k e^{-\lambda \Sigma_k(x)} \mathbb{P}(T_{k+1} > t - \Sigma_k(x)) \mathrm{d}x;$$

here and below, we can always change the order of integration at will by virtue of Fubini's theorem, because all terms appearing are non-negative. Since T_{k+1} is exponentially distributed, $\mathbb{P}(T_{k+1} > r) = e^{-\lambda r}$ for all $r \geqslant 0$ and so

$$\begin{aligned}
\mathbb{P}(N_s = k, N_t - N_s = 0) &= \int_{\mathbb{R}_+^k} \mathbb{1}_{\{\Sigma_k(x) \leqslant s\}} \lambda^k e^{-\lambda \Sigma_k(x)} e^{-\lambda(t - \Sigma_k(x))} \mathrm{d}x \\
&= \lambda^k e^{-\lambda t} \int_{\mathbb{R}_+^k} \mathbb{1}_{\{\Sigma_k(x) \leqslant s\}} \mathrm{d}x \\
&= \lambda^k e^{\lambda t} \frac{s^k}{k!} = \frac{(\lambda s)^k}{k!} e^{-\lambda s} \frac{(\lambda t)^0}{0!} e^{-\lambda(t-s)}
\end{aligned}$$

proving (2.10) in this case; in the intermediate passages, we applied (2.11).

- Next we deal with the case $\ell \geqslant 1$. We start by writing

$$\begin{aligned}
\mathbb{P}(N_s = k, N_t - N_s = \ell) &= \mathbb{P}(N_s = k, N_t = k + \ell) \\
&= \mathbb{P}(S_k \leqslant s < S_{k+1}, S_{k+\ell} \leqslant t < S_{k+\ell+1}).
\end{aligned}$$

Using the same notation for Σ_k as before, and the density of $(T_1, \ldots, T_{k+\ell+1})$, this leads to

$$\begin{aligned}&\mathbb{P}(S_k \leqslant s < S_{k+1}, S_{k+\ell} \leqslant t < S_{k+\ell+1})\\ &\quad= \int_{\mathbb{R}_+^{k+\ell+1}} \mathbb{1}_{\{\Sigma_k(x) \leqslant s < \Sigma_{k+1}(x)\}} \mathbb{1}_{\{\Sigma_{k+\ell}(x) \leqslant t < \Sigma_{k+\ell+1}(x)\}} \lambda^{k+\ell+1} e^{-\lambda \Sigma_{k+\ell+1}(x)} \mathrm{d}x.\end{aligned}$$

The expression is a bit more complicated than before, but we can proceed similarly. We start by integrating out $x_{k+\ell+1}$. We apply the change of variables $z = \Sigma_{k+\ell+1}(x)$ and obtain

$$\begin{aligned}\int_0^\infty \mathbb{1}_{\{\Sigma_{k+\ell}(x) \leqslant t < \Sigma_{k+\ell+1}(x)\}} \lambda e^{-\lambda \Sigma_{k+\ell+1}(x)} \mathrm{d}x_{k+\ell+1} &= \int_{\Sigma_{k+\ell}(x)}^\infty \mathbb{1}_{\{\Sigma_{k+\ell}(x) \leqslant t < z\}} \lambda e^{-\lambda z} \mathrm{d}z\\ &= \mathbb{1}_{\{\Sigma_{k+\ell}(x) \leqslant t\}} e^{-\lambda t},\end{aligned}$$

which leads to

$$\begin{aligned}&\int_{\mathbb{R}_+^{k+\ell+1}} \mathbb{1}_{\{\Sigma_k(x) \leqslant s < \Sigma_{k+1}(x)\}} \mathbb{1}_{\{\Sigma_{k+\ell}(x) \leqslant t < \Sigma_{k+\ell+1}(x)\}} \lambda^{k+\ell+1} e^{-\lambda \Sigma_{k+\ell+1}(x)} \mathrm{d}x\\ &\quad= \lambda^{k+\ell} e^{-\lambda t} \int_{\mathbb{R}_+^{k+\ell}} \mathbb{1}_{\{\Sigma_k(x) \leqslant s < \Sigma_{k+1}(x)\}} \mathbb{1}_{\{\Sigma_{k+\ell}(x) \leqslant t\}} \mathrm{d}x.\end{aligned}$$

Now we perform the change of variables $y_1 = \Sigma_{k+1}(x) - s$, $y_2 = x_{k+2}$, $\ldots$, $y_\ell = x_{k+\ell}$ and obtain

$$\begin{aligned}&\int_{\mathbb{R}_+^{k+\ell}} \mathbb{1}_{\{\Sigma_k(x) \leqslant s < \Sigma_{k+1}(x)\}} \mathbb{1}_{\{\Sigma_{k+\ell}(x) \leqslant t\}} \mathrm{d}x\\ &\quad= \int_{\mathbb{R}_+^k} \mathbb{1}_{\{\Sigma_k(x) \leqslant s\}} \left(\int_{\mathbb{R}_+^\ell} \mathbb{1}_{\{y_1 + \cdots + y_\ell \leqslant t - s\}} \mathrm{d}y \right) \mathrm{d}x\\ &\quad= \frac{s^k}{k!} \cdot \frac{(t-s)^\ell}{\ell!},\end{aligned}$$

where in the last passage we used (2.11). Altogether, we have shown that for $k \in \mathbb{N}_0$ and $\ell \in \mathbb{N}$,

$$\mathbb{P}(N_s = k, N_t - N_s = \ell) = \frac{s^k}{k!} \cdot \frac{(t-s)^\ell}{\ell!} \lambda^{k+\ell} e^{-\lambda t} = \frac{(\lambda s)^k}{k!} e^{-\lambda s} \cdot \frac{(\lambda(t-s))^\ell}{\ell!} e^{-\lambda(t-s)};$$

namely, (2.10) holds. Together with the previous case, this complete the verification of (2.10) for any $k, \ell \in \mathbb{N}_0$.

- We finally prove (2.11). For $n = 1$ we have $\int_{\mathbb{R}_+^1} \mathbb{1}_{\{x_1 \leqslant r\}} \mathrm{d}x = \int_0^r 1 \mathrm{d}x_1 = r^1/1!$ and then by induction

$$\begin{aligned}\int_{\mathbb{R}_+^n} \mathbb{1}_{\{x_1 + \cdots + x_n \leqslant r\}} \mathrm{d}x &= \int_0^r \left(\int_{\mathbb{R}_+^{n-1}} \mathbb{1}_{\{x_1 + \cdots + x_{n-1} \leqslant r - x_n\}} \mathrm{d}x_1 \cdots \mathrm{d}x_{n-1} \right) \mathrm{d}x_n\\ &= \int_0^r \frac{(r - x_n)^{n-1}}{(n-1)!} \mathrm{d}x_n = \frac{r^n}{n!}\end{aligned}$$

which shows (2.11). □

[Comment: in Theorem 2.25 we verified that N_t is a Poisson process "by hand", using elementary but lengthy manipulations. There is an alternative more elegant approach, based on the notion of infinitesimal generator *of the Markov process, but it is outside the scope of these lectures; see Theorem 2.4.3 from [19] for more details.]*

Exercise. Let $(N_t^1)_{t\geqslant 0}$ and $(N_t^2)_{t\geqslant 0}$ be independent Poisson processes with intensity $\lambda_1>0$ respectively $\lambda_2>0$. Show that $N_t:=N_t^1+N_t^2$, $t\geqslant 0$, is a Poisson process with intensity $\lambda_1+\lambda_2$.

3 Filtrations and stopping times

So far we have analysed Brownian motion as one of the most canonical examples of: a) (continuous) Gaussian processes; b) Lévy processes. There are two other fundamental categories of which Brownian motion is a prominent example, which are respectively Markov processes and (continuous) martingales. In order to introduce them, we first need to make a detour on the concept of filtrations, stopping times and progressive processes.

3.1 Filtrations and stopping times

Definition 3.1. (Filtration, right-continuous)

i. A filtration *is an increasing family* $\mathbb{F}=(\mathcal{F}_t)_{t\geqslant 0}$ *of sub sigma-algebras of* $\mathcal{F}$*, i.e. such that* $\mathcal{F}_s\subset\mathcal{F}_t\subset\mathcal{F}$ *for all* $0\leqslant s\leqslant t$*. We write*

$$\mathcal{F}_\infty:=\bigvee_{t\geqslant 0}\mathcal{F}_t=\sigma\Bigg(\bigcup_{t\geqslant 0}\mathcal{F}_t\Bigg),\qquad \mathcal{F}_{t+}:=\bigcap_{s>t}\mathcal{F}_s,\qquad \text{for any } t\geqslant 0.$$

We call $(\Omega,\mathcal{F},\mathbb{F},\mathbb{P})$ *a* filtered probability space.

ii. A filtration $\mathbb{F}$ *is called* right-continuous *if* $\mathcal{F}_{t+}=\mathcal{F}_t$ *for all* $t\geqslant 0$*. We write* $\mathbb{F}^+=(\mathcal{F}_t^+)_{t\geqslant 0}$ *for the smallest right-continuous filtration containing* $\mathbb{F}$*, given by*

$$\mathcal{F}_t^+:=\bigcap_{s>t}\mathcal{F}_s=\mathcal{F}_{t+},\qquad t\geqslant 0.$$

Note that $(\mathbb{F}^+)^+=\mathbb{F}^+$ *for every filtration.*

Exercise. Show the last statement.

—— End of the lecture on November 6 ——

Definition 3.2. (Canonical/natural filtration) *Let* $(X_t)_{t\geqslant 0}$ *be a stochastic process and set* $\mathcal{F}_t:=\sigma(X_s: s\leqslant t)$*. In this case we write* $\mathbb{F}^X:=(\mathcal{F}_t^X)_{t\geqslant 0}:=(\mathcal{F}_t)_{t\geqslant 0}$ *and we call* $\mathbb{F}^X$ *the* canonical filtration *(or* natural filtration*) of* X*. We also write* $\mathbb{F}^{X+}:=(\mathbb{F}^X)^+$.

Note that in general we can have $\mathbb{F}^{X+}\neq\mathbb{F}^X$, even if X is continuous and real-valued: for example

$$A:=\left\{\omega:\lim_{h\downarrow 0}\frac{X_{t+h}(\omega)-X_t(\omega)}{h}\text{ exists}\right\}\in\mathcal{F}_t^{X+},$$

but in general $A\notin\mathcal{F}_t^X$.

Definition 3.3. (Adapted process) *A stochastic process* $(X_t)_{t\geqslant 0}$ *is called* adapted *to a given filtration* $\mathbb{F}$ *if* X_t *is* $\mathcal{F}_t$*–measurable for all* $t\geqslant 0$.

Clearly, any process X is adapted to its canonical filtration $\mathbb{F}^X$.

Definition 3.4. (Negligible sets, complete σ-algebra, completion) *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra.*

i. A set $B \subset \Omega$ is called $\mathbb{P}$-negligible *(with respect to $\mathcal{F}$), or simply* negligible*, if there exists $N \in \mathcal{F}$ with $\mathbb{P}(N) = 0$ such that $B \subset N$. We write $\mathcal{N}^{\mathbb{P}}$ for the $\mathbb{P}$-negligible sets.*

ii. $\mathcal{G}$ is called complete *(with respect to $\mathcal{F}$) if $\mathcal{N}^{\mathbb{P}} \subset \mathcal{G}$.*

iii. The completion of *$\mathcal{G}$ (with respect to $\mathcal{F}$) is $\mathcal{G}^{\mathbb{P}} := \sigma(\mathcal{G} \cup \mathcal{N}^{\mathbb{P}})$.*

Remark 3.5. (Exercise, see also Durrett [6], Theorem A.2.3) The completion $\mathcal{F}^{\mathbb{P}}$ of $\mathcal{F}$ is given by

$$\mathcal{F}^{\mathbb{P}} = \{A \cup B : A \in \mathcal{F}, B \in \mathcal{N}^{\mathbb{P}}\}.$$

Therefore, we can uniquely extend $\mathbb{P}$ from $\mathcal{F}$ to $\mathcal{F}^{\mathbb{P}}$ by setting

$$\mathbb{P}\,(A \cup B) = \mathbb{P}(A).$$

Definition 3.6. (Complete filtration, usual conditions) *Let $\mathbb{F}$ be a filtration.*

i. $\mathbb{F}$ is called complete *if $\mathcal{F}_t$ is complete for all $t \geqslant 0$; equivalently, if $\mathcal{F}_0$ is complete.*

ii. $\mathbb{F}$ satisfies the usual conditions *if it is right-continuous and complete.*

If $\mathbb{F}$ is a filtration, then

$$\mathbb{F}^{+,\mathbb{P}} := (\mathbb{F}^{\mathbb{P}})^+$$

is right-continuous and complete by construction. One can show that it is the smallest filtration containing $\mathbb{F}$ that satisfies the usual conditions. This is called the *usual augmentation (or extension)* of a filtration. Here it is important that we first take the completion, and after that we make the filtration right-continuous: if we took the opposite order, then the resulting filtration $(\mathbb{F}^+)^{\mathbb{P}}$ might not be right-continuous.

Definition 3.7. (Stopping time, events determined until τ) *Let $\mathbb{F}$ be a filtration. An* $\mathbb{F}$-stopping time *(or simply* stopping time*, if there is no ambiguity about the filtration) is a map $\tau: \Omega \to [0, \infty]$ such that $\{\tau \leqslant t\} \in \mathcal{F}_t$ for all $t \geqslant 0$. If τ is a $\mathbb{F}$-stopping time, then we write*

$$\mathcal{F}_\tau := \{A \in \mathcal{F} : A \cap \{\tau \leqslant t\} \in \mathcal{F}_t \text{ for all } t \geqslant 0\} \tag{3.1}$$

for the σ-algebra of events determined until τ.

Stopping times have many useful properties:

Lemma 3.8. *Let $\mathbb{F}$ be a filtration and let $\tau: \Omega \to [0, \infty]$.*

i. If τ is a stopping time, then $\mathcal{F}_\tau$ is indeed a σ-algebra.

ii. If $\tau(\omega) = t$ for all ω, where $t \in [0, \infty]$ is fixed, then τ is a stopping time and $\mathcal{F}_t = \mathcal{F}_\tau$ where $\mathcal{F}_\tau$ is defined in (3.1). So our definitions are consistent.

iii. If τ is a stopping time, then so is $\tau + t$ for any $t \in [0, \infty]$; the same is not necessarily true for $\tau - t$. In particular, $\mathcal{F}_{\tau+t}$ is a σ-algebra for any $t \geqslant 0$ and we can define $\mathcal{F}_{\tau+} := \bigcap_{t>0} \mathcal{F}_{\tau+t}$.

iv. τ is an $\mathbb{F}^+$-stopping time if and only if $\{\tau < t\} \in \mathcal{F}_t$ for all $t > 0$.

v. If τ is a stopping time, then τ is $\mathcal{F}_\tau$–measurable.

vi. If τ_1, τ_2 are stopping times with $\tau_1(\omega) \leqslant \tau_2(\omega)$ for all $\omega \in \Omega$, then $\mathcal{F}_{\tau_1} \subset \mathcal{F}_{\tau_2}$.

vii. If τ_1, τ_2 are stopping times, then $\tau_1 \vee \tau_2$ and $\tau_1 \wedge \tau_2$ are stopping times and

$$\mathcal{F}_{\tau_1 \wedge \tau_2} = \mathcal{F}_{\tau_1} \cap \mathcal{F}_{\tau_2}.$$

Proof. Parts *i.*, *ii.*, *iii.* and *v.* are part of Exercise Sheet 4. Let us prove the rest.

iv. If $\{\tau < r\} \in \mathcal{F}_r$ for all r, then

$$\{\tau \leqslant t\} = \bigcap_{n \in \mathbb{N}} \left\{\tau < t + \frac{1}{n}\right\} \subset \bigcap_{s > t} \mathcal{F}_s = \mathcal{F}_{t+},$$

so τ is an $\mathbb{F}^+$-stopping time. If τ is an $\mathbb{F}^+$-stopping time, then

$$\{\tau < t\} = \bigcup_{n \in \mathbb{N}} \underbrace{\left\{\tau \leqslant t - \frac{1}{n}\right\}}_{\in \mathcal{F}^+_{t-1/n} \subset \mathcal{F}_t} \in \mathcal{F}_t.$$

vi. Let $A \in \mathcal{F}_{\tau_1}$ and $t \geqslant 0$. Then $A \in \mathcal{F}$ and since $\{\tau_1 \leqslant t\} \supset \{\tau_2 \leqslant t\}$, we have

$$A \cap \{\tau_2 \leqslant t\} = \underbrace{(A \cap \{\tau_1 \leqslant t\})}_{\in \mathcal{F}_t} \cap \{\tau_2 \leqslant t\} \in \mathcal{F}_t.$$

therefore $A \in \mathcal{F}_{\tau_2}$ as well.

vii. For all $t \geqslant 0$, we have:

$$\{\tau_1 \vee \tau_2 \leqslant t\} = \{\tau_1 \leqslant t\} \cap \{\tau_2 \leqslant t\} \in \mathcal{F}_t,$$

so $\tau_1 \vee \tau_2$ is a stopping time. Similarly,

$$\{\tau_1 \wedge \tau_2 \leqslant t\} = \{\tau_1 \leqslant t\} \cup \{\tau_2 \leqslant t\} \in \mathcal{F}_t,$$

so $\tau_1 \wedge \tau_2$ is a stopping time.

Since $\tau_1 \wedge \tau_2 \leqslant \tau_i$ for $i = 1, 2$, by *vi.* we know that $\mathcal{F}_{\tau_1 \wedge \tau_2} \subset \mathcal{F}_{\tau_1} \cap \mathcal{F}_{\tau_2}$. Conversely, let $A \in \mathcal{F}_{\tau_1} \cap \mathcal{F}_{\tau_2}$. Then $A \in \mathcal{F}$ and

$$A \cap \{\tau_1 \wedge \tau_2 \leqslant t\} = (A \cap \{\tau_1 \leqslant t\}) \cup (A \cap \{\tau_2 \leqslant t\}) \in \mathcal{F}_t$$

and thus $A \in \mathcal{F}_{\tau_1 \wedge \tau_2}$. □

The most important examples of stopping times are so called hitting times:

Definition 3.9. (Entry/hitting time) *Let $X = (X_t)_{t \geqslant 0}$ be a stochastic process taking values in a measurable space $(S, \mathcal{S})$. For $A \in \mathcal{S}$, we define the* entry time*, or* hitting time *of X into A, as*

$$\tau_A(\omega) := \inf\{t \geqslant 0 \colon X_t(\omega) \in A\},$$

where we adopt the standard convetion $\inf \emptyset = \infty$*.*

If X is adapted to $\mathbb{F}$, then we would expect that by knowing the trajectory of X until time $t \geqslant 0$, we can decide whether X entered A strictly before t:

$$\{\tau_A < t\} = \bigcup_{s < t} \{X_s \in A\}.$$

on the other hand, note that e.g. for an open set A, it is intuitively clear that in general we can only decide if $\tau_A \leqslant t$ if we can "peak a bit into the future". So one might hope that

$$\{\tau_A \leqslant t\} = \bigcap_{\varepsilon > 0} \{\tau_A < t + \varepsilon\} = \bigcap_{\varepsilon > 0} \bigcup_{s < t + \varepsilon} \{X_s \in A\} \in \mathbb{F}^+$$

and so that τ_A is an $\mathbb{F}^+$-stopping time. But there is a problem with these arguments: While $\{X_s \in A\} \in \mathcal{F}_s$, our descriptions of $\{\tau_A < t\}$ and $\{\tau_A \leqslant t\}$ involve unions of uncountably many events, so a priori these sets are not in $\mathcal{F}_t$ or $\mathcal{F}_{t+}$. We could only deduce that τ_A is a stopping time or an $\mathbb{F}^+$-stopping time if we were somehow able to reduce to countably many set operations. Under suitable conditions on A and X, this is possible:

Proposition 3.10. *Let (S, d) be a metric space and let X be a stochastic process with values in S which is adapted to the filtration $\mathbb{F}$.*

i. If $A \subset S$ is open and X is right-continuous or left-continuous, then τ_A is an $\mathbb{F}^+$-stopping time.

ii. If $A \subset S$ is closed and X is continuous, then τ_A is an $\mathbb{F}$-stopping time.

Proof.

i. If A is open and X is left- or right-continuous, we have

$$\{\tau_A < t\} = \bigcup_{s \in \mathbb{Q} \cap [0,t)} \{X_s \in A\} \in \mathcal{F}_t,$$

so by Lemma 3.8 we get that τ_A is an $\mathbb{F}^+$-stopping time.

ii. We get from the continuity of X and closedness of A that infima are always realized as minima, so that

$$\{\tau_A \leqslant t\} = \left\{ \min_{s \in [0,t]} d(X_s, A) = 0 \right\} = \left\{ \inf_{s \in \mathbb{Q} \cap [0,t]} d(X_s, A) = 0 \right\} \in \mathcal{F}_t,$$

where we wrote $d(x, A) = \inf \{d(x, a) : a \in A\}$. □

Exercise. Let X be an adapted, real-valued, increasing process, i.e. such that $t \mapsto X_t(\omega)$ is (not necessarily strictly) increasing for all $\omega \in \Omega$. Show that $\tau_a := \inf \{t \geqslant 0 : X_t \geqslant a\}$ is an $\mathbb{F}^+$-stopping time for all $a \in \mathbb{R}$. If X is additionally right-continuous, then τ_a is even an $\mathbb{F}$-stopping time.

3.2 Progressively measurable processes

We discussed that entry times for (one-sided) continuous and adapted stochastic processes are stopping times. Sometimes the assumption of path continuity is too much, but adaptedness gives us no control at all about the trajectories. The notion of progressive measurability answer this issue: it is stronger than adaptedness and for example (almost) sufficient for a process to serve as a *stochastic integrand* (up to additional technical conditions, as we will see later), but at the same time much less restrictive than continuity of trajectories.

In the following, we are given a filterered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ and another measurable space $(S, \mathcal{S})$.

Definition 3.11. ((Progressively) measurable processes) *A stochastic process $X = (X_t)_{t \geqslant 0}$ taking values in $(S, \mathcal{S})$ is called*

i. measurable *if the map*

$$\Omega \times \mathbb{R}_+ \ni (\omega, t) \mapsto X_t(\omega) \in S$$

is $(\mathcal{F} \otimes \mathcal{B}(\mathbb{R}_+), \mathcal{S})$–measurable;

ii. progressive *(or* progressively measurable*) if for any* $t \geqslant 0$*, the map*

$$\Omega \times [0,t] \ni (\omega, s) \mapsto X_s(\omega) \in S$$

is $(\mathcal{F}_t \otimes \mathcal{B}([0,t]), \mathcal{S})$*–measurable.*

Remark 3.12. We can define the σ-algebra of *progressive sets* by

$$\mathrm{Prog} = \{A \in \mathcal{F} \otimes \mathcal{B}(\mathbb{R}_+) \colon A \cap (\Omega \times [0,t]) \in \mathcal{F}_t \otimes \mathcal{B}([0,t])\};$$

notice that $A \in \mathrm{Prog}$ if and only if $\mathbb{1}_A$ is progressive. It can be shown that $X \colon \Omega \times \mathbb{R}_+ \to S$ is progressively measurable if and only if it is $(\mathrm{Prog}, \mathcal{S})$-measurable; in other words, progressive measurability amounts to measurability w.r.t. the σ-algebra of progressive sets.

In the case when $(S, \mathcal{S}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, this fact has many useful consequences:

- If X and Y are progressive, so are $X+Y$ and $X \cdot Y$;
- If $(X^n)_n$ is a sequence of progressive processes, then $\limsup_n X^n$ and $\liminf_n X^n$ are still progressive; in particular, whenever it exists, $X_t(\omega) := \lim_{n\to\infty} X^n(t,\omega)$ is progressive.

Exercise. Check that Prog is a σ-algebra.

Example 3.13. If $0 \leqslant s < u$ and $Y \in \mathcal{F}_s$, then the process

$$X_t(\omega) = Y(\omega) \mathbb{1}_{[s,u)}(t) \tag{3.2}$$

is progressive; similarly for $\tilde{X}_t(\omega) = Y(\omega) \mathbb{1}_{\{s\}}(t)$.

Exercise. Show that the process defined in (3.2) is progressive.

Lemma 3.14. *Let* $(X_t)_{t \geqslant 0}$ *be a stochastic process taking values in* $(S, \mathcal{S})$*.*

i. If X *is measurable, then the map* $t \mapsto X_t(\omega)$ *is* $(\mathcal{B}(\mathbb{R}_+), \mathcal{S})$*–measurable for all* $\omega \in \Omega$*.*

ii. If X *is progressive, then it is measurable and* $\mathbb{F}$*-adapted.*

iii. If S *is a metric space,* $\mathcal{S} = \mathcal{B}(S)$ *and* X *is right-continuous (or left-continuous) and adapted, then* X *is progressive.*

——— End of the lecture on November 7 ———

Proof. (The proofs of i. and ii. were skipped in the lecture in the interest of time, but for those interested they are included here)

i. The statement is in fact part of Fubini theorem, or if you prefer the fundamental property of product σ-algebras like $\mathcal{F} \otimes \mathcal{B}(\mathbb{R}_+)$ are well-behaved under sections; a self-contained proof can be given by means of the monotone class theorem (Theorem A.11) applied with the π-system $\mathcal{E} = \{A \times B \colon A \in \mathcal{F}, B \in \mathcal{B}(\mathbb{R}_+)\}$.

ii. It is clear that X is measurable. To see that X is adapted, note that for $t \geq 0$, $\Gamma \in \mathcal{S}$ we have

$$\{\omega \colon X_t(\omega) \in \Gamma\} = \{\omega \colon (\omega, t) \in X^{-1}(\Gamma)\},$$

where we interpret X as a map from $\Omega \times [0,t]$ to $\mathcal{S}$. By assumption we have $B = X^{-1}(\Gamma) \in \mathcal{F}_t \otimes \mathcal{B}([0,t])$, and therefore the t-section $\{\omega \colon (\omega, t) \in B\}$ is in $\mathcal{F}_t$.

iii. Let X be right-continuous and fix $t \geqslant 0$. For $n \in \mathbb{N}$, define the processes

$$X^n \colon \Omega \times [0,t] \to S, \qquad X^n_s = \sum_{k=0}^{n-1} \mathbb{1}_{\left[\frac{k}{n}t, \frac{k+1}{n}t\right)}(s) X_{\frac{k+1}{n}t} + \mathbb{1}_{\{t\}}(s) X_t.$$

Since X is right-continuous, we have $\lim_{n\to\infty} X_s^n(\omega) = X_s(\omega)$ for all $s\in[0,t]$ and all $\omega\in\Omega$; so it suffices to show that X^n is $\mathcal{F}_t\otimes\mathcal{B}([0,t])$–measurable for all n. Since X is adapted, $X_{(k+1)t/n}$ is $\mathcal{F}_t$–measurable and so $\mathbb{1}_{[kt/n,(k+1)t/n)}(s)X_{(k+1)t/n}$ is $\mathcal{F}_t\otimes\mathcal{B}([0,t])$–measurable; since this property is preserved under summation, we conclude that X^n is $\mathcal{F}_t\otimes\mathcal{B}([0,t])$–measurable.

If X is left-continuous, we similarly define X^n in a discrete way by approximating X from the left. □

Lemma 3.14 indicates why progressive measurability is a useful property. For instance, it follows by similar arguments that if X is a progressive and bounded process (in the sense that there exists a constant $C>0$ such that $|X_t(\omega)|\leqslant C$ for all (t,ω)), then by the measurability of $t\mapsto X_t(\omega)$ we can construct the time integral $Y_t(\omega)=\int_0^t X_s(\omega)\mathrm{d}s$; it's not too hard to see that the resulting process $(Y_t)_{t\geqslant 0}$ is continuous and $\mathbb{F}$-adapted, therefore by Lemma 3.14 it is progressive. In particular, under mild assumptions, progressive processes are closed under the operation of integration in time. With a bit of work it can be also shown that, if X is progressive, so is the running maximum $X_t^* := \sup_{s\leqslant t} X_t$.

Remark 3.15. Due to continuity of its trajectories, it follows from Lemma 3.14 that Brownian motion $(B_t)_{t\geqslant 0}$ is progressive (w.r.t. its natural filtration), and so in particular it is $(\mathcal{F}\otimes\mathcal{B}(\mathbb{R}_+),\mathbb{R})$–measurable. A similar argument applies to the Poisson process $(N_t)_{t\geqslant 0}$.

Existence of (progressively) measurable modifications: In the lectures, Lemma 3.14 will always suffice for our purposes. The following discussion was not presented in details in the lectures and is not examinable; it is a bit more technical in nature, but useful to get a more complete picture.

We have seen before that Kolmogorov's continuity criterion provides sufficient conditions for the existence of a continuous modification, and at the same time there are stochastic processes to which it does not apply (e.g. Poisson) for which we can still have càdlàg trajectories. It makes sense to wonder whether there are other abstract results guaranteeing the existence of measurable modifications. This is indeed the case, and the standard requirement in the literature is that $(X_t)_{t\geqslant 0}$ is *stochastically continuous*, namely that $X_{t+\varepsilon}$ converges in probability to X_t whenever $\varepsilon\to 0$, for all $t\geqslant 0$; moreover, if $(X_t)_{t\geqslant 0}$ is $\mathbb{F}$-adapted and stochastically continuous, then it admits a *progressively measurable modification* $(\tilde{X}_t)_{t\geqslant 0}$, see Propositions 3.2 and 3.6-3.7 from [3].

Stochastic continuity is sufficient but not necessary for the existence of a measurable modification; the latter instead is equivalent to "stochastic quasi-continuity", see the recent [5]. Moreover, if $(X_t)_{t\geqslant 0}$ is a measurable, $\mathbb{F}$-adapted process, a classical result ensures the existence of a progressively measurable modification $(\tilde{X}_t)_{t\geqslant 0}$, see Theorem IV.30 from [4]; the standard proof however is very demanding, see the recent [21] for a more elementary one. Notice that, if we start with an adapted continuous process $(X_t)_{t\geqslant 0}$, there is no guarantee that the abstract modification $(\tilde{X}_t)_{t\geqslant 0}$ obtained in this way will be still continuous; thus in this case we are better off applying Lemma 3.14 anyway.

The interest in stopping times often comes from looking at "stopped processes", and in particular at computing the statistics of the process $(X_t)_{t\geqslant 0}$ exactly when it is evaluated at the random time τ. Namely, we are interested in the map $\omega\mapsto X_\tau(\omega) := X_{\tau(\omega)}(\omega)$ for a stopping time τ. Since $\tau(\omega)$ may be infinite, this map may not be defined for all ω. So as a convention, we introduce a "cemetery state" $\Delta\notin S$ and set

$$X_\tau(\omega) := \begin{cases} X_{\tau(\omega)}(\omega) & \text{if} \quad \tau(\omega)<\infty,\\ \Delta & \text{if} \quad \tau(\omega)=\infty. \end{cases}$$

On $S\cup\{\Delta\}$, we consider the sigma algebra $\mathcal{S}\cup\{A\cup\{\Delta\}: A\in\mathcal{S}\}$.

Lemma 3.16. *Let X be progressive and let τ be a stopping time. Then X_τ is $\mathcal{F}_\tau$–measurable.*

Proof. We first show that $\{\omega: \tau(\omega)\leqslant t, X_\tau(\omega)\in A\}\in\mathcal{F}_t$ for all $A\in\mathcal{S}$ and $t\geqslant 0$. We introduce

$$\Phi: \{\tau\leqslant t\}\to\Omega\times[0,t] \qquad \Phi(\omega)=(\omega,\tau(\omega)),$$

which is $(\mathcal{F}_t\cap\{\tau\leqslant t\},\mathcal{F}_t\otimes\mathcal{B}([0,t]))$–measurable, where

$$\mathcal{F}_t\cap\{\tau\leqslant t\}=\{A\cap\{\tau\leqslant t\}: A\in\mathcal{F}_t\}\subset\mathcal{F}_t.$$

We also introduce

$$\Psi: \Omega\times[0,t]\to S, \qquad \Psi(\omega,s)=X_s(\omega),$$

which by assumption is $(\mathcal{F}_t\otimes\mathcal{B}([0,t]),\mathcal{S})$–measurable.

So $X_\tau|_{\{\tau\leqslant t\}}$ is a composition of measurable maps,

$$X_\tau|_{\{\tau\leqslant t\}}=\Psi\circ\Phi,$$

and thus it is $(\mathcal{F}_t\cap\{\tau\leqslant t\},\mathcal{S})$–measurable. Therefore,

$$\{\omega: \tau(\omega)\leqslant t, X_\tau(\omega)\in A\}=\{\omega\in\{\tau\leqslant t\}: X_\tau(\omega)\in A\}\in\mathcal{F}_t\cap\{\tau\leqslant t\}\subset\mathcal{F}_t.$$

It remains to show that $\{X_\tau\in A\}$ is $\mathcal{F}$-measurable for all $A\in\mathcal{S}\cup\{B\cup\{\Delta\}: B\in\mathcal{S}\}$. For $A\in\mathcal{S}$, by the preceding we have

$$\{X_\tau\in A\}=\{X_\tau\in A,\tau<\infty\}=\bigcup_{n\in\mathbb{N}}\underbrace{\{X_\tau\in A,\tau\leqslant n\}}_{\in\mathcal{F}_n}\in\mathcal{F}.$$

For $A=B\cup\{\Delta\}$ we have

$$\{X_\tau\in A\}=\{X_\tau\in B,\tau<\infty\}\cup\{\tau=\infty\}\in\mathcal{F}.$$ □

As a consequence of Lemma 3.16, we can deduce that *stopped progressive processes* are still progressive.

Lemma 3.17. *Let X be progressive and τ be a stopping time. Then the* stopped process

$$X_t^\tau(\omega):=X_{t\wedge\tau(\omega)}(\omega),$$

usually abbreviated as $X_t^\tau=X_{t\wedge\tau}$, is also a progressive process.

Proof. See Exercise Sheet 4. □

Another nice property of progressive processes is that, under strong assumptions on the filtration, all entry times are stopping times:

Theorem 3.18. (Debut theorem) *Let $\mathbb{F}$ be a filtration satisfying the usual conditions, let X be $\mathbb{F}$-progressive with values in $(S,\mathcal{S})$, and let $A\in\mathcal{S}$. Then the entry time τ_A is a stopping time.*

Proof. We do not prove this result here. It relies on a deep theorem from measure theory, the so called *Section Theorem*. For a proof see Dellacherie-Meyer [4], Theorem IV.50, or Revuz-Yor [23], Theorem I.4.15. □

An alternative way of stating the Debut theorem is as follows: Let X be a progressively measurable process in an arbitrary filtration (not necessarily satisfying the usual conditions) and let $A \in \mathcal{S}$. Then τ_A is a stopping time with respect to $\mathbb{F}^{+,\mathbb{P}}$, no matter which probability measure $\mathbb{P}$ we choose for completion. This points to the role played by so called "universal completions" of filtrations.

3.3 Applications to Brownian motion

Definition 3.19. (d-dimensional Brownian motion, $\mathbb{F}$-Brownian motion) *Let $B = (B^1, \ldots, B^d)$ be a stochastic process.*

*i. B is called a d-*dimensional Brownian motion *if the B^j, $j = 1, \ldots, d$, are independent (1-dimensional) Brownian motions.*

ii. Let B be a d-dimensional Brownian motion and let $\mathbb{F}$ be a filtration. B is called a (d-dimensional) $\mathbb{F}$-Brownian motion if it is adapted to $\mathbb{F}$ and if for all $t \geqslant 0$ the process $(B_{t+s} - B_t)_{s \geqslant 0}$ is independent of $\mathcal{F}_t$.

If B is a Brownian motion, then it is obviously an $\mathbb{F}^B$-Brownian motion. The reason for introducing the notion of an $\mathbb{F}$-Brownian motion is that it is often desirable and indeed possible to take $\mathbb{F}$ larger than $\mathbb{F}^B$. Think for example of a two-dimensional Brownian motion $B = (B^1, B^2)$. Then B^1 is a one-dimensional $\mathbb{F}^B$-Brownian motion, although $\mathbb{F}^B$ is larger than $\mathbb{F}^{B^1}$.

Theorem 3.20. (Strong Markov property of Brownian motion) *Let $\mathbb{F}$ be a filtration and let B be a d-dimensional $\mathbb{F}$-Brownian motion. Then for any finite stopping time τ the process $B^{(\tau)} = (B_{\tau+t} - B_\tau)_{t \geqslant 0}$ is a d-dimensional Brownian motion and independent of $\mathcal{F}_{\tau+}$.*

In particular, an $\mathbb{F}$-Brownian motion is also an $\mathbb{F}^+$-Brownian motion and a $\mathbb{F}^{+,\mathbb{P}}$-Brownian motion.

Exercise. You might remember the strong Markov property of Markov chains. Does the Brownian motion also satisfy an analogous version of the strong Markov property?

Proof. The process $B^{(\tau)}$ is continuous by definition, so it suffices to show that it is a pre-Brownian motion and independent of $\mathcal{F}_{\tau+}$. For that purpose it suffices to show that

$$\mathbb{E}\,[\mathbb{1}_A \mathbb{1}_C(B^{(\tau)})] = \mathbb{P}(A)\mathbb{E}[\mathbb{1}_C(B)] \tag{3.3}$$

for all $A \in \mathcal{F}_{\tau+}$ and all $C \in \mathcal{B}(\mathbb{R}^d)^{\otimes \mathbb{R}_+}$. By Dynkin's $\pi - \lambda$ theorem it suffices to consider $C = \{f : \mathbb{R}_+ \to \mathbb{R}^d | f(t_1) \in C_1, \ldots, f(t_k) \in C_k\}$ for $k \in \mathbb{N}$ and closed $C_1, \ldots, C_k \subset \mathbb{R}^d$. Moreover, with $C = C_1 \times \cdots \times C_k$ we have for all $x \in \mathbb{R}^k$

$$\mathbb{1}_C(x) = \lim_{n \to \infty} (1 - n \cdot d(x, C)) \vee 0,$$

and since the right hand side is bounded and continuous in x we can replace $\mathbb{1}_C$ by $\varphi(B_{t_1}, \ldots, B_{t_k})$ for a continuous bounded function $\varphi \in C_b((\mathbb{R}^d)^k, \mathbb{R})$.

Next, we approximate τ from above by

$$\tau_n := \sum_{k=0}^{n2^n - 1} (k+2) 2^{-n} \mathbb{1}_{\{\tau \in [k2^{-n}, (k+1)2^{-n})\}} + \infty \mathbb{1}_{\{\tau \geqslant n\}},$$

so that τ_n takes only finitely many values (for simplicity we write $s_1 < \ldots < s_m$ for these values) and $\tau_n \geqslant \tau + 2^{-n}$ while $\lim_{n\to\infty}\tau_n = \tau$. By Lemma 3.8, vi., we have $\mathcal{F}_{\tau+} \subset \mathcal{F}_{\tau+2^{-n}} \subset \mathcal{F}_{\tau_n}$, and therefore $A \cap \{\tau_n = s_j\} \in \mathcal{F}_{s_j}$, which leads to

$$\begin{aligned}\mathbb{E}\,[\mathbb{1}_{\{\tau_n<\infty\}}\mathbb{1}_A\varphi(B^{(\tau_n)}_{t_1},\ldots,B^{(\tau_n)}_{t_k})] &= \sum_{j=1}^{m}\mathbb{E}\,[\mathbb{1}_{\{\tau_n=s_j\}}\mathbb{1}_A\varphi(B^{(\tau_n)}_{t_1},\ldots,B^{(\tau_n)}_{t_k})]\\ &= \sum_{j=1}^{m}\mathbb{E}\,[\mathbb{1}_{A\cap\{\tau_n=s_j\}}\varphi(B_{s_j+t_1}-B_{s_j},\ldots,B_{s_j+t_k}-B_{s_j})]\\ &= \sum_{j=1}^{m}\mathbb{E}[\mathbb{1}_{A\cap\{\tau_n=s_j\}}]\mathbb{E}[\varphi(B_{t_1},\ldots,B_{t_k})]\\ &= \mathbb{P}(A\cap\{\tau_n<\infty\})\mathbb{E}[\varphi(B_{t_1},\ldots,B_{t_k})]\end{aligned}$$

where in the penultimate step we used that $(B_{s_j+t}-B_{s_j})_{t\geqslant 0}$ is a Brownian motion and independent of $\mathcal{F}_{s_j}$ (Proposition 2.11). Recall that τ is finite. So letting n tend to infinity, the left hand side converges to $\mathbb{E}\,[\mathbb{1}_A\varphi(B^{(\tau)}_{t_1},\ldots,B^{(\tau)}_{t_k})]$ (recall that φ is continuous and bounded) and the right hand side to $\mathbb{P}(A)\mathbb{E}[\varphi(B_{t_1},\ldots,B_{t_k})]$. □

∘

——— End of the lecture on November 13 ———

Remark 3.21. Let τ be a not necessarily finite stopping time with $\mathbb{P}(\tau<\infty)>0$. Then the proof of Theorem 3.20 still shows that

$$\mathbb{P}(A\cap\{\tau<\infty\}\cap\{B^{(\tau)}\in C\}) = \mathbb{P}(A\cap\{\tau<\infty\})\mathbb{P}(B\in C)$$

for all $A\in\mathcal{F}_{\tau+}$ and all $C\in\mathcal{B}(\mathbb{R}^d)^{\otimes\mathbb{R}_+}$. Dividing both sides by $\mathbb{P}(\tau<\infty)$, we get

$$\mathbb{P}(A\cap\{B^{(\tau)}\in C\}|\tau<\infty) = \mathbb{P}(A|\tau<\infty)\mathbb{P}(B\in C),$$

which shows that under the conditional probability measure $\mathbb{P}(\cdot|\tau<\infty)$ the process $(B_{\tau+t}-B_\tau)_{t\geqslant 0}$ (defined for example as 0 on the set $\tau=\infty$) is a Brownian motion independent of $\mathcal{F}_{\tau+}$. In particular, the statement of Theorem 3.20 still holds for stopping times that are almost surely finite.

Corollary 3.22. (Blumenthal's 0-1 law) *Let B be a (d-dimensional) Brownian motion and let $A\in\mathcal{F}^B_{0+}$. Then $\mathbb{P}(A)\in\{0,1\}$.*

Intuitively, Blumenthal's 0-1 law says that we cannot learn anything new by peaking a little bit into the future of the Brownian motion.

Proof. If $A\in\mathcal{F}^B_{0+}\subset\mathcal{F}^B_\infty$, then there exists $C\in\mathcal{B}(\mathbb{R}^d)^{\otimes\mathbb{R}_+}$ such that $A=\{\omega: B(\omega)\in C\}$. The strong Markov property applied with $\tau=0$ gives

$$\mathbb{P}(A)=\mathbb{P}(A\cap A)=\mathbb{P}(A\cap\{B\in C\})=\mathbb{P}(A\cap\{B^{(0)}\in C\})=\mathbb{P}(A)\mathbb{P}(B\in C)=\mathbb{P}(A)^2,$$

and the result follows because the only numbers a with $a^2=a$ are $\{0,1\}$ □

Exercise. Let B be a Brownian motion and let $\mathcal{F}^B_{t,\infty}:=\sigma(B_s: s\geqslant t)$. Let $\mathcal{T}^B=\bigcap_{t>0}\mathcal{F}^B_{t,\infty}$ be the tail σ-algebra of B. Show that every $A\in\mathcal{T}^B$ satisfies $\mathbb{P}(A)\in\{0,1\}$. Compare this result to the Kolmogorov 0-1 law (cf. Exercise Sheet 4).

Hint: Consider the time-inversed Brownian motion $\tilde{B}_t = tB_{1/t}$.

Corollary 3.23. *Let B be a Brownian motion. Then with probability 1 we have for all $\varepsilon>0$*

$$\sup_{s\in[0,\varepsilon]} B_s>0, \qquad \inf_{s\in[0,\varepsilon]} B_s<0.$$

Moreover, if for $a\in\mathbb{R}$ *we set* $\tau_a=\inf\{t\geqslant 0\colon B_t=a\}$*, then* $\tau_a<\infty$ *for all* $a\in\mathbb{R}$ *with probability* 1*, so that in particular*

$$-\infty=\liminf_{t\to\infty} B_t<\limsup_{t\to\infty} B_t=\infty. \tag{3.5}$$

At first sight, it is not obvious whether $\sup_{s\in[0,\varepsilon]} B_s$ is measurable. But recall that B is continuous, and therefore $\sup_{s\in[0,\varepsilon]} B_s=\sup_{s\in[0,\varepsilon]\cap\mathbb{Q}} B_s$. In the sequel we will often implicitly use this kind of argument when dealing with continuous (or right- or left-continuous) processes. The last conclusion (3.5) above might not come as a surprise, given our previous discussions and results around Corollary 2.15 and Theorem 2.16; but it is nice to see how Blumenthal's 0-1 law provides a short, elegant proof of that fact.

Proof of Corollary 3.23. Notice that

$$\begin{aligned}\Gamma:=\left\{\omega\in\Omega\colon \sup_{s\in[0,\varepsilon]} B_s(\omega)>0 \text{ for all } \varepsilon>0\right\} &=\bigcap_{n\geqslant 1}\left\{\omega\in\Omega\colon \sup_{s\in[0,1/n]} B_s(\omega)>0\right\}\\ &=\bigcap_{n\geqslant N}\left\{\omega\in\Omega\colon \sup_{s\in[0,1/n]} B_s(\omega)>0\right\}\end{aligned}$$

for any fixed $N\in\mathbb{N}$. In particular, $\Gamma\in\mathcal{F}_{1/N}$ for all $N\in\mathbb{N}$ and so $\Gamma\in\mathcal{F}_{0+}$; by Blumenthal's 0-1 law, $\mathbb{P}(\Gamma)$ must be 0 or 1. On the other hand

$$\begin{aligned}\mathbb{P}(\Gamma) &=\mathbb{P}\left(\bigcap_{n\in\mathbb{N}}\left\{\sup_{s\in[0,1/n]} B_s>0\right\}\right)\\ &=\lim_{n\to\infty}\mathbb{P}\left(\sup_{s\in[0,1/n]} B_s>0\right)\\ &\geqslant\limsup_{n\to\infty}\mathbb{P}\left(B_{1/n}>0\right)=\frac{1}{2}\end{aligned}$$

since $B_{1/n}$ is a centered Gaussian, thus its law is symmetric. Therefore $\mathbb{P}(\Gamma)=1$.

Replacing B by the Brownian motion $-B$, we get the statement about the infimum.

If there exists $a>0$ with $\tau_a=\infty$, then $\sup_{s\geqslant 0} B_s<\infty$. But

$$\begin{aligned}\mathbb{P}\left(\sup_{s\geqslant 0} B_s=\infty\right) &= \lim_{n\to\infty}\mathbb{P}\left(\sup_{s\geqslant 0} B_s>n\right)\\ &= \lim_{n\to\infty}\mathbb{P}\left(\sup_{s\geqslant 0} n^{-2}B_s>\frac{1}{n}\right)\\ &= \lim_{n\to\infty}\mathbb{P}\left(\sup_{s\geqslant 0} \tilde{B}_{n^{-4}s}>\frac{1}{n}\right)\\ &= \mathbb{P}\left(\sup_{s\geqslant 0}\tilde{B}_s>0\right)=1,\end{aligned}$$

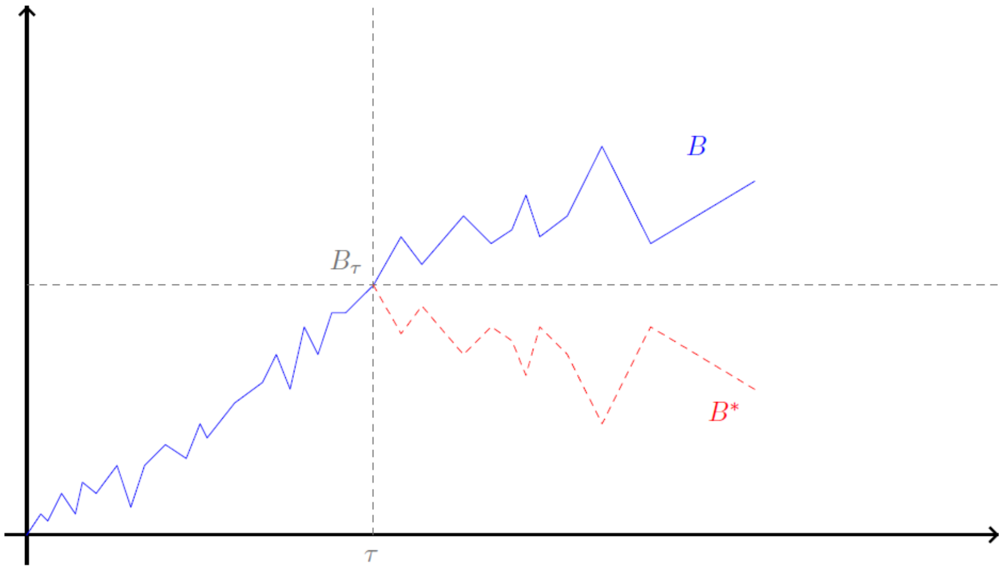
where $\tilde{B}_s:=n^{-2}B_{n^4 s}$ is a new Brownian motion (so that $n^{-2}B_s=\tilde{B}_{n^{-4}s}$) and we used the scaling properties of the Brownian motion, as well as the fact that the supremum is taken over the whole real line $[0,+\infty)$ (so that computing it over the variable $\tilde{s}=n^{-4}s$ instead of the original s doesn't change its value).

As before, up to replacing B with $-B$, we find that $\mathbb{P}(\inf_{s\geqslant 0} B_s=-\infty)=1$ as well. Since B is continuous, so that its sup and inf computed on compact time intervals $[0,T]$ are necessarily finite (for fixed ω), we deduce also the claim about liminf and limsup. $\square$

Proposition 3.24. (Reflection principle) *Let B be a one-dimensional Brownian motion and let τ be a finite stopping time. Then the process*

$$B_t^* := B_t \mathbb{1}_{\{t \leqslant \tau\}} + (2B_\tau - B_t)\mathbb{1}_{\{t > \tau\}}$$

is also a Brownian motion.



Proof. By the strong Markov property, both

$$B^{(\tau)} = (B_{t+\tau} - B_\tau)_{t \geqslant 0}, \qquad -B^{(\tau)} = (B_\tau - B_{\tau+t})_{t \geqslant 0}$$

are Brownian motions that are independent of $\mathcal{F}_\tau$. Using that $(\tau, B_{t \wedge \tau})$ is $\mathcal{F}_\tau$–measurable (by Corollary 3.16) and thus is independent of these two processes, we get that the "glued" process

$$B_t \;=\; B_{t\wedge\tau} + B^{(\tau)}_{t-\tau}\mathbb{1}_{\{t>\tau\}}$$

has the same law as

$$\begin{aligned} B_{t\wedge\tau} - B^{(\tau)}_{t-\tau}\mathbb{1}_{\{t>\tau\}} &= B_{t\wedge\tau} + (B_\tau - B_{\tau+(t-\tau)})\mathbb{1}_{\{t>\tau\}} \\ &= B_t\mathbb{1}_{\{t\leqslant\tau\}} + B_\tau\mathbb{1}_{\{t>\tau\}} + (B_\tau - B_{\tau+(t-\tau)})\mathbb{1}_{\{t>\tau\}} \\ &= B_t^* \end{aligned}$$

and this concludes the proof.

You might (and probably should) feel a bit uncomfortable about plugging $t-\tau$ into $B^{(\tau)}$ resp. $-B^{(\tau)}$ and claiming that we still have the same law. To rigorously see that everything works, we could first assume that τ only takes finitely many values, verify everything there "by hand" conditioning on the values τ can attain, and then use a limiting argument as in the proof of the strong Markov property. Due to the similarity of such passages with those in the proof of Theorem 3.20, we omit them here for simplicity.

∘

□

Corollary 3.25. *Let B be a Brownian motion and let $S_t = \max_{s\in[0,t]} B_s$. Then*

$$\mathbb{P}(S_t \geqslant a) = 2\mathbb{P}(B_t \geqslant a) = \mathbb{P}(|B_t| \geqslant a)$$

for all $a \geqslant 0$.

Proof. Exercise Sheet 5. □

Exercise.

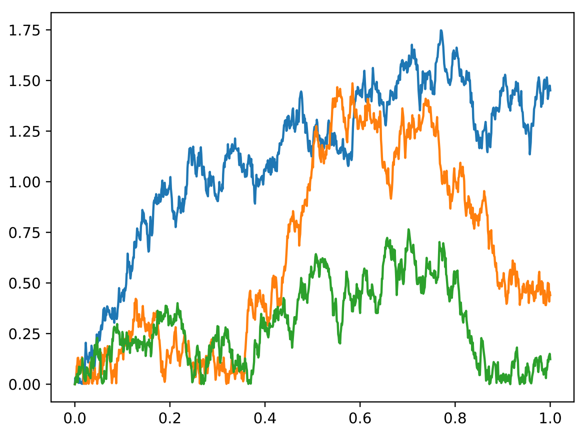
i. Do $(S_t)_{t\geqslant 0}$ and $(|B_t|)_{t\geqslant 0}$ have the same law?

ii. Show that $\mathbb{E}[e^{\lambda S_t}] < \infty$ for all $\lambda \in \mathbb{R}$ and $t > 0$; show that, for any fixed $t > 0$, there exists $\lambda^* = \lambda^*(t) > 0$ small enough such that $\mathbb{E}[e^{\lambda^* |S_t|^2}] < \infty$.

Here are simulations of B, $|B|$ and M. We start with $|B|$, then we draw B so that we can better compare it with M (note that M is the maximum value of B and not the maximum value of $|B|$). Here is $|B|$:

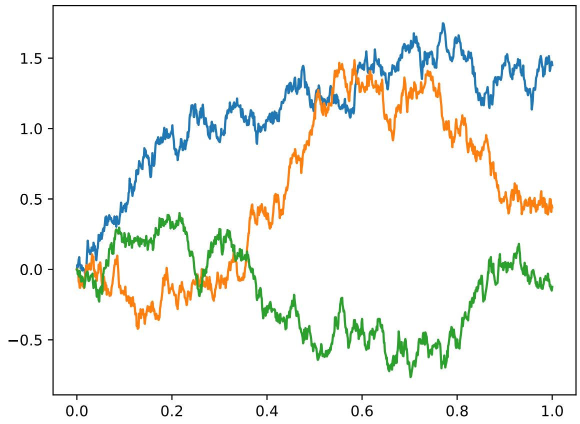
```
Python 3.7.4 [/opt/anaconda3/bin/python3]
Python plugin for TeXmacs.
Please see the documentation in Help -> Plugins -> Python
```

```
>>> import numpy as np
    import matplotlib.pyplot as plt

    T, h = 1, 1e-3
    n = int(T/h)
    k = 3

    time = np.arange(0,T+h,h)
    dB = np.sqrt(h)*(np.random.randn(k,n))
    BM = np.zeros((k,n+1))
    BM[:,1:] = np.cumsum(dB, axis=1)
    BM_norm = np.abs(BM)
    BM_max = np.maximum.accumulate(BM,axis=1)

    plt.clf()

    for i in range(k):
        plt.plot(time,BM_norm[i,:])

    pdf_out(plt.gcf())
```



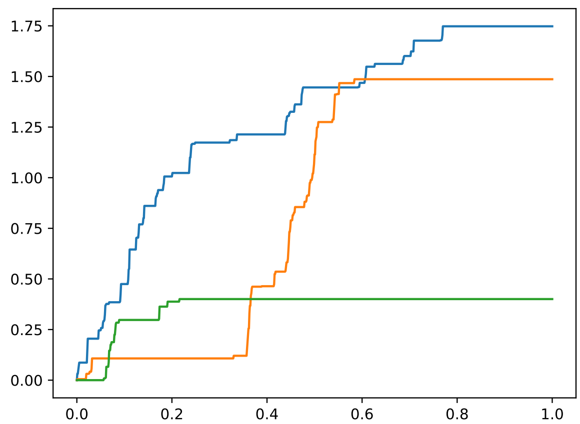
Now we plot the Brownian motion B:

```
>>> plt.clf()

    for i in range(k):
        plt.plot(time,BM[i,:])

    pdf_out(plt.gcf())
```



And here is M. Compare this with the plot of $|B|$ from above!

```
>>> plt.clf()

    for i in range(k):
        plt.plot(time,BM_max[i,:])

    pdf_out(plt.gcf())
```




```
>>>
```

Here is another example of two processes that have the same one-dimensional marginal distributions but which have different distributions as processes: If $X\sim\mathcal{N}(0;1)$, then for each fixed $t>0$ the random variable $\sqrt{t}\,X$ has the same distribution as B_t, but of course the processes $(\sqrt{t}\,X)_{t\geqslant 0}$ and $(B_t)_{t\geqslant 0}$ are nothing alike.

4 Martingales in continuous time

Throughout this section we fix a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$.

4.1 Path regularity

Definition 4.1. (Integrable) *A stochastic process* X *is called* integrable *if* $\mathbb{E}[|X_t|] < \infty$ *for all* $t \geqslant 0$. *For* $p > 0$ *we call* X p-integrable *if* $\mathbb{E}[|X_t|^p] < \infty$ *for all* $t \geqslant 0$. *For* $p = 2$ *we also say* square-integrable.

Definition 4.2. (Martingale) *An adapted, real-valued and integrable process* $X = (X_t)_{t \geqslant 0}$ *is called a*

i. martingale *if* $\mathbb{E}[X_t|\mathcal{F}_s] = X_s$ *for all* $0 \leqslant s \leqslant t$;

ii. supermartingale *if* $\mathbb{E}[X_t|\mathcal{F}_s] \leqslant X_s$ *for all* $0 \leqslant s \leqslant t$;

iii. submartingale *if* $\mathbb{E}[X_t|\mathcal{F}_s] \geqslant X_s$ *for all* $0 \leqslant s \leqslant t$.

Clearly, X is a martingale if and only if it is both a submartingale and a supermartingale; moreover X is a supermartingale if and only if $-X$ is a submartingale. Finally, notice how the martingale property may be rephrased as $\mathbb{E}[X_t - X_s|\mathcal{F}_s] = 0$ for all $0 \leqslant s \leqslant t$ (similarly for super- and sub-martingales, up to replacing $=$ with $\leqslant$ and $\geqslant$).

In the following we will state some of the next results only for supermartingales, but by the above similar variants can be immediately inferred for submartingales as well.

Example 4.3. (Brownian martingales) Let B be an $\mathbb{F}$-Brownian motions. Then:

i. B is a martingale:

$$\mathbb{E}[B_t|\mathcal{F}_s] = \mathbb{E}[B_t - B_s|\mathcal{F}_s] + B_s = \mathbb{E}[B_t - B_s] + B_s = B_s.$$

ii. $X_t = B_t^2 - t$, $t \geqslant 0$, is a martingale:

$$\mathbb{E}[X_t|\mathcal{F}_s] = \mathbb{E}[(B_t - B_s)^2 + 2(B_t - B_s)B_s + B_s^2|\mathcal{F}_s] - t = (t - s) + B_s^2 - t = X_s.$$

iii. For $\lambda \in \mathbb{R}$ the process $Y_t = e^{\lambda B_t - \lambda^2 t/2}$, $t \geqslant 0$, is a martingale:

$$\begin{aligned}\mathbb{E}[Y_t|\mathcal{F}_s] &= \mathbb{E}[e^{\lambda(B_t - B_s)}|\mathcal{F}_s]\, e^{\lambda B_s - \lambda^2 t/2} = \mathbb{E}[e^{\lambda(B_t - B_s)}]\, e^{\lambda B_s - \lambda^2 t/2} \\ &= e^{\lambda^2(t-s)/2}\, e^{\lambda B_s - \lambda^2 t/2} = Y_s;\end{aligned}$$

we used the formula $\mathbb{E}[e^{\lambda U}] = e^{\lambda^2\sigma^2/2}$ for the Laplace transform of $U \sim \mathcal{N}(0, \sigma^2)$.

Similarly (up to extending the definition of martingale to the complex-valued processes), for any $\lambda \in \mathbb{R}$, $Y_t = e^{i\lambda B_t + \lambda^2 t/2}$ is a martingale.

iv. If B, $\tilde{B}$ are independent $\mathbb{F}$-Brownian motions, then $U_t = B_t\tilde{B}_t$, $t \geqslant 0$, is a martingale:

$$\mathbb{E}[U_t|\mathcal{F}_s] = \mathbb{E}[(B_t - B_s)(\tilde{B}_t - \tilde{B}_s) + (B_t - B_s)\tilde{B}_s + B_s(\tilde{B}_t - \tilde{B}_s) + B_s\tilde{B}_s|\mathcal{F}_s] = U_s.$$

——— End of the lecture on November 14 ———

v. Let $f \in L^2(\mathbb{R}_+)$ and set $Z_t = \int_0^t f(s)\mathrm{d}B_s = \int_0^\infty \mathbb{1}_{[0,t]}(s) f(s)\mathrm{d}B_s$, where the right hand side is the Wiener integral; then Z is a martingale. Indeed, first assume that $f(t) = \sum_{k=0}^{n-1} x_k \mathbb{1}_{(t_k, t_{k+1})}(t)$, i.e. $f \in \mathcal{E}$ in the notation of Lemma 1.12. Then

$$Z_t = \sum_{k=0}^{n-1} x_k \int_0^\infty \mathbb{1}_{[0,t]}(s)\mathbb{1}_{(t_k, t_{k+1}]}(s)\mathrm{d}B_s = \sum_{k=0}^{n-1} x_k(B_{t_{k+1}\wedge t} - B_{t_k \wedge t}),$$

which follows by considering the cases $t < t_k$, $t \in [t_k, t_{k+1}]$ and $t > t_{k+1}$ separately. Therefore,

$$\begin{aligned} \mathbb{E}[Z_t|\mathcal{F}_s] &= \sum_{k=0}^{n-1} x_k \mathbb{E}[B_{t_{k+1}\wedge t} - B_{t_k\wedge t}|\mathcal{F}_s] \\ &= \sum_{k=0}^{n-1} x_k (B_{t_{k+1}\wedge s} - B_{t_k\wedge s}) = Z_s. \end{aligned}$$

For general $f \in L^2(\mathbb{R}_+)$ there exists a sequence $(f_n)_n \subset \mathcal{E}$ such that $\|f_n \mathbb{1}_{[0,t]} - f\mathbb{1}_{[0,t]}\|_{L^2(\mathbb{R}_+)} \to 0$, and since the Wiener integral is an isometry we get that $Z_t^n = \int_0^t f_n(s) \mathrm{d}B_s$ converges in $L^2(\mathbb{R}_+)$ to Z_t, and we just saw that Z^n is a martingale for each n. Since we can pull the L^2 limit into the conditional expectation, we get that Z is a martingale.

Example 4.4. (Poisson martingales) Let N be a Poisson process with intensity $\lambda > 0$. Then:

i. $(N_t - \lambda t)_{t \geqslant 0}$ is a martingale in the filtration $\mathbb{F} = \mathbb{F}^N$.

ii. Let $X_t = \sum_{k=1}^{N_t} Y_k$ be a compound Poisson process, where $(Y_k)_{k\in\mathbb{N}}$ is an i.i.d. sequence of integrable random variables that is independent of N and we set $m = \mathbb{E}[Y_1]$. Then the *compensated compound Poisson* process $\tilde{X}_t := X_t - \lambda m t$ is a martingale w.r.t. its canonical filtration (also, note that $\mathbb{F}^X = \mathbb{F}^{\tilde{X}}$). In particular, X is a martingale if $m = 0$.

iii. If additionally $a = \mathbb{E}[Y_1^2] < \infty$, then in the setting of the Point ii., $M_t := |\tilde{X}_t|^2 - \lambda a t$ is also a martingale w.r.t. $\mathbb{F}^X$.

All of the above statements were part of Exercise Sheet 3 (for point i. this follows from N being a Lévy process with $\mathbb{E}[N_t] = \lambda t$).

Exercise.

i. Let N be a Poisson process with intensity $\lambda > 0$ and let $\mu \in \mathbb{R}$. Show that

$$M_t = \exp(\mu N_t - \lambda t(e^{\mu} - 1)), \qquad t \geqslant 0,$$

is a martingale.

ii. Let N^1, N^2 be independent Poisson processes of the same intensity. Show that $N^1 - N^2$ is a martingale in the filtration $\mathbb{F}^{(N^1, N^2)}$.

Remark 4.5. If $(X_t)_{t\geqslant 0}$ is a martingale, then by the tower property of conditional expectiation we immediately have $\mathbb{E}[X_t] = \mathbb{E}[X_0]$ for all $t \geqslant 0$. Similarly, if X is a supermartingale (respectively submartingale) then $t \mapsto \mathbb{E}[X_t]$ is decreasing (resp. increasing).

Recall **conditional Jensen's inequality** (e.g. from Stochastics II): given a real-valued random variable Z and a **convex** function $\varphi: \mathbb{R} \to \mathbb{R}$ such that Z, $\varphi(Z)$ are integrable, it holds

$$\varphi(\mathbb{E}[Z|\mathcal{G}]) \leqslant \mathbb{E}[\varphi(Z)|\mathcal{G}]$$

for any σ-algebra $\mathcal{G} \subset \mathcal{F}$.

Remark 4.6. It follows from conditional Jensen's inequality that:

i. If X is a martingale, $\varphi: \mathbb{R} \to \mathbb{R}$ is convex and $(\varphi(X_t))_{t \geqslant 0}$ is integrable, then $\varphi(X)$ is a submartingale.

ii. If X is a submartingale, φ is convex and increasing, and $\varphi(X)$ is integrable, then $\varphi(X)$ is a submartingale.

iii. In particular, $|X|^p$ is a submartingale if X is a p-integrable martingale and $p \geqslant 1$, and $X^+ = X \vee 0$ is a submartingale if X is a submartingale.

Exercise. Prove/convince yourself of the above statements.

A priori, martingales have no path regularity. Our first aim is to show that they admit nice modifications. In general, in the upcoming results, our strategy will be to leverage as much as possible on the results for discrete-time martingales you have already seen in Stochastics-II, and transfer them to the continuous-time case. To that end, let us recall some discrete-time results first.

Definition 4.7. (Upcrossings) *Let $I \subset \mathbb{R}_+$ and $f: I \to \mathbb{R}$. For $a < b$, the number of upcrossings of f across the interval $[a, b]$ in I is the supremum over all n for which there exist times $s_k, t_k \in I$, $k = 1, \ldots, n$, such that $s_1 < t_1 < s_2 < t_2 < \ldots < s_n < t_n$ with $f(s_k) \leqslant a$ and $f(t_k) \geqslant b$ for all $k = 1, \ldots, n$. We denote it with*

$$U([a, b]; I; f).$$

Lemma 4.8. (Doob's upcrossing inequality) *Let $(X_n)_{n \in \mathbb{N}_0}$ be a discrete time supermartingale. Then we have for all $a < b \in \mathbb{R}$*

$$\mathbb{E}[U([a, b]; \{0, \ldots, n\}; X)] \leqslant \frac{\mathbb{E}[(X_n - a)^-]}{b - a}, \qquad \mathbb{E}[U([a, b]; \mathbb{N}_0; X)] \leqslant \sup_{n \in \mathbb{N}} \frac{\mathbb{E}[(X_n - a)^-]}{b - a}.$$

Lemma 4.9. (Doob's inequalities, discrete time case) *Let $(X_n)_{n \in \mathbb{N}_0}$ is a discrete time martingale, then for all $\lambda > 0$ and $n \in \mathbb{N}$:*

$$\mathbb{P}\left(\max_{k \in \{0, \ldots, n\}} |X_k| \geqslant \lambda \right) \leqslant \frac{1}{\lambda} \mathbb{E}[|X_n|].$$

Moreover for all $p \in (1, \infty)$ we have

$$\mathbb{E}\left[\max_{k \in \{0, \ldots, n\}} |X_k|^p \right] \leqslant \left(\frac{p}{p - 1} \right)^p \mathbb{E}[|X_n|^p].$$

It is also useful to shortly recall the concept of *uniform integrability* of a family of random variables and its properties.

Definition 4.10. *A family of real-valued random variables $(Y_j)_{j \in J}$ is **uniformly integrable** if*

$$\lim_{M \to \infty} \sup_{j \in J} \mathbb{E}[|Y_j| \mathbb{1}_{\{|Y_j| \geqslant M\}}] = 0.$$

Remark 4.11. Recall the following facts about uniform integrability:

i. If $(Y_j)_{j \in J}$ is uniformly integrable, then it is bounded in L^1: $\sup_{j \in J} \mathbb{E}[|Y_j|] < \infty$; the converse is not true.

ii. If $(Y_j)_{j\in J}$ is bounded in L^p for some $p>1$, i.e. $\sup_{j\in J}\mathbb{E}[|Y_j|^p]<\infty$, then it is uniformly integrable.

iii. Given a sequence $(Y_n)_{n\in\mathbb{N}}$, $Y_n\to Y$ in L^1 if and only if $Y_n\to Y$ in probability and $(Y_n)_{n\in\mathbb{N}}$ is uniformly integrable.

iv. If $Y\in L^1$ and $(\mathcal{G}_j)_{j\in J}$ is a family of σ-algebras, then $(\mathbb{E}[Y|\mathcal{G}_j])_{j\in J}$ is uniformly integrable.

With these preparations, we can now show that martingales have càdlàg modifications, the proof of which crucially relies on Lemma 4.8.

Theorem 4.12. *Let X be a martingale and assume that $\mathbb{F}$ satisfies the usual conditions. Then X has an adapted càdlàg modification which still is a martingale.*

Proof. (The proof was only sketched in the lectures and is not examinable)

1. We first show that X restricted to $\mathbb{Q}_+$ almost surely admits limits from the left and right: Let $k\in\mathbb{N}$ and let $(I_n)_{n\in\mathbb{N}}$ be an increasing sequence of finite subsets of $\mathbb{Q}_+\cap[0,k]$ such that $\bigcup_n I_n=\mathbb{Q}_+\cap[0,k]$. We also assume that $k\in I_n$ for all n. Then by the monotone convergence theorem, together with Doob's upcrossing lemma, for all $a,b\in\mathbb{R}$, $a<b$ and all $n\in\mathbb{N}$ it holds that

$$\mathbb{E}\left[U([a,b];\mathbb{Q}_+\cap[0,k];X)\right]=\lim_{n\to\infty}\mathbb{E}[U([a,b];I_n;X)]\leqslant\frac{\mathbb{E}[(X_k-a)^-]}{b-a}<\infty.$$

Similarly, by applying Lemma 4.9 and passing to the limit, we get with Doob's maximal inequality for any $\lambda>0$:

$$\mathbb{P}\left(\sup_{t\in\mathbb{Q}_+\cap[0,k]}|X_t|\geqslant\lambda\right)=\lim_{n\to\infty}\mathbb{P}\left(\sup_{t\in I_n}|X_t|\geqslant\lambda\right)\leqslant\frac{\mathbb{E}[|X_k|]}{\lambda}.$$

Therefore, there exists a null set N such that for all $\omega\in N^c$

$$U([a,b];\mathbb{Q}_+\cap[0,k];X(\omega))<\infty\qquad\text{for all } a<b\in\mathbb{Q}, k\in\mathbb{N},$$

and

$$\sup_{t\in\mathbb{Q}_+\cap[0,k]}|X_t(\omega)|<\infty\qquad\text{for all } k\in\mathbb{N}.$$

From here it is not hard to see that for $\omega\in N^c$ the limits

$$X_{t+}(\omega):=\lim_{s\downarrow t,s\in\mathbb{Q}_+}X_s(\omega)\in\mathbb{R}\qquad\text{and}\qquad X_{t-}(\omega):=\lim_{s\uparrow t,s\in\mathbb{Q}_+}X_s(\omega)\in\mathbb{R}$$

exist for all $t\geqslant 0$ respectively $t>0$.

2. With the null set N and X_{t+} as in step i. we define

$$\tilde{X}_t(\omega)=\begin{cases}X_{t+}(\omega), & \omega\in N^c,\\ 0, & \omega\in N.\end{cases}$$

$\tilde{X}$ is right-continuous by construction, and it also has left limits: For $\omega\in N$ this is clear, so let $\omega\in N^c$, let $t_n\uparrow t$, and consider for all $n\in\mathbb{N}$ a point $s_n\in(t_n,t)\cap\mathbb{Q}$ such that $|X_{t_n+}(\omega)-X_{s_n}(\omega)|<1/n$. Then

$$\lim_{n\to\infty}\tilde{X}_{t_n}(\omega)=\lim_{n\to\infty}X_{t_n+}(\omega)=\lim_{n\to\infty}X_{s_n}(\omega)\overset{s_n\uparrow t,(s_n)\subset\mathbb{Q}}{=}X_{t-}(\omega).$$

Moreover, $\tilde{X}$ is adapted because our filtration satisfies the usual conditions. The family

$$(X_s)_{s\in[t,t+1]\cap\mathbb{Q}_+} = (\mathbb{E}[X_{t+1}|\mathcal{F}_s])_{s\in[t,t+1]\cap\mathbb{Q}_+}$$

is uniformly integrable, because it is given by conditional expectations of $X_{t+1}\in L^1$ (cf. Remark 4.11-iv.). Therefore, we get almost surely

$$\tilde{X}_t = \mathbb{E}[\tilde{X}_t|\mathcal{F}_t] = \mathbb{E}\Big[\lim_{s\downarrow t, s\in\mathbb{Q}_+} X_s\Big|\mathcal{F}_t\Big] = \lim_{s\downarrow t, s\in\mathbb{Q}_+} \mathbb{E}[X_s|\mathcal{F}_t] = \lim_{s\downarrow t, s\in\mathbb{Q}_+} X_t = X_t,$$

i.e. $\tilde{X}$ is a modification of X. □

Remark 4.13. More generally, one can show that any supermartingale X in a filtration satisfying the usual conditions and for which $t\mapsto\mathbb{E}[X_t]$ is right-continuous has a càdlàg adapted modification; see Theorem 3.17 of Le Gall [16].

Exercise. The condition that $t\mapsto\mathbb{E}[X_t]$ is right-continuous is necessary: Find a supermartingale which does not have a càdlàg modification.

——— End of the lecture on November 20 ———

Theorem 4.14. (Martingale convergence theorem) *Let X be a càdlàg supermartingale with $\sup_{t\geqslant 0}\mathbb{E}[X_t^-] < \infty$.*

Then there exists a random variable $X_\infty\in L^1$ with $\lim_{t\to\infty} X_t = X_\infty$ almost surely.

If $(|X_t|^p)_{t\geqslant 0}$ is uniformly integrable, for $p\geqslant 1$, then X_t also converges in L^p to X_∞.

Remark 4.15. Notice that since X is a supermartingale,

$$\sup_{t\geqslant 0}\mathbb{E}[X_t^-] < \infty \quad \Leftrightarrow \quad \sup_{t\geqslant 0}\mathbb{E}[|X_t|] < \infty.$$

One implication is obvious; for the other, we have $\mathbb{E}[X_t^+] - \mathbb{E}[X_t^-] = \mathbb{E}[X_t] \leqslant \mathbb{E}[X_0]$, so that

$$\sup_{t\geqslant 0}\mathbb{E}[X_t^+] \leqslant \sup_{t\geqslant 0}\mathbb{E}[X_t^-] + \mathbb{E}[X_0] \quad \Rightarrow \quad \sup_{t\geqslant 0}\mathbb{E}[|X_t|] \leqslant \sup_{t\geqslant 0}\mathbb{E}[X_t^+] + \sup_{t\geqslant 0}\mathbb{E}[X_t^-] < \infty.$$

Proof. By an approximation argument of $\mathbb{Q}_+$ via finite sets we get (similarly as in Theorem 4.12) for all $a < a' < b' < b \in \mathbb{R}$:

$$\mathbb{E}[U([a,b];\mathbb{R}_+;X)] \leqslant \mathbb{E}[U([a',b'];\mathbb{Q}_+;X)] \leqslant \frac{1}{b'-a'}\left(\sup_{t\geqslant 0}\mathbb{E}[X_t^-] + |a'|\right) < \infty,$$

where the first inequality uses that X is càdlàg. So almost surely $U([a,b];\mathbb{R}_+;X) < \infty$ for all $a,b\in\mathbb{Q}$ with $a<b$, which shows that X_t converges almost surely to a limit X_∞ with values in $[-\infty,\infty]$. Since X_∞ is also the limit of $(X_n)_{n\in\mathbb{N}}$, we get $X_\infty\in L^1$ from the discrete time version of the martingale convergence theorem (Stochastics II).

Suppose now that $(|X_t|^p)_{t\geqslant 0}$ is uniformly integrable; then $|X_t|^p\to|X_\infty|^p$ $\mathbb{P}$-a.s. (thus also in probability), which together with Remark 4.11-iii. implies $|X_t|^p\to|X_\infty|^p$ in L^1, which in turn implies that $X_t\to X_\infty$ in L^p.

○

□

Exercise. Show that any positive càdlàg supermartingale almost surely converges.

Example 4.16.

i. Without a condition like $\sup_{t\geqslant 0}\mathbb{E}[X_t^-]<\infty$, convergence can fail: For example, the Brownian motion almost surely does not converge because it is unbounded from below and from above.

ii. If X is a martingale, then $\mathbb{E}[X_t]=\mathbb{E}[X_0]$ for all $t\geqslant 0$. But even if X converges, we may have $\mathbb{E}[X_\infty]\neq\mathbb{E}[X_0]$. Consider for example $X_t=\exp(B_t-t/2)$, $t\geqslant 0$, which is a positive martingale (cf. Example 4.3-iii.) and therefore it almost surely converges. We know from Corollary 2.15 that for $\alpha\in(1/2,1)$ and for almost every $\omega\in\Omega$ there exists $C(\omega)>0$ with $|B_t(\omega)|\leqslant C(\omega)t^\alpha$ for all $t\geqslant 0$, so that

$$0\leqslant\limsup_{t\to\infty}X_t(\omega)\leqslant\limsup_{t\to\infty}\exp(C(\omega)t^\alpha-t/2)=0,$$

while $\mathbb{E}[X_t]=1$ for all $t\geqslant 0$.

Theorem 4.17. *For a càdlàg martingale X the following conditions are equivalent:*

i. X is uniformly integrable (we say X is a uniformly integrable martingale*);*

ii. X_t converges almost surely and in L^1 to a limit X_∞ as $t\to\infty$;

iii. there exists $Y\in L^1$ with $X_t=\mathbb{E}[Y|\mathcal{F}_t]$ for all $t\geqslant 0$ (we say that X is a closed martingale*).*

In that case we can take $Y=X_\infty$, and for general Y we always have $X_\infty=\mathbb{E}[Y|\mathcal{F}_\infty]$.

Proof. This follows line-by-line by the same arguments as in the discrete time case, see Stochastics II. □

Remark 4.18. Given an integrable r.v. Y and a filtration $\mathbb{F}$, $X_t=\mathbb{E}[Y|\mathcal{F}_t]$ always defines a martingale, by the tower property of conditional expectation; moreover X is uniformly integrable, by Remark 4.11-iv. If additionally $\mathbb{F}$ satisfies the usual assumptions, then we can invoke Theorem 4.12 to deduce that, up to a modification, X has càdlàg paths.

Example 4.19. Let $(X_t)_{t\geqslant 0}$ be a martingale and fix a deterministic $T\in(0,+\infty)$; set

$$X_t^T=X_{t\wedge T}=X_t\mathbb{1}_{t\leqslant T}+X_T\mathbb{1}_{t>T}.$$

It's easy to check that X^T is also a martingale, that it is uniformly integrable, since $X_t=\mathbb{E}[X_T|\mathcal{F}_t]$ for all $t\geqslant 0$.

○

4.2 Martingale inequalities and stopping theorems

Here we transfer some useful properties of discrete time martingales to continuous time.

Theorem 4.20. (Doob's martingale inequalities) *Let $T\in[0,+\infty)$ and $\lambda>0$.*

i. If X is a càdlàg submartingale, then

$$\mathbb{P}\left(\sup_{t\in[0,T]}X_t\geqslant\lambda\right)\leqslant\frac{1}{\lambda}\mathbb{E}[X_T^+],\qquad\mathbb{P}\left(\sup_{t\geqslant 0}X_t\geqslant\lambda\right)\leqslant\frac{1}{\lambda}\sup_{t\geqslant 0}\mathbb{E}[X_t^+].$$

ii. If X is a càdlàg martingale, then

$$\mathbb{P}\left(\sup_{t\in[0,T]}|X_t|\geqslant\lambda\right)\leqslant\frac{1}{\lambda}\mathbb{E}[|X_T|],\qquad\mathbb{P}\left(\sup_{t\geqslant 0}|X_t|\geqslant\lambda\right)\leqslant\frac{1}{\lambda}\sup_{t\geqslant 0}\mathbb{E}[|X_t|],$$

iii. If X is a càdlàg martingale, then for all $p\in(1,\infty)$

$$\mathbb{E}\left[\sup_{t\in[0,T]}|X_t|^p\right]\leqslant\left(\frac{p}{p-1}\right)^p\mathbb{E}[|X_T|^p],\qquad\mathbb{E}\left[\sup_{t\geqslant 0}|X_t|^p\right]\leqslant\left(\frac{p}{p-1}\right)^p\sup_{t\geqslant 0}\mathbb{E}[|X_t|^p].$$

Proof. Inequalities *i.–ii.* for $T=\infty$ are obtained by sending $T\to\infty$ in the corresponding finite time inequalities, and applying the monotone convergence theorem. There is a small subtlety though, because we may have $\sup_{t\geqslant 0}X_t=\lambda$ and yet $\sup_{t\in[0,T]}X_t<\lambda$ for all $T>0$; we can circumvent this with a small trick. Assume we have proved the finite-time statement in *i.*; then for any $\varepsilon\in(0,1)$, we have:

$$\begin{aligned}\mathbb{P}\left(\sup_{t\geqslant 0}X_t\geqslant\lambda\right)&\leqslant\mathbb{P}\left(\sup_{t\geqslant 0}X_t>\lambda(1-\varepsilon)\right)=\lim_{T\to\infty}\mathbb{P}\left(\sup_{t\in[0,T]}X_t>\lambda(1-\varepsilon)\right)\\&\leqslant\frac{1}{\lambda(1-\varepsilon)}\sup_{t\geqslant 0}\mathbb{E}[X_t^+].\end{aligned}$$

Since the left hand side does not depend on ε, we can then send $\varepsilon\to 0$ to get the claim. The argument works similarly for ii.

To derive the inequality *i.* in finite time, let $(I_n)_{n\in\mathbb{N}}$ be an increasing sequence of finite subsets of $\mathbb{Q}_+\cap[0,T]$ such that $\bigcup_{n\in\mathbb{N}}I_n=\mathbb{Q}_+\cap[0,T]$ and such that $T\in I_n$ for all n. From the σ-continuity of $\mathbb{P}$ and the right-continuity of X, we get

$$\mathbb{P}\left(\sup_{t\in[0,T]}X_t>\lambda\right)=\lim_{n\to\infty}\mathbb{P}\left(\sup_{t\in I_n}X_t>\lambda\right)\leqslant\lim_{n\to\infty}\frac{1}{\lambda}\mathbb{E}[X_T^+]=\frac{1}{\lambda}\mathbb{E}[X_T^+],$$

where we applied Doob's inequality in discrete time, Lemma 4.9. With the same argument as above, we can replace $\mathbb{P}\left(\sup_{t\in[0,T]}X_t>\lambda\right)$ by $\mathbb{P}\left(\sup_{t\in[0,T]}X_t\geqslant\lambda\right)$ on the left hand side.

If X is a càdlàg martingale, then $\tilde{X}_t=|X_t|$ is a càdlàg submartingale, so that *ii.* follows from *i.* applied to $\tilde{X}$.

The inequalities in *iii.* are obtained using similar arguments, relying on Lemma 4.9 and the monotone convergence theorem (instead of σ-continuity). ☐

Exercise. Show that in each of the infinite-time inequalities in Theorem 4.20 we have

$$\sup_{t\geqslant 0}\mathbb{E}[\varphi(X_t)]=\lim_{t\to\infty}\mathbb{E}[\varphi(X_t)].$$

Hint: recall Remark 4.6.

Remark 4.21. The constant $\left(\frac{p}{p-1}\right)^p$ in Doob's maximal L^p-inequality (inequality *iii.* in Theorem 4.20) is optimal. Notice that it diverges for $p\to 1$; indeed, the inequality is false for $p=1$ and we cannot control $\mathbb{E}[\sup_{t\in[0,T]}|X_t|]$ in terms of $\mathbb{E}[|X_T|]$ (the corresponding inequality already fails in finite discrete time).

To control the supremum, one needs X_T to belong to $L\log L$: if X is a càdlàg martingale, then it holds

$$\mathbb{E}\left[\sup_{t\in[0,T]}|X_t|\right]\leqslant\frac{e}{e-1}(1+\mathbb{E}[|X_T|\log|X_T|])$$

see for instance Exercise II.1.16 in [23].

If τ is a stopping time and X_t almost surely converges to X_∞ as $t\to\infty$, then we define

$$X_\tau(\omega) := \mathbb{1}_{\{\tau(\omega)<\infty\}} X_{\tau(\omega)}(\omega) + \mathbb{1}_{\{\tau(\omega)=\infty\}} X_\infty(\omega).$$

Theorem 4.22. (Optional Sampling Theorem) *Let X be a càdlàg martingale and let $\sigma \leqslant \tau$ be stopping times. Assume that either*

i. $\tau \leqslant C < \infty$ almost surely, where $C>0$ is a deterministic constant;

ii. or X is uniformly integrable.

Then X_σ and X_τ are in L^1 and

$$\mathbb{E}[X_\tau|\mathcal{F}_\sigma] = X_\sigma.$$

We momentarily postpone the proof of Theorem 4.22 (similarly for Corollary 4.24 below) in order to present some applications of interest to Brownian motion (cf. Corollary 4.25).

Theorem 4.23. *Let X be a positive càdlàg supermartingale (which almost surely converges to some X_∞ by Proposition 4.14) and let $\sigma \leqslant \tau$ be stopping times. Then X_σ and X_τ are in L^1 and*

$$\mathbb{E}[X_\tau|\mathcal{F}_\sigma] \leqslant X_\sigma.$$

Proof. The proof is similar to the one of Theorem 4.22, but in some places a bit more technical, so we skip it; for a reference, see Theorem 3.25 of [16]. □

Corollary 4.24. (Stopping theorem) *Let X be a càdlàg martingale and let τ be a stopping time. Then the stopped process $X_t^\tau = X_{t\wedge\tau}$, $t\geqslant 0$, is a càdlàg martingale. If X is uniformly integrable, then X^τ is as well and we have*

$$X_t^\tau = \mathbb{E}[X_\tau|\mathcal{F}_t] \quad \forall t\geqslant 0. \tag{4.1}$$

∘

Corollary 4.25. *Let B be a Brownian motion and write $\tau_x = \inf\{t\geqslant 0: B_t = x\}$ for $x\in\mathbb{R}$. Let $a,b>0$. Then*

$$\mathbb{P}(\tau_{-a}<\tau_b) = \frac{b}{a+b}, \qquad \mathbb{P}(\tau_{-a}>\tau_b) = \frac{a}{a+b}.$$

Proof. By the stopping theorem $B^{\tau_{-a}\wedge\tau_b}$ is a martingale, which has uniformly bounded trajectories due to the very definition of $\tau_{-a}\wedge\tau_b$: it holds $\sup_{t\geqslant 0}|B_t^{\tau_{-a}\wedge\tau_b}| \leqslant |a|\vee|b|$; therefore $B^{\tau_{-a}\wedge\tau_b}$ is uniformly integrable. By Corollary 3.23, the stopping time $\tau_{-a}\wedge\tau_b$ is almost surely finite and we get

$$\begin{aligned}0 =& \mathbb{E}[B_0^{\tau_{-a}\wedge\tau_b}] = \mathbb{E}[B_{\tau_{-a}\wedge\tau_b}]\\ =& \mathbb{E}[B_{\tau_{-a}}\mathbb{1}_{\tau_{-a}\leqslant\tau_b} + B_{\tau_b}\mathbb{1}_{\tau_a>\tau_b}]\\ =& -a\,\mathbb{P}(\tau_{-a}\leqslant\tau_b) + b\,\mathbb{P}(\tau_{-a}>\tau_b)\end{aligned}$$

Now observe that $\mathbb{P}(\tau_{-a}\leqslant\tau_b) = 1-\mathbb{P}(\tau_{-a}>\tau_b)$, and therefore

$$0 = -a(1-\mathbb{P}(\tau_{-a}>\tau_b)) + b\,\mathbb{P}(\tau_{-a}>\tau_b) = (a+b)\mathbb{P}(\tau_{-a}>\tau_b) - a,$$

from where the claim follows (notice that by definition $\tau_{-a} \neq \tau_b$). □

——— End of the lecture on November 21 ———

Proof of Theorem 4.22. It suffices to show *ii.*, because then we obtain *i.* by considering the uniformly integrable martingale $(X_{t\wedge C})_{t\geqslant 0}$ (cf. Example 4.19).

To show *ii.*, define for $n\in\mathbb{N}$

$$\sigma_n = \sum_{k=0}^{\infty} (k+1)\, 2^{-n}\, \mathbb{1}_{\{\sigma\in[k2^{-n},(k+1)2^{-n})\}} + \infty\, \mathbb{1}_{\{\sigma=\infty\}}$$

and similarly for τ_n. Then σ_n and τ_n are stopping times and decrease to σ and τ, respectively, as $n\to\infty$. It is not hard to see that σ_n and τ_n are also stopping times with respect to the discrete time filtration $(\mathcal{F}_{k2^{-n}})_{k\geqslant 0}$; moreover, the σ-algebra $\mathcal{F}_{\sigma_n}$ defined w.r.t. $(\mathcal{F}_{k2^{-n}})_{k\geqslant 0}$ coincide with the one defined w.r.t. $(\mathcal{F}_t)_{t\geqslant 0}$, similarly for τ_n. We can therefore apply the discrete time Optional Sampling Theorem for uniformly integrable martingales (Stochastics II) to obtain

$$\mathbb{E}[X_{\tau_n}|\mathcal{F}_{\sigma_n}] = X_{\sigma_n}.$$

Conditioning both sides on $\mathcal{F}_\sigma$ and using that $\sigma\leqslant\sigma_n$, so that $\mathcal{F}_\sigma\subset\mathcal{F}_{\sigma_n}$, we obtain

$$\mathbb{E}[X_{\tau_n}|\mathcal{F}_\sigma] = \mathbb{E}[\mathbb{E}[X_{\tau_n}|\mathcal{F}_{\sigma_n}]|\mathcal{F}_\sigma] = \mathbb{E}[X_{\sigma_n}|\mathcal{F}_\sigma].$$

Since X is right-continuous, $(X_{\tau_n})_n$ converges almost surely to X_τ. By the discrete time stopping theorem we know that $X_{\tau_n} = \mathbb{E}[X_\infty|\mathcal{F}_{\tau_n}]$, and therefore $(X_{\tau_n})_n$ is uniformly integrable and the convergence also holds in L^1. Similarly, $(X_{\sigma_n})_n$ converges in L^1 to X_σ, and therefore

$$\mathbb{E}[X_\tau|\mathcal{F}_\sigma] = \lim_{n\to\infty} \mathbb{E}[X_{\tau_n}|\mathcal{F}_\sigma] = \lim_{n\to\infty} \mathbb{E}[X_{\sigma_n}|\mathcal{F}_\sigma] = \mathbb{E}[X_\sigma|\mathcal{F}_\sigma] = X_\sigma.$$

Note that $X_\tau\in L^1$ as we just showed that it is the limit in L^1 of the sequence X_{τ_n}; similarly for X_σ. □

Exercise. Which parts of the proof of Theorem 4.22 also work for supermartingales, and where did we use the martingale property of X (as opposed to the supermartingale property)? Compare with the statement of Theorem 4.23.

Proof of Corollary 4.22. Since X is càdlag, X^τ is also càdlàg due to its definition.

First assume that X is uniformly integrable; note that, once we prove (4.1), both martingale and uniform integrability properties follow from Remark 4.18 (and the fact that $X_\tau\in L^1$, by Theorem 4.22). We have

$$\begin{aligned}\mathbb{E}[X_\tau|\mathcal{F}_t] &= \mathbb{E}[X_\tau\,\mathbb{1}_{\tau\leqslant t}|\mathcal{F}_t] + \mathbb{E}[X_\tau\,\mathbb{1}_{\tau>t}|\mathcal{F}_t]\\ &= X_\tau\,\mathbb{1}_{\tau\leqslant t} + \mathbb{E}[X_{\tau\vee t}\,\mathbb{1}_{\tau>t}|\mathcal{F}_t]\\ &= X_\tau\,\mathbb{1}_{\tau\leqslant t} + \mathbb{E}[X_{\tau\vee t}\,|\mathcal{F}_t]\,\mathbb{1}_{\tau>t}\end{aligned}$$

where in the second step we used that $X_\tau\,\mathbb{1}_{\tau\leqslant t}$ is $\mathcal{F}_t$-measurable (cf. Exercise Sheet 4) and similarly in the last step that $\mathbb{1}_{\tau>t}$ is $\mathcal{F}_t$-measurable. Since $\tau\vee t$ is a stopping time, $\tau\vee t\geqslant t$, and X is uniformly integrable, by Theorem 4.22 we get

$$\mathbb{E}[X_\tau|\mathcal{F}_t] = X_\tau\,\mathbb{1}_{\tau\leqslant t} + X_t\,\mathbb{1}_{\tau>t} = X_{t\wedge\tau} = X_t^\tau.$$

Now let X be any càdlàg martingale and fix $t \geqslant 0$. Then X^t is a uniformly integrable martingale (cf. Example 4.19), $(X^t)^\tau = X^{t\wedge\tau}$; so for any $s \leqslant t$ we get

$$\mathbb{E}[X_t^\tau | \mathcal{F}_s] = \mathbb{E}[X_t^{t\wedge\tau} | \mathcal{F}_s] = X_s^{t\wedge\tau} = X_s^\tau. \qquad \square$$

Exercise. (A random walk embedded in Brownian motion, this exercise is a bit technical) Let B be a Brownian motion and consider the stopping times $\tau_0 := 0$ and

$$\tau_{n+1} := \inf\{t \geqslant \tau_n : |B_t - B_{\tau_n}| = 1\} = \inf\{t \geqslant 0 : \mathbb{1}_{\{t \geqslant \tau_n\}} |B_t - B_{\tau_n}| = 1\}.$$

Show that $X_n := B_{\tau_n}$, $n \in \mathbb{N}_0$, is a simple symmetric random walk.

We skipped this remark in the lectures: With the help of the *Skorokhod embedding theorem*, one can show that for *any* centered and square-integrable random walk $(Y_n)_{n\in\mathbb{N}_0}$ (i.e. Y_n is the sum of i.i.d. centered and square-integrable random variables) there exist integrable stopping times $(\tau_n)_{n\in\mathbb{N}_0}$ such that $(B_{\tau_n})_{n\in\mathbb{N}_0}$ has the same distribution as $(Y_n)_{n\in\mathbb{N}_0}$. Once this result is shown, it leads to a relatively simple proof of Donsker's invariance principle. See for example Chapters 1.10 and 1.11 of Liggett [17].

5 Continuous semimartingales

We saw that the Brownian motion is nowhere differentiable. But we would like to make sense of stochastic differential equations such as

$$\partial_t Y_t = b(Y_t) + \sigma(Y_t) \partial_t B_t, \qquad Y_0 = y_0.$$

One way of doing so is to integrate both sides from 0 to t, formally obtaining

$$Y_t = y_0 + \int_0^t b(Y_s) \mathrm{d}s + \int_0^t \sigma(Y_s) \mathrm{d}B_s. \tag{5.1}$$

In order to make sense of the equation, we construct the stochastic integral $\int H_s \,\mathrm{d}X_s$ for suitable processes H and X; in particular, we would like not only to be able to make sense of $\int H_s \,\mathrm{d}B_s$, but also of objects of the form $\int H_s \,\mathrm{d}Y_s$ for processes Y which are themselves solutions to (5.1).

A good class of integrators X turn out to be *semimartingales*, i.e. processes X which can be decomposed as $X = M + A$, where M is a *local martingale* and A has *paths of finite variation*. In this lecture we will restrict our attention to continuous semimartingales, and we start by studying some basic properties of semimartingales.

Assumption: from now on until the end of the lecture notes we will assume that our filtration $\mathbb{F}$ satisfies the usual conditions – unless explicitly stated otherwise. It is possible (and sometimes crucial) to develop the theory that follows without this assumption, see e.g. Jacod-Shiryaev [13], Section I.4, where the filtration is only assumed to be right-continuous (which does not change too much), or von Weizsäcker-Winkler [28], Chapters 5 and 6, where not even right-continuity is needed (for this we have to be very careful). As the material is already technical enough as it is, we prefer to simplify our life as much as possible and thus we work under the usual conditions.

Let us also stress that, in the Brownian case, by Blumenthal's 0-1 law (Corollary 3.22), $\mathcal{F}_t^B$ and $\mathcal{F}_{t+}^B$ can only differ by trivial sets, i.e. having either probability 0 or 1. In particular, if we take the completion of $\mathbb{F}^B$ w.r.t. the Wiener measure $\mathbb{P}$, it is automatically right-continuous.

Recall the notation for the increment of X from s to t: $X_{s,t} := X_t - X_s$.

5.1 Processes of finite variation

Definition 5.1. (Finite variation, total variation) *Let $T\in(0,+\infty)$. We say that a continuous function $a\colon[0,T]\to\mathbb{R}$ is of* finite variation *on $[0,T]$, i.e. $a\in\mathrm{TV}([0,T])$, if $a(0)=0$ and its* total variation *on $[0,T]$ is finite:*

$$\|a\|_{\mathrm{TV}([0,T])}:=\sup\left\{\sum_{k=0}^{n-1}|a(t_{k+1})-a(t_k)|: n\in\mathbb{N}, 0=t_0<\ldots<t_n=T\right\}<\infty.$$

We say that a continuous function $a\colon\mathbb{R}_+\to\mathbb{R}$ is of finite variation*, notation $a\in\mathrm{TV}(\mathbb{R}_+)$, if $a|_{[0,T]}$ is of finite variation for all $T\in(0,\infty)$. In that case we write*

$$V(a)(t):=\|a\|_{\mathrm{TV}([0,t])},\quad\forall t\geqslant 0.$$

Example 5.2.

i. Any $a\in C^1(\mathbb{R}_+)$ with $a(0)=0$ is of finite variation: If $0=t_0<\ldots<t_n=t$, then

$$\sum_{k=0}^{n-1}|a(t_{k+1})-a(t_k)|\leqslant\sum_{k=0}^{n-1}\max_{s\in[t_k,t_{k+1}]}|a'(s)|\times|t_{k+1}-t_k|\leqslant\max_{s\in[0,t]}|a'(s)|\cdot t.$$

ii. More generally, any absolutely continuous $a\colon\mathbb{R}_+\to\mathbb{R}$ with $a(0)=0$ is of finite variation, because

$$\sum_{k=0}^{n-1}|a(t_{k+1})-a(t_k)|=\sum_{k=0}^{n-1}\left|\int_{t_k}^{t_{k+1}}a'(s)ds\right|\leqslant\sum_{k=0}^{n-1}\int_{t_k}^{t_{k+1}}|a'(s)|ds\leqslant\int_0^t|a'(s)|ds.$$

iii. Any increasing function $a\colon\mathbb{R}_+\to\mathbb{R}$ with $a(0)=0$ is of finite variation, because

$$\sum_{k=0}^{n-1}|a(t_{k+1})-a(t_k)|=\sum_{k=0}^{n-1}(a(t_{k+1})-a(t_k))=a(t)-a(0)=a(t).$$

iv. By the previous example, there exist continuous functions a of finite variation which are not absolutely continuous: one example is the *devil's staircase*, which is constant outside of the Cantor set (which has Lebesgue measure 0).

v. If a,b are of finite variation, then also $a+b$ is of finite variation (follows from the triangle inequality); if a is of finite variation, so is λa for any $\lambda\in\mathbb{R}$. In particular, $\mathrm{TV}([0,T])$ is a vector space.

The last two examples together show that if a_+,a_- are increasing functions starting from 0 and $a=a_+-a_-$, then a is of finite variation. Next, we will see that the existence of such a decomposition is also necessary for a to be of finite variation.

Proposition 5.3. *Let $a\in C(\mathbb{R}_+)$ be such that $a(0)=0$. The following conditions are equivalent:*

i. $a\in\mathrm{TV}(\mathbb{R}_+)$.

ii. There exist two measures μ_+ and μ_- on $\mathcal{B}(\mathbb{R}_+)$ such that $\mu_\pm([0,T])<\infty$ for all $T>0$, such that μ_+ and μ_- are mutually singular (i.e. there exists $D_+\in\mathcal{B}(\mathbb{R}_+)$ with $\mu_+=\mu_+(\cdot\cap D_+)$ and $\mu_-(D_+)=0$), and such that

$$a(t)=\mu([0,t]):=\mu_+([0,t])-\mu_-([0,t]).$$

iii. a can be written as the difference of two increasing functions a_+ and a_-, $a(t)=a_+(t)-a_-(t)$.

In this case μ_+ and μ_- are unique, we write $|\mu| = \mu_+ + \mu_-$, and we have

$$|\mu|([0,t]) = V(a)(t).$$

We call μ the (signed) measure associated with *a.*

Remark 5.4. Note that, since a is continuous, and μ_+ and μ_- are mutually singular, they must be non-atomic: $\mu_\pm(\{t\}) = 0$ for all $t \geqslant 0$.

Proof. (Sketch of proof):

- *i.* $\Rightarrow$ *ii.*: One can show that $V(a)$ is continuous, see e.g. Friz-Victoir [10], Proposition 1.12. Define

  $$a_+(t) := \frac{1}{2}(V(a)(t) + a(t)), \qquad a_-(t) := \frac{1}{2}(V(a)(t) - a(t)).$$

  Then

  $$a(t) = a_+(t) - a_-(t), \qquad t \geqslant 0,$$

  and $a_\pm$ are positive, increasing continuous functions and therefore they are the "distribution functions" of two measures on $\mathcal{B}(\mathbb{R}_+)$, determined via

  $$\mu_+([0,t]) = a_+(t), \qquad \mu_-([0,t]) = a_-(t).$$

  By construction we have $a(t) = \mu_+([0,t]) - \mu_-([0,t])$ and one can show that μ_+ and μ_- are mutually singular and the unique mutually singular measures with this property.

  (Sketch of the argument, you may skip this): Use the Jordan decomposition $\mu = \nu_+ - \nu_-$ of the signed measure $\mu := \mu_+ - \mu_-$, where $\nu_\pm$ are mutually singular by definition of the Jordan decomposition, and use that

  $$(\mu_+ + \mu_-)((s,t]) = V(a)(t) - V(a)(s) \leqslant (\nu_+ + \nu_-)((s,t]) \leqslant (\mu_+ + \mu_-)((s,t]),$$

  where the second inequality holds by definition of $V(a)$ as a supremum, and the last inequality holds because the Jordan decomposition is minimal. Since also $\mu_+ - \mu_- = \nu_+ - \nu_-$ we get $\mu_\pm = \nu_\pm$, and then that μ_+ and μ_- are mutually singular because ν_+ and ν_- are mutually singular.
  Uniqueness: Also follows from the Jordan decomposition.

- *ii.* $\Rightarrow$ *iii.*: Set $a_\pm(t) := \mu_\pm([0,t])$.

- *iii.* $\Rightarrow$ *i.*: This is exactly the discussion in Example 5.2-iii-iv. □

Definition 5.5. (Lebesgue-Stieltjes integration) *Let $a \in C(\mathbb{R}_+) \cap \mathrm{TV}(\mathbb{R}_+)$ and let $h: \mathbb{R}_+ \to \mathbb{R}$ be measurable with $\int_0^T |h(t)||\mu|(\mathrm{d}t) < \infty$ for all $T \geqslant 0$. Then we define*

$$\int_0^t h(s)\mathrm{d}a(s) := \int_{[0,t]} h(s)\mu(\mathrm{d}s) = \int_{[0,t]} h(s)\mu_+(\mathrm{d}s) - \int_{[0,t]} h(s)\mu_-(\mathrm{d}s), \qquad t \geqslant 0,$$

and

$$\int_0^t h(s)\mathrm{d}V(a)(s) := \int_{[0,t]} h(s)|\mu|(\mathrm{d}s), \qquad t \geqslant 0.$$

Both $\int_0^\cdot h(s)\mathrm{d}a(s)$ and $\int_0^\cdot h(s)\mathrm{d}V(a)(s)$ are continuous functions (by dominated convergence and the atomless property of $|\mu|$), have finite variation and the associated measures are $h\mu$ and $h|\mu|$.

——— End of the lecture on November 27 ———

Example 5.6. If $a\in C^1(\mathbb{R}_+)$, then

$$\int_0^t h(s)\mathrm{d}a(s)=\int_0^t h(s)a'(s)\mathrm{d}s, \qquad \int_0^t h(s)\,\mathrm{d}V(a)(s)=\int_0^t h(s)|a'(s)|\mathrm{d}s.$$

Indeed, for $h=\mathbb{1}_{(u,v]}$ this follows from the fundamental theorem of calculus, and for more general h we apply the usual approximation arguments ("measure-theoretic induction", e.g. the monotone class theorem, cf. Theorem A.11).

Exercise. Let $a\in C(\mathbb{R}_+)\cap \mathrm{TV}(\mathbb{R}_+)$ and let $h\in C(\mathbb{R}_+)$. Show that the integral $\int_0^t h(s)\mathrm{d}a(s)$ can be computed as limit of Riemann sums:

$$\int_0^t h(s)\mathrm{d}a(s)=\lim_{n\to\infty}\sum_{k=0}^{n-1} h\Big(\frac{kt}{n}\Big)\Big(a\Big(\frac{(k+1)t}{n}\Big)-a\Big(\frac{kt}{n}\Big)\Big).$$

Now we introduce randomness:

Definition 5.7. (Process of finite variation) *A stochastic process $A=(A_t)_{t\geqslant 0}$ is a* process of finite variation *if it is adapted, continuous, and $A(\omega)\in \mathrm{TV}(\mathbb{R}_+)$ for all $\omega\in\Omega$. In that case we write $A\in\mathcal{A}$. If furthermore $A(\omega)$ is increasing for all $\omega\in\Omega$, we write $A\in\mathcal{A}^+$.*

If $A\in\mathcal{A}$, then $V(A)$ is a continuous increasing process and also adapted, so $V(A)\in\mathcal{A}^+$: Indeed, it is possible (exercise!) to show that

$$V(A)_t=\lim_{n\to\infty}\sum_{k=0}^{n-1}\Big|A_{\frac{k+1}{n}t}-A_{\frac{k}{n}t}\Big| \qquad \forall t\geqslant 0$$

and for fixed n the sum on the right hand side is obviously $\mathcal{F}_t$–measurable.

Proposition 5.8. *Let $A\in\mathcal{A}$ and let H be a progressively measurable process such that almost surely*

$$\int_0^T |H_s|\mathrm{d}V(A)_s<\infty \qquad \forall T\geqslant 0.$$

Then

$$\Big(\int_0^t H_s\,\mathrm{d}A_s\Big)(\omega):=\begin{cases}\int_0^t H_s(\omega)\mathrm{d}A_s(\omega) & \text{if } \int_0^T |H_s|(\omega)\mathrm{d}V(A)_s(\omega)<\infty \text{ for all } T\geqslant 0,\\ 0 & \text{otherwise,}\end{cases}$$

defines (ω-wise) a (progressively measureable) process of finite variation.

Proof. Note that by Lemma 3.14 the trajectories of H are measurable, so both the condition $\int_0^t |H_s|(\omega)\mathrm{d}V(A)_s(\omega)<\infty$ and the term $\int_0^t H_s(\omega)\mathrm{d}A_s(\omega)$ make sense. By definition, the trajectories

$$t\mapsto\Big(\int_0^t H_s\,\mathrm{d}A_s\Big)(\omega)$$

are continuous finite variation functions for all ω, so we only have to show that $\int_0^\cdot H_s\,\mathrm{d}A_s$ is adapted. Upon modifying H on a null set (recall that our filtration satisfies the usual conditions!) we may assume that $\int_0^t |H_s|(\omega)\mathrm{d}V(A)_s(\omega)<\infty$ and $(H\cdot A)_t(\omega)=\int_0^t H_s(\omega)\mathrm{d}A_s(\omega)$ for all $\omega\in\Omega$ and all $t\geqslant 0$.

Fix $t \geqslant 0$. We first show that $\omega \mapsto \int_0^t \mathbb{1}_B(\omega, s) \mathrm{d}A_s(\omega)$ is $\mathcal{F}_t$-measurable for all $B \in \mathcal{F}_t \otimes \mathcal{B}([0,t])$. For $B = B_1 \times (a,b]$ with $B_1 \in \mathcal{F}_t$ and $0 \leqslant a < b \leqslant t$ or $B = B_1 \times \{0\}$ this is clear. But these sets are stable by intersection and they generate $\mathcal{F}_t \otimes \mathcal{B}([0,t])$. Moreover, the set of all $B \in \mathcal{F}_t \otimes \mathcal{B}([0,t])$ for which $\omega \mapsto \int_0^t \mathbb{1}_B(\omega, s) \mathrm{d}A_s(\omega)$ is $\mathcal{F}_t$-measurable is a λ-system since measurability is preserved under pointwise convergence. Therefore, this holds for all $B \in \mathcal{F}_t \otimes \mathcal{B}([0,t])$. By the usual approximation argument ("measure-theoretic induction", first consider $H = \sum_{k=1}^n x_k \mathbb{1}_{B_k}$, then positive H, then differences of positive functions; equivalently the monotone class theorem) it follows that, for all progressive H such that $\int_0^t |H_s| \mathrm{d}V(A)_s < \infty$, the random variable $\omega \mapsto \int_0^t H_s(\omega) \mathrm{d}A_s(\omega)$ is $\mathcal{F}_t$-measurable. □

Remark 5.9. (Associativity of the Lebesgue-Stieltjes integral) If, for progressive H, G, almost surely $\int_0^t |H_s| \mathrm{d}V(A)_s < \infty$ and $\int_0^t |G_s H_s| \mathrm{d}V(A)_s < \infty$ for all $t \geqslant 0$, then we have

$$\int_0^\cdot G_s \,\mathrm{d}\left(\int_0^s H_r \,\mathrm{d}A_r \right) = \int_0^\cdot G_s H_s \,\mathrm{d}A_s,$$

because $\int_0^\cdot H_s \,\mathrm{d}A_s$ is a finite variation process associated to the measure $H \mathrm{d}\mu$, where μ is the measure associated to A.

Exercise. Let $A(\omega) \in C^1$ for all ω, and let $G(\omega), H(\omega)$ be continuous for all ω. Verify the identity $\int_0^\cdot G_s \mathrm{d}(\int_0^\cdot H_s \mathrm{d}A_s) = \int_0^\cdot G_s H_s \mathrm{d}A_s$ using only elementary (Analysis I) arguments and Example 5.6.

5.2 Brownian motion and prelude to stochastic integration

Disclaimer: from now on we will always (unless specified) consider processes $(X_t)_{t \geqslant 0}$ with continuous time. Whenever talking about continuous martingales, we therefore mean martingales with $\mathbb{P}$-a.s. continuous paths, not just martingales indexed over continuous time $t \in [0, \infty)$.

Let now B be a Brownian motion. If we would have $B \in \mathcal{A}$, then we could use the results from the last section to construct $\int_0^t H_s \,\mathrm{d}B_s$ for suitable integrands H. However, we have the following negative result, informing us that there is no hope to accomplish this strategy for *any* continuous martingale!

Lemma 5.10. (Continuous finite variation martingales are constant) *Let $M \in \mathcal{A}$ be a continuous martingale of finite variation. Then almost surely $M_t = 0$ for all $t \geqslant 0$.*

To prove the lemma, we first need a couple of remarks.

Remark 5.11. Let M be a square integrable martingale, then it enjoys the property of *orthogonality of increments*: for any $t_1 < t_2 < t_3 < t_4$, it holds

$$\begin{aligned}
\mathbb{E}[(M_{t_4} - M_{t_3})(M_{t_2} - M_{t_1})] &= \mathbb{E}[\mathbb{E}[(M_{t_4} - M_{t_3})(M_{t_2} - M_{t_1}) | \mathcal{F}_{t_2}]] \\
&= \mathbb{E}[(M_{t_2} - M_{t_1}) \,\mathbb{E}[(M_{t_4} - M_{t_3}) | \mathcal{F}_{t_2}]] \\
&= 0.
\end{aligned}$$

Exercise. Let A be a process of finite variation, τ be a stopping time, $A_t^\tau = A_{t \wedge \tau}$ be the stopped process. Show that $V(A^\tau)_t = V(A)_{t \wedge \tau} = V(A)_t^\tau$.

Proof of Lemma 5.10. Since $M \in \mathcal{A}$, the process $V(M)$ is continuous and $V(M)_0 = 0$.

Without loss of generality, we may assume M and $V(M)$ to be uniformly bounded. Indeed, if they weren't, define the stopping time $\tau_m = \inf \{t \geqslant 0 : V(M)_t \geqslant m\}$ with $m \in \mathbb{N}$; note that

$$|M_t^{\tau_m}| \leqslant V(M^{\tau_m})_t = V(M)_t^{\tau_m} \leqslant m \quad \forall t \geqslant 0,$$

where we used the above exercise. By the stopping theorem, M^{τ_m} is a continuous, bounded martingale of finite variation; one we show the statement for M^{τ_m}, so that $M^{\tau_m} \equiv 0$, we can send $m \to \infty$ (note that $\mathbb{P}$-a.s. $\tau_m \uparrow +\infty$ by definition, since we assume $M \in \mathcal{A}$) to conclude that $\mathbb{P}$-a.s. $M \equiv 0$ as well.

So now assume $|M|$, $V(M)$ uniformly bounded by m and fix $t > 0$; for $n \in \mathbb{N}$, set $t_k = \frac{kt}{n}$. By Remark 5.11, we have

$$\begin{aligned}
\mathbb{E}[M_t^2] &= \mathbb{E}\left[\left(\sum_{k=0}^{n-1} (M_{t_{k+1}} - M_{t_k})\right)^2\right] \\
&= \sum_{k=0}^{n-1} \mathbb{E}[(M_{t_{k+1}} - M_{t_k})^2] + 2\sum_{k \neq l} \mathbb{E}[(M_{t_{k+1}} - M_{t_k})(M_{t_{\ell+1}} - M_{t_\ell})] \\
&= \sum_{k=0}^{n-1} \mathbb{E}[(M_{t_{k+1}} - M_{t_k})^2].
\end{aligned}$$

By the definition of total variation we find

$$\begin{aligned}
\mathbb{E}[M_t^2] &\leqslant \mathbb{E}\left[V(M)_t \cdot \sup_{k=0,\ldots,n-1} |M_{t_{k+1}} - M_{t_k}|\right] \\
&\leqslant m\,\mathbb{E}\left[\sup_{k=0,\ldots,n-1} |M_{t_{k+1}} - M_{t_k}|\right].
\end{aligned}$$

For $n \to \infty$, the term inside the expectation goes to zero because M is continuous (thus uniformly continuous on $[0,t]$). Since moreover $\sup_{t \geqslant 0} |M_t| \leqslant m$, by dominated convergence we get $\mathbb{E}[M_t^2] = 0$. Since $t > 0$ was arbitrary, the proof is complete. ☐

Remark 5.12. Lemma 5.10 shows that any nontrivial continuous martingale is almost surely of infinite variation. For discontinuous martingales this is not true: recall for instance the compensated Poisson process of Example 4.4.

Exercise. Show that the compensated Poisson process is almost surely of finite variation (with the same definition of "finite variation" that we used for continuous functions).

Since Brownian motion is a continuous martingale and not constant, we deduce that it is not in $\mathcal{A}$. In fact, it follows from the previous proof that for any square-integrable martingale M and any partition $0 = t_0 < t_1 < \ldots < t_n = t$ of $[0,t]$, we have:

$$\mathbb{E}[(M_t - M_0)^2] = \mathbb{E}\left[\sum_{k=0}^{n-1} (M_{t_k, t_{k+1}})^2\right]; \tag{5.2}$$

so it seems more reasonable to expect that $\sum_{k=0}^{n-1} (M_{t_k,t_{k+1}})^2$ converges as $n \to \infty$, rather than hoping for convergence of $\sum_{k=0}^{n-1} |M_{t_k,t_{k+1}}|$. We start by showing this explicitly in the Brownian case.

Lemma 5.13. (Quadratic variation of Brownian motion) *Let $t > 0$ and let $0 = t_0^n < t_1^n < \ldots < t_{k_n}^n = t$ be a sequence of deterministic partitions of $[0,t]$ with $\max_{0 \leqslant i < k_n} |t_{i+1}^n - t_i^n|$ converging to zero as $n \to \infty$. Then*

$$\lim_{n \to \infty} \sum_{i=0}^{k_n - 1} (B_{t_{i+1}^n} - B_{t_i^n})^2 = t$$

where the convergence takes place in $L^2(\Omega, \mathbb{P})$.

We call $\langle B\rangle_t := t$, $t \geqslant 0$, the quadratic variation *of B. It is the unique (up to indistinguishability) process in $\mathcal{A}^+$ such that $B_t^2 - \langle B\rangle_t$, $t \geqslant 0$, is a martingale.*

Proof. The variables $(B_{t_{i+1}^n} - B_{t_i^n})^2_{i=0,\ldots,k_n-1}$ are independent and $\mathbb{E}[(B_{t_{i+1}^n} - B_{t_i^n})^2] = (t_{i+1}^n - t_i^n)$; therefore

$$\begin{aligned}
\left[\left(\sum_{i=0}^{k_n-1} (B_{t_{i+1}^n} - B_{t_i^n})^2 - t\right)^2\right] &= \mathrm{var}\left(\sum_{i=0}^{k_n-1} (B_{t_{i+1}^n} - B_{t_i^n})^2\right)\\
&= \sum_{i=0}^{k_n-1} \mathrm{var}((B_{t_{i+1}^n} - B_{t_i^n})^2)\\
&\leqslant \sum_{i=0}^{k_n-1} \mathbb{E}[(B_{t_{i+1}^n} - B_{t_i^n})^4]\\
\left(B_{t_{i+1}^n - t_i^n} \sim \sqrt{t_{i+1}^n - t_i^n}\, N,\ \mathbb{E}[N^4] = 3 \text{ for } N \sim \mathcal{N}(0,1)\right) &= 3 \sum_{i=0}^{k_n-1} (t_{i+1}^n - t_i^n)^2\\
&\lesssim \max_{0 \leqslant k < k_n} |t_{k+1}^n - t_k^n| \sum_{i=0}^{k_n-1} (t_{i+1}^n - t_i^n)\\
&= t \cdot \max_{0 \leqslant k < k_n} |t_{k+1}^n - t_k^n| \to 0
\end{aligned}$$

as $n \to \infty$ by assumption.

We already saw in Example 4.3 that $B_t^2 - t$ is a martingale. If $A \in \mathcal{A}^+$ is another process for which $B_t^2 - A_t$ is a martingale, then by linearity $t - A_t \in \mathcal{A}$ is a continuous martingale of finite variation; therefore $t - A_t$ is indistinguishable from 0 by Lemma 5.10. □

Exercise. Let $f \in C^\alpha([0,t])$ with $\alpha \in (\frac{1}{2}, 1]$. Show that for any sequence of partitions as in Lemma 5.13, it holds

$$\lim_{n\to\infty} \sum_{i=0}^{k_n-1} (f(t_{i+1}^n) - f(t_i^n))^2 = 0.$$

——— End of the lecture on November 28 ———

Refining the arguments from Lemmas 5.10 and 5.13 (see also Exercise Sheet 6) it's easy to show that

$$\mathbb{P}(\|B\|_{\mathrm{TV}([s,t])} = +\infty \ \text{ for all } 0 \leqslant s < t < \infty) = 1$$

(technically we only defined $\mathrm{TV}([0,t])$, but the definition immediately generalizes to any interval $[s,t]$). As a consequence, we cannot hope to integrate general continuous stochastic processes against B naively. Indeed, for any n and any partition $\pi^n = \{0 = t_0 < t_1 < \ldots < t_{k_n}^n = t\}$, we can find a continuous process $H^{(n)}$ such that $|H_t^{(n)}(\omega)| \leqslant 1$ for all t and ω and such that

$$\sum_{k=0}^{n-1} H_{t_k^n}^{(n)} B_{t_k^n, t_{k+1}^n} = \sum_{k=0}^{n-1} |B_{t_k^n, t_{k+1}^n}|,$$

implying that this quantity can grow without control as $n \to \infty$ and not converge at all. $H^{(n)}$ can be constructed by enforcing $H_{t_k^n}^{(n)} = \mathrm{sgn}(B_{t_k^n, t_{k+1}^n})$ and then extending to all other values of t by piecewise linear interpolation (or splines, Legendre polynomials, or the numerical scheme you prefer).

However, to construct $H^{(n)}$ one already needs to know $\mathrm{sgn}(B_{t_k^n,t_{k+1}^n})$ at time t_k^n, so we have to peak into the future. The brilliant idea of Itô in the 1940s was to restrict the space of integrands to those that are adapted (more precisely, progressively measurable) with respect to the past of B (or more generally with respect to our filtration $\mathbb{F}$). Moreover, unlike for the integral against processes of finite variation A, we will not freeze the realization $B(\omega)$ and construct a pathwise integral using real analysis, but rather we construct the integral as a limit in L^2, so using functional analysis instead. The point is that we want to exploit to the fullest all the stochastic cancellations coming from the special structure of the process B, similarly to what we did in the construction of the Wiener integral. Like therein, we will in fact obtain an exact *isometry* between L^2-spaces.

We will present the construction of stochastic integrals for the much richer class of continuous (local) martingales, which requires to understand the key role played by their *quadratic variation*. In view of this, let us first run some preliminary computations, starting from candidate Stjeltes-type integrals: given a filtration $\pi=\{0=t_0<t_1<\ldots<t_n=t\}$, and a collection of bounded random variables $\{H_{t_k}\}_{k=1}^n$ such that H_{t_k} is $\mathcal{F}_{t_k}$-measurable, consider

$$(H\cdot B)_t=\sum_{k=0}^{n-1} H_{t_k} B_{t_k,t_{k+1}}=\text{``}\int_0^t \sum_{k=0}^{n} H_{t_k}\mathbb{1}_{[t_k,t_{k+1})}(s)\,\mathrm{d}B_s\text{''}$$

and let's try to compute its L^2-norm:

$$\mathbb{E}[(H\cdot B)_t^2]=\sum_{k=0}^{n-1} \mathbb{E}[|H_{t_k}|^2\,|B_{t_k,t_{k+1}}|^2]+2\sum_{k\neq l} \mathbb{E}[H_{t_k} B_{t_k,t_{k+1}} H_{t_l} B_{t_l,t_{l+1}}].$$

By the martingale property, like in Remark 5.11, the terms with $k\neq l$ vanish: suppose wlog $k<l$, then

$$\mathbb{E}[H_{t_k} B_{t_k,t_{k+1}} H_{t_l} B_{t_l,t_{l+1}}]=\mathbb{E}[H_{t_k} B_{t_k,t_{k+1}} H_{t_l}\mathbb{E}[B_{t_l,t_{l+1}}|\mathcal{F}_{t_l}]]=0.$$

Combined with the independence of increments and stationarity properties of B, we get

$$\begin{aligned}\mathbb{E}[(H\cdot B)_t^2]&=\sum_{k=0}^{n-1}\mathbb{E}[|H_{t_k}|^2](t_{k+1}-t_k)\\&=\mathbb{E}\left[\sum_{k=0}^{n-1}|H_{t_k}|^2(t_{k+1}-t_k)\right]\\&=\mathbb{E}\left[\int_0^t\sum_{k=0}^{n}H_{t_k}^2\,\mathbb{1}_{[t_k,t_{k+1})}(s)\mathrm{d}s\right]=\mathbb{E}\left[\left\|\sum_{k=0}^{n}H_{t_k}\,\mathbb{1}_{[t_k,t_{k+1})}\right\|_{L^2([0,t])}^2\right]\end{aligned}$$

which already looks very promising, since all passages were exact equalities!

For general square integrable martingales M, if we similarly define $(H\cdot M)_t$, the same procedure however will not reach a conclusion. We can still push it as far as possible given our current theory:

$$\begin{aligned}\mathbb{E}[(H\cdot M)_t^2]&=\sum_{k=0}^{n-1}\mathbb{E}[|H_{t_k}|^2\,(M_{t_k,t_{k+1}})^2]\\&=\sum_{k=0}^{n-1}\mathbb{E}[|H_{t_k}|^2\,\mathbb{E}[(M_{t_k,t_{k+1}})^2|\mathcal{F}_{t_k}]]\\&=\mathbb{E}\left[\sum_{k=0}^{n-1}|H_{t_k}|^2\,\mathbb{E}[(M_{t_k,t_{k+1}})^2|\mathcal{F}_{t_k}]\right].\end{aligned}$$

Notice again that, by the martingale property, it holds

$$\begin{aligned}\mathbb{E}[(M_{t_k,t_{k+1}})^2|\mathcal{F}_{t_k}] =& \mathbb{E}[M_{t_{k+1}}^2|\mathcal{F}_{t_k}] - 2\mathbb{E}[M_{t_{k+1}}M_{t_k}|\mathcal{F}_{t_k}] + \mathbb{E}[M_{t_k}^2|\mathcal{F}_{t_k}]\\ =& \mathbb{E}[M_{t_{k+1}}^2|\mathcal{F}_{t_k}] - 2\,M_{t_k}\,\mathbb{E}[M_{t_{k+1}}|\mathcal{F}_{t_k}] + M_{t_k}^2\\ =& \mathbb{E}[M_{t_{k+1}}^2|\mathcal{F}_{t_k}] - M_{t_k}^2\\ =& \mathbb{E}[M_{t_{k+1}}^2 - M_{t_k}^2|\mathcal{F}_{t_k}].\end{aligned}$$

Inserting this identity in the above we arrive at

$$\mathbb{E}[(H\cdot M)_t^2] = \mathbb{E}\left[\sum_{k=0}^{n-1} |H_{t_k}|^2\,\mathbb{E}[M_{t_{k+1}}^2 - M_{t_k}^2|\mathcal{F}_{t_k}]\right].$$

Even though M is not of finite variation, the process $t\mapsto M_t^2$ seems to have some monotonicity property, at least w.r.t. conditional expectation: we just saw that

$$\mathbb{E}[M_{t_{k+1}}^2|\mathcal{F}_{t_k}] = M_{t_k}^2 + \mathbb{E}[(M_{t_k,t_{k+1}})^2|\mathcal{F}_{t_k}] \geqslant M_{t_k}^2.$$

Based on this intuition, *if we could* replace $\mathbb{E}[M_{t_{k+1}}^2 - M_{t_k}^2|\mathcal{F}_{t_k}]$ with $\mathbb{E}[A_{t_k,t_{k+1}}|\mathcal{F}_{t_k}]$, where $A_{t_k,t_{k+1}}$ denote the increments of a stochastic process $A\in\mathcal{A}_+$, we would get

$$\mathbb{E}[(H\cdot M)_t^2] = \mathbb{E}\left[\sum_{k=0}^{n-1} |H_{t_k}|^2\,A_{t_k,t_{k+1}}\right] = \mathbb{E}\left[\int_0^t H_{t_k}^2\,\mathbb{1}_{[t_k,t_{k+1})}(s)\,\mathrm{d}A_s\right]$$

which would somewhat restore the desired isometry property, at the price of replacing the standard Lebesgue integration $\mathrm{d}t$ we got in the Brownian case with $\mathrm{d}A_t$. The goal of the next section is to show the existence of A for a large class of continuous martingales.

5.3 Continuous martingales and quadratic variations

It turns out that L^p-bounded, continuous martingales have a nice Banach space structure, at least for $p\in(1,\infty)$. To this end, let us define

$$\mathcal{H}^{p,c} := \left\{(M_t)_{t\geqslant 0} : M \text{ is a continuous martingale, } \|M\|_{\mathcal{H}^{p,c}} := \sup_{t\geqslant 0}\mathbb{E}[|M_t|^p]^{\frac{1}{p}} < \infty\right\};$$

notice that, if $M\in\mathcal{H}^{p,c}$, then it is L^1-bounded, so that by the martingale convergence theorem its $\mathbb{P}$-a.s. limit M_∞ exists almost surely.

Proposition 5.14. *For any $p\in(1,\infty)$, $(\mathcal{H}^{p,c},\|\cdot\|_{\mathcal{H}^{p,c}})$ is a Banach space; moreover it is isometric to a closed linear subspace of $L^p(\Omega)$, with linear isometry J given by $JM := M_\infty$:*

$$\|M\|_{\mathcal{H}^{p,c}}^p := \sup_{t\geqslant 0}\mathbb{E}[|M_t|^p] = \mathbb{E}[|M_\infty|^p] = \|M_\infty\|_{L^p}.$$

Moreover, $M_t\to M_\infty$ in L^p as $t\to\infty$. An equivalent norm $\|\cdot\tilde{\|}_{\mathcal{H}^{p,c}}$ for $\mathcal{H}^{p,c}$ is given by

$$\|M\tilde{\|}_{\mathcal{H}^{p,c}}^p := \mathbb{E}[\|M\|_{C_b(\mathbb{R}_+)}^p] := \mathbb{E}\left[\sup_{t\geqslant 0}|M_t|^p\right].$$

For $p=2$, $\mathcal{H}^{2,c}$ has a Hilbert space structure, with inner product given by

$$(M,N)_{\mathcal{H}^{2,c}} := \mathbb{E}[M_\infty N_\infty]. \tag{5.3}$$

Proof. Exercise Sheet 7. □

We will see later that, due to its Hilbert structure and the nice scalar product given by (5.3), the space $\mathcal{H}^{2,c}$ is a natural one where to develop a stochastic integration theory. Convergence in $\mathcal{H}^{2,c}$ is however very strong, and it implies weaker (but still extremely useful) notions of convergence like the following.

Definition 5.15. (UCP convergence) *Let $\{f^n\}_n$, f be deterministic functions from $\mathbb{R}_+$ to $\mathbb{R}$; we say that f^n converge to f* uniformly on compact sets *if*

$$\lim_{n\to\infty} \sup_{t\in[0,T]} |f_t^n - f_t| = 0 \quad \forall T \in (0, +\infty).$$

Let $\{X^n\}_n$, X be jointly measurable real-valued stochastic processes; we say that X^n convergence to X uniformly on compacts in probability *if*

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{t\in[0,T]} |X_t^n - X_t| > \varepsilon\right) = 0$$

for all $\varepsilon > 0$ and $T \in (0, +\infty)$. In this case, we write $X^n \to X$ in ucp*, and we refere to above as* ucp convergence*.*

The space of continuous stochastic processes (possibly adapted to some reference filtration), endowed to the ucp convergence, has a complete metric structure.

Lemma 5.16. *The space of continuous stochastic processes (possibly adapted to a reference filtration $\mathbb{F}$) is a complete metric space when endowed with the distance*

$$D^{\mathrm{ucp}}(X, Y) := \mathbb{E}[D(X, Y)] := \sum_{k=1}^{\infty} \mathbb{E}\left[2^{-k} \wedge \sup_{t\in[0,k]} |X_t - Y_t|\right]$$

which induces the ucp convergence.

Moreover if $X^n \to X$ in ucp, then there exists a subsequence $\{X^{n_j}\}_j$ such that, for $\mathbb{P}$-a.e. ω, $X^{n_j}(\omega) \to X(\omega)$ uniformly on compact sets.

Proof. We skipped the proof in the lectures, it is included here for completeness.

It's easy to check that by construction D (resp. D^{ucp}) is a metric on the space $C(\mathbb{R}_+)$ (resp. the space of continuous processes); moreover D induces the uniform convergence on compact sets.

Notice that $D(\cdot, \cdot) \leqslant 1$ by construction and that

$$\mathbb{E}\left[1 \wedge \sup_{t\in[0,k]} |X_t - Y_t|\right] \leqslant 2^k D^{\mathrm{ucp}}(X, Y) \quad \forall k \in \mathbb{N}.$$

By Markov's inequality, if $D^{\mathrm{ucp}}(X^n, X) \to 0$, then $\sup_{t\in[0,k]} |X_t^n - X_t| \to 0$ in probability; as the argument holds for any k, it follows that $X^n \to X$ in ucp. Conversely, if $X^n \to X$ in ucp, then each k-term in the series defining $D^{\mathrm{ucp}}(X^n, X)$ converges to 0, and by dominated convergence so does the whole series.

Suppose now that $\{X^n\}_n$ is a Cauchy series w.r.t. D^{ucp}, then we can extract a subsequence such that $D^{\mathrm{ucp}}(X^{n_j}, X^{n_{j+1}}) \leqslant 2^{-j}$, so that

$$\mathbb{E}\left[\sum_{j=1}^{\infty} D(X^{n_j}, X^{n_{j+1}})\right] = \sum_{j=1}^{\infty} D^{\mathrm{ucp}}(X^{n_j}, X^{n_{j+1}}) \leqslant 1.$$

It follows that, for $\mathbb{P}$-a.e. ω, $\{X^{n_j}(\omega)\}_{j\in\mathbb{N}}$ is a Cauchy sequence in $(C(\mathbb{R}_+), D)$ and therefore admits a unique limit in $C(\mathbb{R}_+)$, which we denote by $X(\omega)$:

$$\lim_{j\to\infty} D(X^{n_j}(\omega), X(\omega)) = 0 \quad \text{for } \mathbb{P}\text{-a.e. } \omega.$$

X is a stochastic process since it is the $\mathbb{P}$-a.s. limit of processes, and has continuous trajectories (up to relabelling it on a zero probability set) since $(C(\mathbb{R}_+), D)$ is a complete metric space. Again by the definition of D^{ucp} and dominated convergence it's easy to see that

$$\lim_{j\to\infty} D^{\text{upc}}(X^{n_j}, X) = \lim_{j\to\infty} \mathbb{E}[D(X^{n_j}, X)] = 0$$

and then using the fact that $\{X^n\}_n$ is Cauchy one can deduce by triangular inequality that $\lim_{n\to\infty} D^{\text{upc}}(X^n, X) = 0$. □

Definition 5.17. ($\boldsymbol{\mathcal{M}^{2,c}}$, quadratic variation) *We denote by $\mathcal{M}^{2,c}$ the space of continuous, square integrable martingales $(M_t)_{t\geqslant 0}$. Given $M \in \mathcal{M}^{2,c}$, IF there exists a process $A \in \mathcal{A}_+$ such that*

$$M_t^2 - M_0^2 - A_t$$

is a martingale, then we refer to A as the quadratic variation *of M, and denote it by $\langle M\rangle$.*

Remark 5.18. If $M \in \mathcal{M}^{2,c}$ and $\langle M\rangle$ exists, then it is an integrable process: setting $N_t = M_t^2 - M_0^2 - \langle M\rangle_t$, it holds $\langle M\rangle_t \leqslant M_t^2 + M_0^2 + |N_t|$, where M^2 is integrable since $M \in \mathcal{M}^{2,c}$ and N is integrable by definition of being a martingale. Moreover if $M_0 = 0$, then

$$\mathbb{E}[N_t] = \mathbb{E}[N_0] = 0 \quad\Rightarrow\quad \mathbb{E}[M_t^2] = \mathbb{E}[\langle M\rangle_t] \quad \forall t \geqslant 0.$$

Exercise. Show that, if $\langle M\rangle$ exists, so does $\langle M - M_0\rangle$ and we have $\langle M - M_0\rangle = \langle M\rangle$.

—— End of the lecture on December 4 ——

The next basic result guarantees that Definition 5.17 is meaningful.

Lemma 5.19. *Let $M \in \mathcal{M}^{2,c}$. If $\langle M\rangle$ exists, then it is unique (up to indistinguishability). Moreover, for any stopping time τ it holds that*

$$\langle M\rangle^\tau = \langle M^\tau\rangle.$$

Proof. Like in Lemma 5.13, if A and $\tilde{A}$ are both processes in $\mathcal{A}_+$ satisfying the definition of $\langle M\rangle$, then

$$A - \tilde{A} = (M^2 - M_0^2 - \tilde{A}) - (M^2 - M_0^2 - A)$$

is both a continuous martingale and a process in $\mathcal{A}$, thus it is $\mathbb{P}$-a.s. identically zero by Lemma 5.10; namely $A \equiv \tilde{A}$ $\mathbb{P}$-a.s.

If $M \in \mathcal{M}^{2,c}$ is such that $\langle M\rangle$ exists, then

$$(M^2 - M_0^2 - \langle M\rangle)^\tau = (M^\tau)^2 - M_0^2 - \langle M\rangle^\tau$$

is a martingale by the stopping theorem, which by uniqueness implies $\langle M^\tau\rangle = \langle M\rangle^\tau$. □

The main goal of this section is to ensure the existence of $\langle M\rangle$ and to provide a practical way to construct it. To this end, we need to introduce some terminology.

We can identify a partition π of the real line $\mathbb{R}_+$ with an increasing (possibly unbounded) sequence of ordered points $0 = t_0 < t_1 < \ldots < t_n < \ldots$ and thus we can regard π as a (finite or countable) subset of $\mathbb{R}_+$. We will say that π is *locally finite* if $\pi \cap [0, T]$ is a finite set for every $T \in (0, +\infty)$, equivalently if π has no accumulation points (apart from possibly $+\infty$). The *mesh* of a partition π is defined by $|\pi| = \sup_{k\geqslant 0} |t_{k+1} - t_k|$. Given two partitions π^1 and π^2, we say that π^2 is a *refinement* of π^1 if $\pi^1 \subset \pi^2$.

Given a sequence of partitions $\{\pi^n\}_n$ of $\mathbb{R}_+$, we say that: i) the sequence is *increasing* if $\pi^n \subset \pi^{n+1}$ for all n; ii) the sequence is of *infinitesimal mesh* if $\lim_{n\to\infty} |\pi^n| = 0$.

Theorem 5.20. *Let $M \in \mathcal{M}^{2,c}$, then its quadratic variation $\langle M \rangle$ exists. Moreover, for any deterministic sequence of locally finite partitions $\{\pi_n\}_n$ of $\mathbb{R}_+$ with infinitesimal mesh, upon defining (for $\pi_n = \{t_k^n\}_{k \geqslant 0}$)*

$$A_t^n := \sum_{k \geqslant 0} (M_{t \wedge t_k^n, t \wedge t_{k+1}^n})^2$$

it holds that

$$A^n \to \langle M \rangle \quad \text{in upc as } n \to \infty. \tag{5.4}$$

For any locally finite partition $\pi = \{t_k\}_{k \geqslant 0}$ with $|\pi| < \infty$, setting $A^\pi := \sum_{k \geqslant 0} (M_{t \wedge t_k, t \wedge t_{k+1}})^2$, it holds

$$\mathbb{E}\left[\sup_{t \in [0,T]} |A_t^\pi - \langle M \rangle_t|^2\right] \lesssim \mathbb{E}\left[\left(\sup_{s,t \in [0,T] : |t-s| \leqslant |\pi|} |M_{s,t}|^2\right) \langle M \rangle_T\right] \tag{5.5}$$

for any $T \in (0, +\infty)$ (up to allowing both sides in (5.5) to take value $+\infty$).

In order to give the proof, we need some preliminary lemmas.

Lemma 5.21. *Let $M \in \mathcal{M}^{2,c}$, $\pi = \{t_k\}_{k \in \mathbb{N}}$ deterministic and locally finite. Then*

$$M_t^2 - M_0^2 - A_t^\pi = 2 \sum_{k=0}^{\infty} M_{t \wedge t_k} (M_{t \wedge t_k, t \wedge t_{k+1}}) =: 2 J_t^\pi \tag{5.6}$$

for all $t \geqslant 0$, where the process J^π is a continuous martingale.

Proof. In the case of a finite partition of the interval $[0, T]$, this is the content of Exercise Sheet 8; the proof here proceeds identically. Notice in particular that, under the assumption that π is locally finite, the series appearing in (5.6) is finite for any fixed $t \geqslant 0$. □

Lemma 5.22. *Let $M \in \mathcal{H}^{4,c}$, $0 = t_0 < t_1 < \ldots < t_n = t$. Then*

$$\left\| M_0 \, M_{0,t} - \sum_{i=0}^{n-1} M_{t_i} (M_{t_i, t_{i+1}}) \right\|_{L^2}^2 \lesssim \mathbb{E}\left[\sup_{s \in [0,t]} |M_{0,s}|^2 \sum_{i=0}^{n-1} (M_{t_i, t_{i+1}})^2\right].$$

Proof. Notice tha all terms appearing have above have the right integrability under the assumption $M \in \mathcal{H}^{4,c}$. It holds

$$(*) := M_0 \, M_{0,t} - \sum_{i=0}^{n-1} M_{t_i} (M_{t_i, t_{i+1}}) = -\sum_{i=0}^{n-1} M_{0,t_i} (M_{t_i, t_{i+1}})$$

so that

$$\|(*)\|_{L^2}^2 = \sum_{i=0}^{n-1} \mathbb{E}[(M_{0,t_i})^2 (M_{t_i, t_{i+1}})^2] + 2 \sum_{i<j} \mathbb{E}[M_{0,t_i} (M_{t_i, t_{i+1}}) M_{0,t_j} (M_{t_j, t_{j+1}})]$$

By the martingale property, for any fixed $i < j$, it holds

$$\mathbb{E}[M_{0,t_i} (M_{t_i, t_{i+1}}) M_{0,t_j} (M_{t_j, t_{j+1}})] = \mathbb{E}[M_{0,t_i} (M_{t_i, t_{i+1}}) M_{0,t_j} \, \mathbb{E}[M_{t_j, t_{j+1}} | \mathcal{F}_{t_j}]] = 0. \tag{5.7}$$

Combined with the trivial estimate $|M_{0,t_i}|^2 \leqslant \sup_{s \in [0,t]} |M_{0,s}|^2$, this yields the conclusion. □

With these preparations, we can now present the

Proof of Theorem 5.20. We divide the proof in several steps. To avoid being too repetitive, whenever partitions π appear in the following, they are taken deterministic and locally finite partitions of $\mathbb{R}_+$.

Step 1. We first assume M to be a bounded martingale, namely that there exists a deterministic $C>0$ such that $\mathbb{P}$-a.s. $\sup_{t\geqslant 0}|M_t|\leqslant C$; we will remove this assumption at the very end of the proof. Let π_1, π_2 be partitions with $\pi^1\subset\pi^2$ and consider A^{π_i} as defined above. By Lemma 5.21, taking the difference of the two identities given by (5.6), we have

$$A_t^{\pi_1}-A_t^{\pi_2}=-2(J_t^{\pi_1}-J_t^{\pi_2})$$

where $J^{\pi_1}-J^{\pi_2}$ is a continous martingale. By Doob's inequality, we have

$$\mathbb{E}\left[\sup_{t\in[0,T]}|J_t^{\pi_1}-J_t^{\pi_2}|^2\right]\lesssim\mathbb{E}[|J_T^{\pi_1}-J_T^{\pi_2}|^2]\quad\forall T\in(0,+\infty).$$

Step 2. For fixed T, we now want to estimate the above quantity. Since $\pi^1\subset\pi^2$, without loss of generality we can assume $T\in\pi^1$ (easy to see if you draw a picture). Notice that, since $\pi^1\subset\pi^2$, for any interval $[t_k,t_{k+1}]$ coming from π^1, we find a finite partition by considering

$$[t_k,t_{k+1}]\cap\pi^2=\{t_k=s_0^k<s_1^k<\ldots<s_{n_k}^k=t_{k+1}\}.$$

Therefore

$$J_T^{\pi_1}-J_T^{\pi_2}=\sum_k J_k:=\sum_k\left(M_{t_k^n}M_{t_k^n,t_{k+1}^n}-\sum_{j=0}^{n_k-1}M_{s_j^k}M_{s_j^k,s_{j+1}^k}\right).$$

Arguing as in (5.7), it's easy to check that $\mathbb{E}[J_k J_\ell]=0$ whenever $k\neq\ell$. On the other hand, for fixed k, by Lemma 5.22 (for $\tilde{M}_s=M_{t_k^n+s}$, w.r.t. the shifted filtration $\tilde{\mathcal{F}}_s=\mathcal{F}_{s+t}$) we get

$$\begin{aligned}\|J_k\|_{L^2}^2&\leqslant\mathbb{E}\left[\sup_{s\in[t_k^n,t_{k+1}^n]}|M_{t_k^n,s}|^2\sum_{j=0}^{n_k-1}\left|M_{s_j^k,s_{j+1}^k}\right|^2\right]\\&\leqslant\mathbb{E}\left[\sup_{u,s\in[0,T]:|u-s|\leqslant|\pi^1|}|M_{u,s}|^2\sum_{j=0}^{n_k-1}\left|M_{s_j^k,s_{j+1}^k}\right|^2\right]\end{aligned}$$

so that

$$\begin{aligned}\|J_T^{\pi_1}-J_T^{\pi_2}\|_{L^2}^2&=\sum_k\|J_k\|^2\\&\leqslant\mathbb{E}\left[\sup_{u,s\in[0,T]:|u-s|\leqslant|\pi^1|}|M_{u,s}|^2\sum_k\sum_{j=0}^{n_k-1}\left|M_{s_j^k,s_{j+1}^k}\right|^2\right]\\&=\mathbb{E}\left[\sup_{u,s\in[0,T]:|u-s|\leqslant|\pi^1|}|M_{u,s}|^2A_T^{\pi_2}\right].\end{aligned}\tag{5.8}$$

Step 3. Consider now the increasing sequence of dyadic partitions $\pi^n=\{2^{-n}k\}_{k\in\mathbb{N}}$, which is of infinitesimal mesh. We want to show that $\{A^{\pi_n}\}_n$ is a Cauchy sequence in the ucp topology; in fact we will check something slightly stronger, namely that

$$\lim_{n\to\infty}\sup_{m\geqslant n}\mathbb{E}\left[\sup_{t\in[0,T]}|A_t^{\pi_n}-A_t^{\pi_m}|^2\right]=0\quad\forall T\in(0,+\infty).\tag{5.9}$$

By the previous manipulations, Doob's inequality and (5.8) we arrive at

$$\begin{aligned}\mathbb{E}\left[\sup_{t\in[0,T]}|A_t^{\pi_n}-A_t^{\pi_m}|^2\right]&\lesssim\mathbb{E}\left[\sup_{u,s\in[0,T]:|u-s|\leqslant 2^{-n}}|M_{u,s}|^2A_T^{\pi_{n+m}}\right]\\&\leqslant\mathbb{E}\left[\sup_{u,s\in[0,T]:|u-s|\leqslant 2^{-n}}|M_{u,s}|^4\right]^{1/2}\mathbb{E}[(A_T^{\pi_{n+m}})^2]^{1/2}.\end{aligned}$$

The first term is bounded and thus infinitesimal by dominated convergence, since M is bounded by assumption and uniformly continuous on compact sets; so we only need to provide a uniform-in-m estimate for the second term.

Since M is bounded and (again by Lemma 5.21) $A_T^{\pi_{n+m}} = M_T^2 - M_0^2 - 2J_T^{\pi_{n+m}}$, by triangular inequality this is the same as bounding J_T^{n+m} in L^2. By the "orthogonality of increments" property (5.7), we have

$$\begin{aligned}\mathbb{E}[|J_T^{\pi_{n+m}}|^2] =& \mathbb{E}\left[\sum_{k\geqslant 0} \left|M_{t_k^{n+m}\wedge T}\right|^2 \left|M_{t_k^{n+m}\wedge T, t_{k+1}^{n+m}\wedge T}\right|^2\right]\\ \leqslant& C^2\,\mathbb{E}\left[\sum_{k\geqslant 0} \left|M_{t_k^{n+m}\wedge T, t_{k+1}^{n+m}\wedge T}\right|^2\right]\\ =& C^2\,\mathbb{E}[|M_T - M_0|^2] \lesssim C^4\end{aligned}$$

where in the intermediate passage we used (5.2). Overall we conclude that

$$\sup_{m\geqslant n} \mathbb{E}\left[\sup_{t\in[0,T]} |A_t^{\pi_n} - A_t^{\pi_m}|^2\right] \lesssim C^4\,\mathbb{E}\left[\sup_{u,s\in[0,T]:|u-s|\leqslant 2^{-n}} |M_{u,s}|^4\right]^{1/2}$$

which yields (5.9). Therefore $(A^{\pi_n})_n$ is Cauchy and admits a unique limit in ucp, as well as in $L^2(\Omega; C([0,T]))$ for any fixed $T\in(0,+\infty)$; denote it by A. For fixed n, by construction

$$A_0^{\pi_m} = 0, \quad A_{(k+1)2^{-n}}^{\pi_{n+m}} \geqslant A_{k2^{-n}}^{\pi_{n+m}} \quad \forall m\geqslant 0$$

therefore passing to the limit the same properties hold true for A; since dyadic points densify on $\mathbb{R}_+$ and A is continuous, we deduce that A is increasing and so $A\in\mathcal{A}_+$.

Step 4. Since $A_t^{\pi_n}\to A_t$ in L^2 and $M^2 - M_0^2 - A^{\pi_n}$ is a martingale for each n, by passing to the limit we deduce that $M^2 - M_0^2 - A$ is a martingale as well. Since $A\in\mathcal{A}_+$, this proves that A is the (unique) quadratic variation of M, $A = \langle M\rangle$.

Step 5. Proof of (5.5): given a partition π, we can always find an increasing sequence $\{\pi^n\}_n$ with infinitesimal mesh with $\pi^1 = \pi$; the previous estimates (cf. (5.8)) yield

$$\mathbb{E}\left[\sup_{t\in[0,T]} |A_t^{\pi} - A_t^{\pi_n}|^2\right] \lesssim \mathbb{E}\left[\left(\sup_{s,t\in[0,T]:|t-s|\leqslant|\pi|} |M_{s,t}|^2\right) |A_T^{\pi_n}|\right]$$

and we can now pass to the limit as $n\to\infty$, since $A^{\pi_n}\to\langle M\rangle$, to deduce that (5.5) holds. As a consequence of (5.5), given any sequence of partitions π^n of infinitesimal mesh, not necessarily increasing, we have

$$\limsup_{n\to\infty} \mathbb{E}\left[\sup_{t\in[0,T]} |A_t^{\pi_n} - \langle M\rangle_t|^2\right] \lesssim \limsup_{n\to\infty} \mathbb{E}\left[\left(\sup_{s,t\in[0,T]:|t-s|\leqslant|\pi^n|} |M_{s,t}|^2\right) \langle M\rangle_T\right] = 0.$$

Step 6. Finally, we remove the assumption that M is bounded by a *localization argument*. Let $M\in\mathcal{M}^{2,c}$ and define the stopping times $\tau_n = \inf\{t\geqslant 0: |M_t|\geqslant n\}$. Clearly τ_n is an increasing sequence, and moreover for $\mathbb{P}$-a.e. ω we have $\tau_n(\omega)\uparrow+\infty$ since $M(\omega)$ is continuous (thus bounded on compact sets). By the stopping theorem, M^{τ_n} is a bounded continuous martingale, so the previous steps apply and $\langle M^{\tau_n}\rangle\in\mathcal{A}_+$ is well-defined. Moreover, for $m\geqslant n$, by Remark 5.18 it holds $\langle M^{\tau_m}\rangle^{\tau_n} = \langle M^{\tau_n}\rangle$ and so we can consistently define

$$\langle M\rangle_t(\omega) = \begin{cases} \langle M^{\tau_n}\rangle_t(\omega) & \text{if there exists } n \text{ such that } t\leqslant\tau_n(\omega)\\ 0 & \text{otherwise.}\end{cases}.$$

It's easy to check that $\langle M^{\tau_n}\rangle \to \langle M\rangle$ in ucp since

$$\mathbb{P}\left(\sup_{t\in[0,T]} |\langle M\rangle_t - \langle M^{\tau_n}\rangle_t| > \varepsilon\right) \leqslant \mathbb{P}(\tau_n < T) \to 0 \text{ as } n\to\infty.$$

By properties of ucp convergence, $\langle M\rangle \in \mathcal{A}_+$ since $\langle M\rangle^{\tau_n}$ do so. Taking $m\to\infty$ in the above relation, we also get $\langle M\rangle^{\tau_n} = \langle M^{\tau_n}\rangle$; since $\langle M\rangle$ is increasing, by monotone convergence and Remark 5.18 we get

$$\mathbb{E}[\langle M\rangle_t] = \lim_{n\to\infty} \mathbb{E}[\langle M\rangle_{t\wedge\tau_n}] = \lim_{n\to\infty} \mathbb{E}[\langle M\rangle_t^{\tau_n}] = \lim_{n\to\infty} \mathbb{E}[M^2_{\tau_n\wedge t}] = \mathbb{E}[M_t^2]$$

where in the last step we used the fact that $(M^2_{\tau_n\wedge t})_n$ is uniformly integrable since by Doob's inequality

$$M^2_{\tau_n\wedge t} \leqslant \sup_{s\in[0,t]} M_s^2 \quad \forall n, \quad \mathbb{E}\left[\sup_{s\in[0,t]} M_s^2\right] \lesssim \mathbb{E}[M_t^2] < \infty.$$

The same arguments show that, for any fixed t, $\langle M\rangle_t^{\tau_n} \to \langle M\rangle_t$ and $(M_t^{\tau_n})^2 \to M_t^2$ in L^1, so that from

$$\mathbb{E}[(M_t^{\tau_n})^2 - M_0^2 - \langle M\rangle_t^{\tau_n} | \mathcal{F}_s] = (M_s^{\tau_n})^2 - M_0^2 - \langle M\rangle_s^{\tau_n}$$

we can pass to the limit to conclude that $M^2 - M_0^2 - \langle M\rangle$ is a martingale. □

Extra comment: The proof actually yields the relation

$$M_t^2 - M_0^2 - \langle M\rangle_t = \lim_{n\to\infty} 2\sum_k M_{t_k^n}(M_{t_{k+1}^n\wedge t} - M_{t_k^n}) = 2\int_0^t M_s\,\mathrm{d}M_s \tag{5.10}$$

where the last term is a *stochastic integral*, as we will see later; equation (5.10) will be a prototypical example of the *Itô formula*.

Equation (5.10) can also be interpreted as a failure of the *classical chain rule*: if M were classically differentiable, by the fundamental theorem of calculus we should have found

$$M_t^2 - M_0^2 = 2\int_0^t M_s\, M_s'\,\mathrm{d}s = 2\int_0^t M_s\,\mathrm{d}M_s$$

without any term $\langle M\rangle$ appearing. The fact that M is not C^1 is not surprising (if it were, it would be of finite variation, thus $M\equiv 0$ by Lemma 5.10), but (5.10) actually tells us that *standard rules of calculus* do not apply for martingales. This is the reason why we will need to develop a theory of *stochastic calculus* instead, which is nontrivial and truly requires to exploit the cancellations coming from a probabilistic framework (i.e. the martingale property).

In the proof, in order to show that $(A^n)_n$ is Cauchy, it was actually easier to work with $(J^n)_n$, the corresponding approximations of the stochastic integral. With the combined information coming from $(M, \langle M\rangle)$ at hand, we will be able to set up a far reaching theory of stochastic integration.

This fact bears a strong similarity with much more recent and pathwise theories of integration coming from *rough path theory*, see the monographs [10, 9]: loosely speaking, given some signal X, in order to rigorously define objects of the form $\int_0^\cdot F(X_s)\,\mathrm{d}X_s$, one first needs to *enhance* the signal by *postulating* (or, in our case, proving) the existence of the iterated integral $\mathbb{X}_t := \int_0^\cdot X_s\,\mathrm{d}X_s$; once the pair $(X, \mathbb{X})$ is fixed, then the integrals $\int_0^\cdot F(X_s)\,\mathrm{d}X_s$ are also uniquely determined for any smooth bounded function F.

—— End of the lecture on December 5 ——

Definition 5.23. **(Quadratic covariation)** *Let* $M,N\in\mathcal{M}^{2,c}$*. We define the* quadratic covariation *of* M *and* N *as the process*

$$\langle M,N\rangle:=\frac{1}{4}\langle M+N\rangle-\frac{1}{4}\langle M-N\rangle.$$

Notice the analogy of (5.23) with a polarization formula (recall the discussion in the proof of Lemma 1.12).

Proposition 5.24. *Let* M*,* $N\in\mathcal{M}^{2,c}$*. Then:*

a) $\langle M,N\rangle$ *is the unique (up to indistinguishability) process in* $\mathcal{A}$ *such that*

$$MN-M_0N_0-\langle M,N\rangle$$

is a martingale.

b) For any sequence of deterministic locally finite partitions $\pi^n=\{t_k^n\}_{k\in\mathbb{N}}$ *of* $\mathbb{R}_+$ *with infinitesimal mesh* $\lim_{n\to\infty}|\pi^n|=0$*, it holds that*

$$\sum_{k=0}^{\infty}M_{t_k^n\wedge t,t_{k+1}^n\wedge t}N_{t_k^n\wedge t,t_{k+1}^n\wedge t}\to\langle M,N\rangle_t\qquad\text{in ucp.}$$

c) The map $(M,N)\mapsto\langle M,N\rangle$ *is bilinear and symmetric, and for any bounded stopping time* τ *it holds that*

$$\langle M,N\rangle^\tau=\langle M^\tau,N^\tau\rangle=\langle M^\tau,N\rangle.$$

Proof. Exercise Sheet 8. □

Part a) of the above Proposition (and possibly a passage to the limit procedure $t\to\infty$) immediately imply the following.

Corollary 5.25. *If* $M,N\in\mathcal{M}^{2,c}$ *and either* $M_0=0$ *or* $N_0=0$*, then for any* $t\in[0,\infty)$ *it holds* $\mathbb{E}[\langle M,N\rangle_t]=\mathbb{E}[M_tN_t]$*. If additionally* $M,N\in\mathcal{H}^{2,c}$*, then*

$$\mathbb{E}[\langle M,N\rangle_\infty]=\mathbb{E}[M_\infty N_\infty]=(M,N)_{\mathcal{H}^{2,c}}.$$

Remark 5.26. If B^1 and B^2 are independent Brownian motions, we have seen in Example 4.3 that B^1B^2 is a martingale, so that

$$\langle B^1,B^2\rangle\equiv0.$$

In fact more generally, if $M,N\in\mathcal{M}^{2,c}$ are independent, then necessarily $\langle M,N\rangle\equiv0$.

Exercise. Let $f:\mathbb{R}_+\to\mathbb{R}_+$ be a deterministic continuous increasing function and let $M_t=B_{f(t)}$ for a Brownian motion B. Show that M is a continuous martingale in its own filtration and $\langle M\rangle_t=f(t)$.

5.4 Continuous local martingales

We have already seen in the proofs of Lemma 5.10 and Theorem 5.20 that whenever dealing with martingales, it is very convenient to employ some *localization procedures* by introducing convenient families of stopping times. It might therefore not be surprising that several of the results presented so far can be extended to a larger class of objects than just continuous, square integrable martingales.

Definition 5.27. ($\mathcal{M}^c$, localizing sequence, local martingale)

i. *We write $\mathcal{M}^c$ for the set of all continuous martingales.*

ii. *A* localizing sequence *is an increasing sequence of stopping times $(\tau_n)_n$ with $\lim_{n\to\infty}\tau_n=\infty$ almost surely.*

iii. *An adapted process M is called a* local martingale *if there exists a localizing sequence $(\tau_n)_n$ such that, for each $n\in\mathbb{N}$, the stopped process*

$$(M-M_0)^{\tau_n}=M^{\tau_n}-M_0=(M^{\tau_n}-M_0)\mathbb{1}_{\tau_n>0}$$

is a martingale. If in addition M is continuous, we say that M is a continuous local martingale*, we write $M\in\mathcal{M}^c_{\mathrm{loc}}$ and we call $(\tau_n)_n$ a* localizing sequence for M*.*

Exercise. Show that every martingale is a local martingale.

Remark 5.28.

i. We do not require local martingales to be integrable. Indeed, not even M_0 needs to be integrable, and even in the case $M_0=0$, it might happen that M_t is not integrable.

ii. Every martingale is a local martingale. (Exercise above)

iii. *Stability under stopping*: if $M\in\mathcal{M}^c_{\mathrm{loc}}$ (respectively $M\in\mathcal{M}^c$) and τ is a stopping time, then $M^\tau\in\mathcal{M}^c_{\mathrm{loc}}$ (respectively $M^\tau\in\mathcal{M}^c$). Indeed note that $(M^\tau)^{\tau_n}=(M^{\tau_n})^\tau=M^{\tau_n\wedge\tau}$ and then apply the stopping theorem.

iv. Linear combinations of local martingales are martingales: if $M,N\in\mathcal{M}^c_{\mathrm{loc}}$, then $\lambda M+N\in\mathcal{M}^c_{\mathrm{loc}}$ for all $\lambda\in\mathbb{R}$. (Exercise!)

v. For every $M\in\mathcal{M}^c_{\mathrm{loc}}$ there exists a localizing sequence of stopping times $(\tau_n)_n$ such that $(M-M_0)^{\tau_n}$ is a uniformly integrable martingale for all $n\in\mathbb{N}$: just replace τ_n by $\tau_n\wedge n$. The same argument shows that we can assume τ_n to be a bounded stopping time for each fixed n.

Exercise. Show that M is a local martingale if and only if there exists (another) localizing sequence $(\tilde{\tau}^n)_n$ such that $M^{\tilde{\tau}_n}\mathbb{1}_{\tilde{\tau}_n>0}$ is a martingale.

Lemma 5.29. *Let $M\in\mathcal{M}^c_{\mathrm{loc}}$.*

i. *If M is non-negative (i.e. $\mathbb{P}$-a.s. $M_t\geqslant 0$ for every $t\geqslant 0$) and $M_0\in L^1$, then M is a supermartingale.*

ii. *If $\sup_{t\geqslant 0}|M_t|\in L^1$, then M is a uniformly integrable martingale.*

iii. *If $M\in\mathcal{M}^c_{\mathrm{loc}}$, then $\tilde{\tau}_n:=\tau_n\wedge n$, where $\tau_n:=\inf\{t\geqslant 0: |M_t-M_0|\geqslant n\}$, is a localizing sequence for M.*

iv. *If X_0 is $\mathcal{F}_0$-measurable, then $\tilde{M}_t=M_tX_0$ satisfies $\tilde{M}\in\mathcal{M}^c_{\mathrm{loc}}$.*

v. *If $M\in\mathcal{M}^c_{\mathrm{loc}}$ and X_0 is $\mathcal{F}_0$-measurable, then $M-X_0\in\mathcal{M}^c_{\mathrm{loc}}$.*

Proof. Exercise Sheet 9. □

Remark 5.30. Given point *ii.* of Lemma 5.29, we might feel tempted to guess that every uniformly integrable local martingale is a martingale. But this is not true, and we will see a counterexample later. In fact, one can even construct example of $M\in\mathcal{M}^c_{\mathrm{loc}}$ such that $\sup_{t\geqslant 0}\mathbb{E}[|M_t|^2]<\infty$ but $M\notin\mathcal{M}^c$.

Exercise. (hard) Show that, if $\{M^n\}_n$ is a sequence in $\mathcal{M}_{\text{loc}}^c$ such that $M^n \to M$ in ucp, then $M \in \mathcal{M}_{\text{loc}}^c$. In other words, $\mathcal{M}_{\text{loc}}^c$ is closed under convergence in ucp. (*Hint: Lemma 5.16 might come in handy*)

Several of the results we have presented so far for martingales readily extend to local martingales by *localization arguments*.

Lemma 5.31. *Let $M \in \mathcal{A} \cap \mathcal{M}_{\text{loc}}^c$, then $\mathbb{P}$-almost surely $M_t = 0$ for all $t \geqslant 0$ (which we write compactly as $\mathbb{P}$-a.s. $M \equiv 0$).*

Proof. It follows from the corresponding result for martingales (Lemma 5.10) by localization. Indeed, let $(\tau_n)_n$ be a localizing sequence for M, then $M^{\tau_n} \in \mathcal{A} \cap \mathcal{M}^c$ (recall that $V(M^{\tau_n}) = V(M)^{\tau_n}$) and thus $M^{\tau_n} \equiv 0$ for all $n \in \mathbb{N}$. Letting $\tau_n \to \infty$ we get $M \equiv 0$. □

Proposition 5.32. *Let $M \in \mathcal{M}_{\text{loc}}^c$. Then there exists an increasing process $\langle M \rangle \in \mathcal{A}^+$ with $\langle M \rangle_0 = 0$, unique up to indistinguishability, such that*

$$M^2 - M_0^2 - \langle M \rangle \in \mathcal{M}_{\text{loc}}^c \tag{5.11}$$

which we call the quadratic variation *of M. Moreover for any sequence of deterministic locally finite partitions $\pi^m = \{t_k^m\}_{k \in \mathbb{N}}$ of $\mathbb{R}_+$ with infinitesimal mesh, it holds that*

$$\sum_{k=0}^{\infty} (M_{t_k^m \wedge t, t_{k+1}^m \wedge t})^2 \to \langle M \rangle_t \qquad \text{in ucp.} \tag{5.12}$$

Proof. The proof is mostly a variation on the one of Theorem 5.20, up to localization arguments (already contained therein).

It suffices to consider the case $M_0 = 0$, as the general one follows from the relation $\langle M - M_0 \rangle = \langle M \rangle$, cf. Exercise Sheet 9. Let $(\tau_n)_n$ be a localizing sequence for M, which we may assume wlog to be such that $M^{\tau_n} \in \mathcal{M}^{2,c}$ by Lemma 5.29-iii. Then for each τ_n, $\langle M^{\tau_n} \rangle$ exists and by monotonicity of τ_n it provides an increasing-in-n family of processes, so that

$$\langle M \rangle_t = \lim_{n \to \infty} \langle M^{\tau_n} \rangle_t$$

exists and satisfies $\langle M \rangle^{\tau_n} = \langle M^{\tau_n} \rangle$; from this property and the fact that

$$(M^{\tau_n})^2 - \langle M^{\tau_n} \rangle = (M^2 - \langle M \rangle)^{\tau_n}$$

is a martingale, we conclude that $\langle M \rangle$ is the desired quadratic variation. By Theorem 5.20 we also get

$$\sum_{k=0}^{\infty} \left(M^{\tau_n}_{t_k^m \wedge t, t_{k+1}^m \wedge t}\right)^2 = \sum_{k=0}^{\infty} (M_{t_k^m \wedge t \wedge \tau_n, t_{k+1}^m \wedge t \wedge \tau_n})^2 \to \langle M^{\tau_n} \rangle_t = \langle M \rangle_{t \wedge \tau_n} \qquad \text{in ucp}$$

from which we can deduce (5.12) thanks to the property that $\mathbb{P}$-a.s. $\tau_n \uparrow +\infty$ (for any fixed interval $[0, T]$, we can find n large enough such that $\mathbb{P}(\tau_n < T) < \varepsilon$, and apply the ucp convergence for $\langle M^{\tau_n} \rangle$ on the event $\{\tau_n \geqslant T\}$). □

As in the case of square integrable $M, N \in \mathcal{M}^c$, we define the quadratic covariation of $M, N \in \mathcal{M}_{\text{loc}}^c$ by formula

$$\langle M, N \rangle := \frac{1}{4} (\langle M + N \rangle - \langle M - N \rangle) \in \mathcal{A}.$$

In analogy with Proposition 5.24, we have the following result.

Proposition 5.33. *Let* $M, N \in \mathcal{M}_{\mathrm{loc}}^c$.

i. $\langle M, N\rangle$ *is the unique (up to indistinguishability) process in* $\mathcal{A}$ *such that* $MN - M_0 N_0 - \langle M, N\rangle$ *is a local martingale.*

ii. The map $(M, N) \mapsto \langle M, N\rangle$ *is bilinear and symmetric.*

iii. For any sequence of deterministic locally finite partitions $\pi^m = \{t_k^m\}_{k\in\mathbb{N}}$ *of* $\mathbb{R}_+$ *with infinitesimal mesh,* $\sum_{k=0}^{\infty} (M_{t_k^m\wedge t, t_{k+1}^m\wedge t})(N_{t_k^m\wedge t, t_{k+1}^m\wedge t}) \to \langle M, N\rangle_t$ *in ucp.*

iv. If τ *is a stopping time, we have* $\langle M, N\rangle^\tau = \langle M^\tau, N^\tau\rangle = \langle M^\tau, N\rangle$.

v. $\langle M, N\rangle = \langle M - M_0, N - N_0\rangle$.

Proof. The proof is almost identical to that of Proposition 5.24, up to localization arguments. Only notice the key differences: in *i.* we can only deduce that we still have a local martingale, while in *iv.* we can allow any stopping time τ, not necessarily bounded (because we don't have the problem of possible failure of integrability, coming from the requirement of martingality). Part *v.* is new and based on the basic property that $\langle M - M_0\rangle = \langle M\rangle$ (which holds true for martingales, cf. the exercise after Remark 5.18, thus also true for local martingales by localization); cf. also Exercise Sheet 9. □

Lemma 5.34. (Kunita-Watanabe inequality, first version) *Let* $M, N \in \mathcal{M}_{\mathrm{loc}}^c$. *Then* $\mathbb{P}$*-almost surely*

$$|\langle M, N\rangle_{s,t}| \leqslant V(\langle M, N\rangle)_{s,t} \leqslant \langle M\rangle_{s,t}^{1/2} \langle N\rangle_{s,t}^{1/2} \quad \forall s \leqslant t < \infty.$$

Similarly, $\mathbb{P}$*-a.s.*

$$\limsup_{t\to\infty} |\langle M, N\rangle_t| \leqslant V(\langle M, N\rangle)_\infty \leqslant \langle M\rangle_\infty^{1/2} \langle N\rangle_\infty^{1/2}$$

where the second and third terms are always defined by monotonicity (possibly as $+\infty$*).*

——— End of the lecture on December 11 ———

Proof. Fix any $s < t$, then we can construct a series of increasing partitions π^n with infinitesimal mesh such that $\pi^1 = \{s, t\}$. In this case it follows that, for each fixed n, $\pi^n \cap [s, t]$ is a partition of $[s, t]$ and so by Proposition (5.32) $\mathbb{P}$-a.s. it holds

$$\begin{aligned}
|\langle M, N\rangle_{s,t}| &= \lim_{n\to\infty} \left| \sum_k M_{t_k^n, t_k^{n+1}} N_{t_k^n, t_k^{n+1}} \right| \\
&\leqslant \lim_{n\to\infty} \left(\sum_k (M_{t_k^n, t_k^{n+1}})^2 \right)^{1/2} \left(\sum_k (N_{t_k^n, t_k^{n+1}})^2 \right)^{1/2} \\
&= \langle M\rangle_{s,t}^{1/2} \langle N\rangle_{s,t}^{1/2}.
\end{aligned}$$

Having shown the result for fixed $s < t$, we can find a set of full probability where the inequality holds for all rational $s < t$ and finally extend to all $s < t$ by continuity of the paths of $\langle M, N\rangle$, $\langle M\rangle$ and $\langle N\rangle$.

Having established the first inequality, again by Cauchy–Schwarz, for $s = t_0 < \ldots < t_n = t$ it holds:

$$\sum_{k=0}^{n-1} |\langle M, N\rangle_{t_k, t_{k+1}}| \leqslant \sum_{k=0}^{n-1} \langle M\rangle_{t_k, t_{k+1}}^{1/2} \langle N\rangle_{t_k, t_{k+1}}^{1/2} \leqslant \langle M\rangle_{s,t}^{1/2} \langle N\rangle_{s,t}^{1/2}$$

which by definition of total variation yields

$$V(\langle M, N\rangle)_{s,t} \leqslant \langle M\rangle_{s,t}^{1/2} \langle N\rangle_{s,t}^{1/2}.$$

The final inequality follows by taking $s = 0$ and sending $t \to \infty$. □

We will shortly see a very powerful result relating moment bounds for $M \in \mathcal{M}_{\mathrm{loc}}^{\mathrm{c}}$ to moment bounds for $\langle M\rangle$, see Theorem 5.38 below. To this end, we first need some preparations.

Definition 5.35. (Lenglart's domination relation) *Let $(X_t)_{t\geqslant 0}$, $(G_t)_{t\geqslant 0}$ be progressive, non-negative processes. We say that X is* dominated *by G if*

$$\mathbb{E}[X_\tau] \leqslant \mathbb{E}[G_\tau] \quad \text{for all bounded stopping times } \tau \tag{5.13}$$

with the convention that $+\infty \leqslant +\infty$.

In the following, given a continuous stochastic process X, it will be convenient to denote by X^* its running supremum (in modulus), namely $X_t^* = \sup_{s\in[0,t]} |X_s|$.

Lemma 5.36. (Lenglart's inequalities) *Let X be a non-negative, continuous adapted process, and let G be a non-negative, increasing, continuous adapted process; assume that X is dominated by G, in the sense of Definition 5.35. Then:*

i. For any $a, b > 0$, it holds

$$\mathbb{P}\left(\sup_{t\geqslant 0} X_t \geqslant a\right) \leqslant \frac{1}{a}\,\mathbb{E}\left[\sup_{t\geqslant 0} G_t \wedge b\right] + \mathbb{P}\left(\sup_{t\geqslant 0} G_t > b\right). \tag{5.14}$$

ii. For any $\theta \in (0,1)$, it holds

$$\mathbb{E}[(X_\infty^*)^\theta] \leqslant \frac{\theta^{-\theta}}{1-\theta}\,\mathbb{E}[G_\infty^\theta]. \tag{5.15}$$

Proof. Exercise Sheet 8. □

Remark 5.37. Often inequality (5.14) is stated in its weaker but more practical version

$$\mathbb{P}\left(\sup_{t\geqslant 0} X_t \geqslant a\right) \leqslant \frac{b}{a} + \mathbb{P}\left(\sup_{t\geqslant 0} G_t > b\right). \tag{5.16}$$

We are now ready to state and partially prove the following result. It is one of the most useful martingale inequalities, with countless applications.

Theorem 5.38. (Burkholder–Davis–Gundy inequality, BDG for short)
For any $p \in (0,\infty)$, there exist universal constants $c_p, C_p > 0$ such that for any $M \in \mathcal{M}_{\mathrm{loc}}^{\mathrm{c}}$ with $M_0 = 0$, setting $M_t^* = \sup_{s\leqslant t} |M_s|$, it holds

$$c_p\,\mathbb{E}\big[\langle M\rangle_\infty^{p/2}\big] \leqslant \mathbb{E}[(M_\infty^*)^p] \leqslant C_p\,\mathbb{E}\big[\langle M\rangle_\infty^{p/2}\big] \tag{5.17}$$

with the convention that $+\infty \leqslant +\infty$.

Comment: It is customary to say that the constants c_p and C_p are "universal" *because they can be taken the same for all local martingales on any probability space whatsoever.*

Proof. We will only prove the inequality for $p\in(0,4]$, see the comments after the proof for the case $p>4$.

We claim that, once we show the estimate for $p=4$, all the other cases will follow. Indeed, by applying the inequality for $p=4$ for M replaced by M^τ, with τ a finite stopping time, we obtain

$$\mathbb{E}[c_4\langle M\rangle_\tau^2]\leqslant\mathbb{E}[(M_\tau^*)^4]\leqslant\mathbb{E}[C_4\langle M\rangle_\tau^2];$$

in other words, $c_4\langle M\rangle_\tau^2$ is dominated (in the sense of Definition 5.35) by $(M_\tau^*)^4$, which in turn is dominated by $C_4\langle M\rangle_\tau^2$. Therefore by Lenglart's inequality (5.15) we find

$$c_4^\theta\left(\frac{\theta^{-\theta}}{1-\theta}\right)^{-1}\mathbb{E}[\langle M\rangle_\infty^{2\theta}]\leqslant\mathbb{E}[(M_\infty^*)^{4\theta}]\leqslant\frac{\theta^{-\theta}}{1-\theta}C_4^\theta\,\mathbb{E}[\langle M\rangle_\infty^{2\theta}]\quad\forall\theta\in(0,1)$$

which yields the conclusion upon taking $p=4\theta$ for $p\in(0,4)$.

It remains to consider the case $p=4$. Up to localization, we may assume M and $\langle M\rangle$ to be bounded processes; arguing as in Lemmas 5.21 and 5.22, for any fixed partition π of $\mathbb{R}_+$ and any $t\geqslant 0$, we have the relation

$$M_t^2-A_t^\pi=2\sum_{k=0}^\infty M_{t\wedge t_k}(M_{t\wedge t_k,t\wedge t_{k+1}})=:2J_t^\pi$$

with J^π being a martingale, and by Doob's inequality and usual computations based on martingale increments we have

$$\begin{aligned}\mathbb{E}\left[\sup_{s\in[0,t]}|M_s^2-A_s^\pi|\right]&\lesssim\mathbb{E}[|J_t^\pi|^2]=\mathbb{E}\left[\sum_{k=0}^\infty|M_{t\wedge t_k}|^2(M_{t\wedge t_k,t\wedge t_{k+1}})^2\right]\\&\leqslant\mathbb{E}\left[|M_t^*|^2\sum_{k=0}^\infty(M_{t\wedge t_k,t\wedge t_{k+1}})^2\right]\end{aligned}$$

so that after passing to the limit as we take a sequence of partitions we find

$$\mathbb{E}\left[\sup_{s\in[0,t]}|M_s^2-\langle M\rangle_s|^2\right]\lesssim\mathbb{E}[|M_t^*|^2\,\langle M\rangle_t]$$

and now sending $t\to\infty$ we get

$$\mathbb{E}\left[\sup_{t\geqslant 0}|M_t^2-\langle M\rangle_t|^2\right]\leqslant C\mathbb{E}[|M_\infty^*|^2\,\langle M\rangle_\infty]$$

for some $C>0$. We show how to get one of the estimates, the other case being identical upon inverting the roles of $|M^*|^2$ and $\langle M\rangle$. By the basic inequality $(a+b)^2\leqslant 2(a^2+b^2)$

$$\begin{aligned}\mathbb{E}[|M_\infty^*|^4]&=\mathbb{E}\left[\sup_{t\geqslant 0}|M_t^2\pm\langle M\rangle_t|^2\right]\\&\leqslant 2\,\mathbb{E}\left[\sup_{t\geqslant 0}|M_t^2-\langle M\rangle_t|^2\right]+2\mathbb{E}[\langle M\rangle_\infty^2]\\&\leqslant 2C\mathbb{E}[|M_\infty^*|^2\,\langle M\rangle_\infty]+2\mathbb{E}[\langle M\rangle_\infty^2]\\&\leqslant\mathbb{E}\left[\frac{1}{2}|M_\infty^*|^4+2C^2\langle M\rangle_\infty^2\right]+2\mathbb{E}[\langle M\rangle_\infty^2]\\&=\frac{1}{2}\mathbb{E}[|M_\infty^*|^4]+(2C^2+2)\mathbb{E}[\langle M\rangle_\infty^2]\end{aligned}$$

where in the intermediate passage we used the basic inequality $ab \leqslant \frac{a^2}{2} + \frac{b^2}{2}$, for $a = |M^*_\infty|^2$ and $b = 2C\langle M\rangle_\infty$. After rearraging the last inequality so that $\mathbb{E}[|M^*_\infty|^4]$ only appears on the l.h.s., we get the conclusion. □

Comments and bibliography:

- In terms of applications, the second inequality in (5.17) is the truly relevant one. The first inequality in (5.17) informs us that by estimating $\mathbb{E}[(M^*_\infty)^p]$ by $\mathbb{E}\big[\langle M\rangle_\infty^{p/2}\big]$ we are not "losing too much information" as the two quantities are comparable.
- For discontinuous martingales, the BDG inequalities are still true, but only for the range of exponents $p \in [1, \infty)$; note however that for discontinuous processes we cannot apply Lenglart's inequalities, so the above proof does not work.
- We might see later (possibly in Exercise sheets), once we have access to stochastic calculus tools, how to prove BDG inequality for $p > 4$. For proofs not relying on stochastic calculus, see Chapter IV.42 from [24]; the proof therein, solely requires the existence of $\langle M\rangle$ and is based on so called "*good λ inequalities*". It is fairly robust as it not only holds for $x \mapsto |x|^p$, but more generally for *moderate* functions $F(x)$.
- There exist more recent and stronger *pathwise versions* of BDG inequalities, valid for both discrete and càdlàg martingales; see the works [1, 25].

From Theorem 5.38 and localization arguments, we immediately deduce the following.

Corollary 5.39. (Local BDG inequality) *For any $p \in (0, \infty)$ there exist universal constants $c_p, C_p > 0$ such that for any $M \in \mathcal{M}^c_{\mathrm{loc}}$ with $M_0 = 0$ and any stopping time τ, setting $M^*_t = \sup_{s \leqslant t} |M_s|$, it holds*

$$c_p\, \mathbb{E}\big[\langle M\rangle_\tau^{p/2}\big] \leqslant \mathbb{E}[(M^*_\tau)^p] \leqslant C_p\, \mathbb{E}\big[\langle M\rangle_\tau^{p/2}\big]$$

with the convention that $+\infty \leqslant +\infty$.

Corollary 5.40. *Let $M \in \mathcal{M}^c_{\mathrm{loc}}$ with $M_0 = 0$ and let $p \in (1, \infty)$.*

i. The following are equivalent:

a) $M \in \mathcal{H}^{p,c}$;

b) $\mathbb{E}[\langle M\rangle_\infty^{p/2}] < \infty$;

c) $\mathbb{E}[\sup_{t \geqslant 0} |M_t|^p] < \infty$.

If additionally $p \geqslant 2$, then under either of the above conditions, $M^2 - \langle M\rangle$ is a uniformly integrable martingale and in particular $\mathbb{E}[M^2_\infty] = \mathbb{E}[\langle M\rangle_\infty]$.

ii. The following are equivalent:

a) M is a p-integrable martingale;

b) $\mathbb{E}[\langle M\rangle_t^{p/2}] < \infty$ for all $t \geqslant 0$;

c) $\mathbb{E}[\sup_{s \in [0,t]} |M_s|^p] < \infty$ for all $t \geqslant 0$.

If additionally $p \geqslant 2$, then under either of the above conditions, $M^2 - \langle M\rangle$ is a martingale and in particular $\mathbb{E}[M^2_t] = \mathbb{E}[\langle M\rangle_t]$ for all $t \geqslant 0$.

N.b.: compare the above result with Remark 5.30; note that the statement is not true anymore if we try to replace point c) by $\sup_{t \geqslant 0} \mathbb{E}[|M_t|^p] < \infty$ (resp $\sup_{s \in [0,t]} \mathbb{E}[|M_s|^p] < \infty$).

Proof. Equivalence between a) and b) in Part *i.* comes from Exercise Sheet 9; equivalence between a) and c) was already stated in Proposition 5.14. Uniform integrability is again part of Exercise Sheet 9, while $\mathbb{E}[M_\infty^2] = \mathbb{E}[\langle M\rangle_\infty]$ comes from Corollary 5.25. Concerning part *ii.*, it suffices to apply i. to the uniformly integrable martingales $(M_{t\wedge n})_{t\geqslant 0}$, for $n \in \mathbb{N}$. □

Lenglart's inequality allows to derive useful criteria for convergence in the ucp topology.

Corollary 5.41. *Let $\{M^n\}_n$, $M \in \mathcal{M}_{\text{loc}}^c$ with $M_0^n = M_0$ for all n. Then $\langle M^n - M\rangle \to 0$ in ucp if and only if $M^n \to M$ in ucp.*

Proof. Exercise Sheet 9. □

——— **End of the lecture on December 12** ———

5.5 Continuous semimartingales

Definition 5.42. *An adapted process $X = (X_t)_{t\geq 0}$ is a* continuous semimartingale *if it has a decomposition*

$$X = X_0 + M + A \tag{5.18}$$

into a continuous local martingale $M \in \mathcal{M}_{\text{loc}}^c$ and a continuous process $A \in \mathcal{A}$ of finite variation, both with $M_0 = A_0 = 0$.

Exercise. Show that continuous semimartingales form a vector space, i.e. if X and Y are continuous semimartingales then so is $X + \lambda Y$ for every $\lambda \in \mathbb{R}$.

Lemma 5.43. *The decomposition (5.18) of a continuous semimartingale X into M and A is unique (up to indistinguishability).*

Proof. If $X - X_0 = M + A = N + B$ are two decompositions, then $M - N = B - A$ is in $\mathcal{M}_{\text{loc}}^c \cap \mathcal{A}$ and hence indistinguishable from the 0 process by Lemma 5.31. □

Definition 5.44. *If $X = X_0 + M + A$ and $Y = Y_0 + N + B$ are continuous semimartingales, we define the quadratic covariation of X and Y as $\langle X, Y\rangle := \langle M, N\rangle$. In particular, the quadratic variation of X is $\langle X\rangle := \langle M\rangle$.*

Note that $\langle X\rangle$ is uniquely defined thanks to Lemma 5.43, it is consistent with our definition for the quadratic covariation for local martingales, and

$$\langle X, Y\rangle = \langle X - X_0, Y - Y_0\rangle, \qquad \langle X, Y\rangle^\tau = \langle X^\tau, Y^\tau\rangle = \langle X^\tau, Y\rangle$$

in agreement with Proposition 5.33.

Exercise. According to the above definition, deduce from the Kunita–Watanabe inequality (Lemma 5.34) that

$$|\langle X, Y\rangle_{s,t}| \leqslant V(\langle X, Y\rangle)_{s,t} \leqslant \langle X\rangle_{s,t}^{1/2} \langle Y\rangle_{s,t}^{1/2} \quad \forall s \leqslant t < \infty$$

holds in this case we well.

Lemma 5.45. *For any continuous semimartingales X,Y and any sequence of deterministic locally finite partitions $\pi^m=\{t_k^m\}_{k\in\mathbb{N}}$ of $\mathbb{R}_+$ with infinitesimal mesh, it holds that*

$$\sum_{k=0}^{\infty} X_{t_k^m\wedge t,t_{k+1}^m\wedge t}Y_{t_k^m\wedge t,t_{k+1}^m\wedge t}\to\langle X,Y\rangle_t \qquad \textit{in ucp}. \tag{5.19}$$

Proof. By polarization we can reduce to $X=Y$; we can also assume $X_0=0$, so that $X=M+A$. We already know that

$$\sum_{k=0}^{\infty}(M_{t_k^m\wedge t,t_{k+1}^m\wedge t})^2\to\langle M\rangle_t \qquad \text{in ucp}$$

so by developing the square it only remaines to show that

$$\sum_{k=0}^{\infty} M_{t_k^m\wedge t,t_{k+1}^m\wedge t}A_{t_k^m\wedge t,t_{k+1}^m\wedge t}+\sum_{k=0}^{\infty}(A_{t_k^m\wedge t,t_{k+1}^m\wedge t})^2\to 0 \qquad \text{in ucp}. \tag{5.20}$$

Notice that by Cauchy

$$\sum_{k=0}^{\infty} M_{t_k^m\wedge t,t_{k+1}^m\wedge t}A_{t_k^m\wedge t,t_{k+1}^m\wedge t}\leqslant\left(\sum_{k=0}^{\infty}(M_{t_k^m\wedge t,t_{k+1}^m\wedge t})^2\right)^{1/2}\left(\sum_{k=0}^{\infty}(A_{t_k^m\wedge t,t_{k+1}^m\wedge t})^2\right)^{1/2}$$

where we already have convergence in upc to $\langle M\rangle$ for the first term; so (5.20) follows one we show that the second term is infinitesimal. This is actually a result which you already solved in Exercise Sheet 7, but let us give the proof for completeness.

Fix $T\in(0,+\infty)$. Since A is continuous, it is uniformly continuous on $[0,T]$; on the other hand it is of finite variation, therefore for any $t\in[0,T]$ we have

$$\begin{aligned}\sum_{k=0}^{\infty}(A_{t_k^m\wedge t,t_{k+1}^m\wedge t})^2 &\leqslant \sup_k|A_{t_k^m\wedge t,t_{k+1}^m\wedge t}|\sum_{k=0}^{\infty}|A_{t_k^m\wedge t,t_{k+1}^m\wedge t}|\\ &\leqslant\left(\sup_{\substack{0\leqslant u\leqslant s\leqslant T\\ |u-s|\leqslant|\pi^m|}}|A_{u,s}|\right)\|A\|_{\mathrm{TV}([0,T])}\end{aligned}$$

where the estimate holds ω-wise, is uniform in $t\in[0,T]$, and the first term vanishes as $m\to\infty$ by the assumption and uniform continuity. □

Exercise. Let $M\in\mathcal{M}_{\mathrm{loc}}^c$. Show that M^2 is a continuous semimartingale.

Exercise. (very hard given our current tools at disposal) Let X be a continuous semimartingale. Show that X^2 is a semimartingale.

6 Stochastic integration

We now have in place all the ingredients to finally construct the stochastic Itô integral $\int_0^\cdot H_s\mathrm{d}X_s$, for suitable stochastic integrands H and with respect to (henceforth abbreviated *w.r.t.*) continuous semimartingales X as integrators. Although the resulting theory is already quite rich under these assumptions, requiring continuity of X is mostly for technical convenience; see for example Protter [22] for a theory of stochastic integration w.r.t. càdlàg semimartingales.

To develop the theory, we will proceed by steps: we first define $\int_0^\cdot H_s \,\mathrm{d}M_s$ for *elementary processes* H and martingales $M \in \mathcal{M}^{2,c}$, then extend the definition to general H by *Itô isometry*, and then further extend it to $M \in \mathcal{M}^c_{\mathrm{loc}}$ by *localization*. We finally extend the definition to continuous semimartingales $X = M + A$ by linearity. Along the way, special attention will be given to the case where $M = B$ Brownian motion.

Recall that in the following we are always on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ satisfying the usual conditions.

6.1 Stochastic integrals of simple processes

We start by defining stochastic integrals in the simplest possible case, where we can "guess" a meaningful definition just by enforcing linearity of the integration and the property that $\int_s^t \mathrm{d}M_u = M_{s,t}$ whenever $s < t$.

Definition 6.1. (Bounded elementary processes) *We denote by $b\mathcal{E}$ the set of* bounded elementary processes*, namely processes H of the form*

$$H_t(\omega) = \sum_{k=0}^{n-1} h_k(\omega) \mathbb{1}_{(t_k, t_{k+1}]}(t) \tag{6.1}$$

for some given $n \in \mathbb{N}$, $0 \leqslant t_0 < t_1 < \ldots < t_n$, and random variables $\{h_k\}_{k=0}^{n-1}$ such that $h_k \in L^\infty(\mathcal{F}_{t_k})$ for all $k \leq n$.

Remark 6.2. Note that H is left-continuous and adapted, and therefore progressively measurable. Moreover, it is easy to check that $b\mathcal{E}$ is a linear vector space.

Recall previously introduced notations: $\mathcal{M}^{2,c}$ denote square integrable continuous martingales, while $\mathcal{H}^{2,c}$ denote L^2-bounded continuous martingales (the latter is a Hilbert space by Proposition 5.14.

Proposition 6.3. (Itô isometry for bounded elementary processes) *Let $M \in \mathcal{M}^{2,c}$ and $H \in b\mathcal{E}$; we define the* stochastic integral *of H with respect to M as the process*

$$\int_0^t H_s \,\mathrm{d}M_s := \sum_{k=0}^{n-1} h_k \, M_{t_k \wedge t, t_{k+1} \wedge t} \qquad \forall t \geqslant 0.$$

The process $\int_0^\cdot H_s \,\mathrm{d}M_s \in \mathcal{H}^{2,c}$ and has quadratic variation given by

$$\left\langle \int_0^\cdot H_s \,\mathrm{d}M_s \right\rangle_t = \sum_{k=0}^{n-1} h_k^2 \langle M \rangle_{t_k \wedge t, t_{k+1} \wedge t} = \int_0^t H_s^2 \,\mathrm{d}\langle M \rangle_s. \tag{6.2}$$

In particular, we have the Itô isometry

$$\left\| \int_0^\cdot H_s \mathrm{d}M_s \right\|_{\mathcal{H}^{2,c}}^2 = \mathbb{E}\left[\left(\int_0^\infty H_s \mathrm{d}M_s\right)^2\right] = \mathbb{E}\left[\int_0^\infty H_s^2 \mathrm{d}\langle M \rangle_s\right] = \mathbb{E}\left[\sum_{k=0}^{n-1} h_k^2 \langle M \rangle_{t_k, t_{k+1}}\right]. \tag{6.3}$$

Remark 6.4. Note that the definition of $\int_0^\cdot H_s \,\mathrm{d}M_s$ in (6.3) does not depend on the specific choice of the representation (6.1) (namely if we change the choice of $\{t_k\}_{k=0}^n$ by further refining the partition, we get the same process (6.3)). Moreover, it is easy to check that the stochastic integral is *linear in H*, in the sense that (as stochastic processes)

$$\int_0^\cdot (\lambda H_s + K_s) \,\mathrm{d}M_s = \lambda \int_0^\cdot H_s \,\mathrm{d}M_s + \int_0^\cdot K_s \,\mathrm{d}M_s$$

for all $\lambda \in \mathbb{R}$ and $H, K \in b\mathcal{E}$.

Proof. It is clear from the definition that $N_t := \int_0^t H_s \, \mathrm{d}M_s$ is continuous. Noticing that $h_k \in L^\infty$, the sum in (6.3) is finite and by construction $N_t = N_{t_k}$ for $t \geqslant t_n$, it follows that N is integrable and L^2-bounded (since $M \in \mathcal{M}^{2,c}$). Adaptedness is also immediate.

Once we show that N is a martingale and (6.2) holds, the isometry (6.3) follows by taking expectation since $N_0 = 0$ (cf. Corollary 5.25).

To verify the martingale property, by linearity it suffices to show that the process $N_t^k := h_k \, M_{t_k \wedge t, t_{k+1} \wedge t}$ is a martingale, for each fixed $k \in \{0, \ldots, n-1\}$. To this end, fix $s \leqslant t$. If $t \leqslant t_k$, then $N_t^k = N_s^k = 0$ and there is nothing to prove. If $s \geqslant t_k$, then

$$\begin{aligned}\mathbb{E}[N_t^k | \mathcal{F}_s] = \mathbb{E}[h_k \, M_{t_k \wedge t, t_{k+1} \wedge t} | \mathcal{F}_s] &= h_k \, \mathbb{E}[M_{t_{k+1} \wedge t} - M_{t_k \wedge t} | \mathcal{F}_s] \\ &= h_k (M_{t_{k+1} \wedge t \wedge s} - M_{t_k \wedge t \wedge s}) = N_s^k\end{aligned}$$

where in the intermediate passage we used the stopping theorem. If $s \leqslant t_k < t$, then by the tower property and the previous step

$$\mathbb{E}[N_t^k | \mathcal{F}_s] = \mathbb{E}[\mathbb{E}[N_t^k | \mathcal{F}_{t_k}] | \mathcal{F}_s] = \mathbb{E}[N_{t_k}^k | \mathcal{F}_s] = 0 = N_s^k.$$

Overall this shows that N^k is a martingale and by linearity so is N.

It remains to show (6.2), namely that

$$\begin{aligned}\tilde{N}_t &:= \left(\sum_{k=0}^{n-1} h_k \, M_{t_k \wedge t, t_{k+1} \wedge t} \right)^2 - \sum_{k=0}^{n-1} h_k^2 \langle M \rangle_{t_k \wedge t, t_{k+1} \wedge t} \\ &= \sum_{k=0}^{n-1} h_k^2 [(M_{t_k \wedge t, t_{k+1} \wedge t})^2 - \langle M \rangle_{t_k \wedge t, t_{k+1} \wedge t}] + 2 \sum_{j<k} h_k h_j \, M_{t_k \wedge t, t_{k+1} \wedge t} \, M_{t_j \wedge t, t_{j+1} \wedge t}\end{aligned}$$

is again a martingale. We will show that, for each k, the process

$$\tilde{N}_t^k := h_k^2 [(M_{t_k \wedge t, t_{k+1} \wedge t})^2 - \langle M \rangle_{t_k \wedge t, t_{k+1} \wedge t}]$$

is a martingale; a similar argument works for the cross-terms related to $j < k$, which by linearity implies that $\tilde{N}$ is a martingale.

As before, if $s \leqslant t \leqslant t_k$, $\tilde{N}_t^k = \tilde{N}_s^k = 0$; once we have shown the martingale property for $t_k \leqslant s \leqslant t$, the intermediate case $s \leqslant t_k \leqslant t$ follows by conditioning w.r.t. $\mathcal{F}_{t_k}$ first. So we may assume $t_k \leqslant s \leqslant t$. Notice that in this case

$$\begin{aligned}(M_{t_k \wedge t, t_{k+1} \wedge t})^2 - \langle M \rangle_{t_k \wedge t, t_{k+1} \wedge t} &= (M_{t_{k+1} \wedge t} - M_{t_{k+1} \wedge t_k})^2 - \langle M \rangle_{t_{k+1} \wedge t} + \langle M \rangle_{t_{k+1} \wedge t_k} \\ &= (M_{t_k, t}^{t_{k+1}})^2 - \langle M^{t_{k+1}} \rangle_{t_k, t}\end{aligned}$$

where we used the fact that $\langle M \rangle_{t_{k+1} \wedge u} = \langle M^{t_{k+1}} \rangle_u$ by Lemma 5.19. $\tilde{M}_t := M_t^{t_{k+1}}$ is again a continuous martingale by the stopping theorem, therefore for $t_k \leqslant s \leqslant t$ we have

$$\begin{aligned}\mathbb{E}[\tilde{N}_t^k | \mathcal{F}_s] &= h_k^2 \, \mathbb{E}[(\tilde{M}_{t_k, t})^2 - \langle \tilde{M} \rangle_{t_k, t} | \mathcal{F}_s] \\ &= h_k^2 \, \mathbb{E}[\tilde{M}_t^2 - \langle \tilde{M} \rangle_t - 2 \tilde{M}_t \, \tilde{M}_{t_k} + \tilde{M}_{t_k}^2 + \langle \tilde{M} \rangle_{t_k} | \mathcal{F}_s] \\ &= h_k^2 \, [\tilde{M}_s^2 - \langle \tilde{M} \rangle_s - 2 \tilde{M}_{t_k} \, \mathbb{E}[\tilde{M}_t | \mathcal{F}_s] + \tilde{M}_{t_k}^2 + \langle \tilde{M} \rangle_{t_k}] \\ &= h_k^2 \, [\tilde{M}_s^2 - 2 \tilde{M}_{t_k} \tilde{M}_s + \tilde{M}_{t_k}^2 - \langle \tilde{M} \rangle_s + \langle \tilde{M} \rangle_{t_k}] = \tilde{N}_s^k\end{aligned}$$

which concludes the proof. □

Exercise. Complete the proof of Proposition 6.3, i.e. show similarly that the processes

$$t \mapsto h_k h_j \, M_{t_k \wedge t, t_{k+1} \wedge t} \, M_{t_j \wedge t, t_{j+1} \wedge t}$$

with $j<k$ are martingales.

—— End of the lecture on December 18 ——

6.2 Stochastic integration w.r.t. Brownian motion

We now specialize the previous result to the case where $M=B$ Brownian motion. Recall the concept of progressively measurable processes coming from Definition 3.11, and its link to measurability w.r.t. to the σ-algebra Prog of progressive events coming from Remark 3.12.

Definition 6.5. *Let* $\overline{\Omega}:=\Omega\times\mathbb{R}_+$ *and set*

$$\mathbb{P}_B(\mathrm{d}\omega,\mathrm{d}t):=\mathrm{d}t\otimes\mathbb{P}(\mathrm{d}\omega),\qquad L^2(B):=L^2(\overline{\Omega},\mathrm{Prog},\mathbb{P}_B);$$

namely

$$L^2(B)=\left\{H:\Omega\times\mathbb{R}_+\to\mathbb{R}\,\middle|\,H \text{ is progressive, } \|H\|_{L^2(B)}^2:=\mathbb{E}\left[\int_0^\infty H_t^2\,\mathrm{d}t\right]<\infty\right\}.$$

As usual with L^p spaces, we identify two processes H, $\tilde{H}$ if $\|H-\tilde{H}\|_{L^2(B)}=0$.

Note that the definition of $L^2(B)$ is meaningful, since trajectories of H are measurable by Lemma 3.14 and so $\int_0^\infty H_t^2\,\mathrm{d}t$ is a well-defined random variable (with values in $[0,+\infty]$). Since $L^2(B)=L^2(\overline{\Omega},\mathrm{Prog},\mathbb{P}_B)$ is an L^2-space, it is complete and Hilbert; it is immediate to check that, if $H\in b\mathcal{E}$, then $H\in L^2(B)$ as well, namely

$$b\mathcal{E}\subset L^2(B).$$

Since $\langle B\rangle_t=t$, Proposition 6.3 can be restated as follows: the stochastic integral

$$I_B: b\mathcal{E}\ni H\mapsto I_B H:=\int_0^\cdot H_s\,\mathrm{d}B_s\in\mathcal{H}^{2,c}$$

is a linear isometry between $(b\mathcal{E},\|\cdot\|_{L^2(B)})$ and $\mathcal{H}^{2,c}$, since by (6.3) we have

$$\|I_B H\|_{\mathcal{H}^{2,c}}^2=\mathbb{E}\left[\int_0^\infty H_s^2\,\mathrm{d}\langle B\rangle_s\right]=\mathbb{E}\left[\int_0^\infty H_s^2\,\mathrm{d}s\right]=\|H\|_{L^2(B)}^2.$$

Since $\mathcal{H}^{2,c}$ is a Hilbert space (in particular it is complete, cf. Proposition 5.14), it follows that I_B extends uniquely to a isometry defined on the closure of $b\mathcal{E}$ in $L^2(B)$; compare to the construction of Wiener integral from Lemma 1.12. It remains to identify such closure.

Lemma 6.6. *The space* $b\mathcal{E}$ *is dense in* $L^2(B)$.

Proof. To show it, we invoke the following result from functional analysis:

Criterion for density in Hilbert spaces: Let E be a Hilbert space, V be a linear subspace of E. Then V is dense in E if and only if, for any element $w\in E$ *orthogonal* to V, namely such that $(w,v)_E=0$ for all $v\in V$, it must hold $w=0$. In other words, V is dense if and only if $V^\perp=\{0\}$.

Since $L^2(B)$ is Hilbert and $b\mathcal{E}\subset L^2(B)$ is a linear subspace, it suffices to show that $b\mathcal{E}^\perp=\{0\}$. So let $K\in b\mathcal{E}^\perp$ and consider

$$X:=\int_0^\cdot K_r\,\mathrm{d}r;$$

noting that $K\in L^2(B)$, so that $\mathbb{P}$-a.s. $\int_0^{+\infty}|K_t|^2\,\mathrm{d}t<\infty$, Cauchy's inequality implies that $\mathbb{P}$-a.s. $\int_0^T|K_t|\,\mathrm{d}t<\infty$ for all $T<\infty$. In particular, $X\in\mathcal{A}$.

We claim that X is a (continuous) martingale. If that is the case, then by Lemma 5.31, $X\equiv 0$ $\mathbb{P}$-a.s., namely there exists a null set $\mathcal{N}\subset\Omega$ such that for all $\omega\in\mathcal{N}^c$

$$\int_s^t K_r(\omega)\mathrm{d}r=0 \qquad \forall 0\leqslant s<t.$$

Since intervals $[s,t]$ generale $\mathcal{B}(\mathbb{R}_+)$, it then follows from Dynkin's lemma that for such ω we have $K_t(\omega)=0$ for Lebesgue-almost all $t\geqslant 0$, hence $K=0$ in $L^2(B)$ by Fubini.

It remains to show the claim that X is a martingale. X is adapted (since K is progressively and integrable), and integrable by the Cauchy-Schwarz inequality (applied twice):

$$\mathbb{E}[|X_t|]\leqslant\mathbb{E}\bigg[\int_0^t K_s^2\,\mathrm{d}s\bigg]^{1/2}\sqrt{t}\leqslant\|K\|_{L^2(B)}\sqrt{t}.$$

Consider now $H=\mathbb{1}_{(s,t]}\mathbb{1}_A$ with $s<t$ and $A\in\mathcal{F}_s$. Then $H\in b\mathcal{E}$ and since $K\perp b\mathcal{E}$, we have

$$0=(H,K)_{L^2(B)}=\mathbb{E}\bigg[\int_0^\infty H_rK_r\mathrm{d}r\bigg]=\mathbb{E}\,[\mathbb{1}_A(X_t-X_s)].$$

Since $A\in\mathcal{F}_s$ was arbitrary, this proves that X is a martingale, concluding the proof. □

Theorem 6.7. (Itô integral and Itô isometry w.r.t. Brownian motion) *Let $H\in L^2(B)$. Then there exists a unique (up to indistinguishability) element of $\mathcal{H}^{2,c}$, which we denote by $(\int_0^t H_s\mathrm{d}B_s)_{t\geqslant 0}$, such that for any sequence $(H^{(n)})_n\subset b\mathcal{E}$ with $H^{(n)}\to H$ in $L^2(B)$ we have*

$$\lim_{n\to\infty}\mathbb{E}\bigg[\sup_{t\geqslant 0}\bigg|\int_0^t H_s\,\mathrm{d}B_s-\int_0^t H_s^{(n)}\mathrm{d}B_s\bigg|^2\bigg]=0.$$

We call $\int_0^\cdot H_s\mathrm{d}B_s$ the Itô integral*, or the* stochastic integral*, of H w.r.t. B. Moreover, the map $L^2(B)\ni H\mapsto\int_0^\cdot H_s\mathrm{d}B_s\in\mathcal{H}^{2,c}$ is a linear isometry, namely* Itô's isometry *holds:*

$$\mathbb{E}\bigg[\bigg(\int_0^\infty H_s\mathrm{d}B_s\bigg)^2\bigg]=\mathbb{E}\bigg[\int_0^\infty H_s^2\mathrm{d}s\bigg]. \tag{6.4}$$

Moreover the quadratic variation of $\int_0^\cdot H_s\,\mathrm{d}B_s$ is given by

$$\bigg\langle\int_0^\cdot H_s\,\mathrm{d}B_s\bigg\rangle_t=\int_0^t H_s^2\,\mathrm{d}s \qquad \forall t\in[0,\infty]. \tag{6.5}$$

Proof. Let $(H^{(n)})_n\subset b\mathcal{E}$ be such that $\|H-H^{(n)}\|_{L^2(B)}\to 0$ (since $b\mathcal{E}$ is dense in $L^2(B)$, such a sequence must exist). Since the stochastic integral on $b\mathcal{E}$ is linear

$$\mathbb{E}\bigg[\sup_{t\geqslant 0}\bigg(\int_0^t H_s^{(n)}\mathrm{d}B_s-\int_0^t H_s^{(m)}\mathrm{d}B_s\bigg)^2\bigg]\leqslant 4\|H^{(n)}-H^{(m)}\|_{L^2(B)}\to 0$$

as $m,n\to\infty$. As a consequence, the sequence $(I_BH^{(n)})_n$ is Cauchy in $\mathcal{H}^{2,c}$, which is complete, and so it must admit a limit point in $\mathcal{H}^{2,c}$; denote it by $\int_0^\cdot H_s\mathrm{d}B_s$.

It then holds

$$\left\| \int_0^{\cdot} H_s \mathrm{d}B_s \right\|_{\mathcal{H}^{2,c}}^2 = \lim_{n\to\infty} \left\| \int_0^{\cdot} H_s^n \mathrm{d}B_s \right\|_{\mathcal{H}^{2,c}}^2 = \lim_{n\to\infty} \|H^n\|_{L^2(B)} = \|H\|_{L^2(B)}$$

proving (6.4). We now want to show that the definition of $\int_0^{\cdot} H_s \mathrm{d}B_s$ does not depend on the chosen sequence $(H^{(n)})_n \subset b\mathcal{E}$; to this end, let $(\tilde{H}^{(n)})_n \subset b\mathcal{E}$ be another sequence converging to H in $L^2(B)$ (possibly $\tilde{H}^{(n)} = H^{(n)}$). Then by linearity of the stochastic integral and Itô isometry (valid on $b\mathcal{E}$)

$$\begin{aligned}\lim_{n\to\infty} \left\| \int_0^{\cdot} H_s^{(n)} \mathrm{d}B_s - \int_0^{\cdot} \tilde{H}_s^{(n)} \mathrm{d}B_s \right\|_{\mathcal{H}^{2,c}}^2 &= \lim_{n\to\infty} \left\| \int_0^{\cdot} (H_s^{(n)} - \tilde{H}_s^{(n)}) \mathrm{d}B_s \right\|_{\mathcal{H}^{2,c}}^2 \\ &= \lim_{n\to\infty} \|H^{(n)} - \tilde{H}^{(n)}\|_{L^2(B)}^2 \\ &\leqslant \lim_{n\to\infty} (\|H^{(n)} - H\|_{L^2(B)} + \|H - \tilde{H}^{(n)}\|_{L^2(B)})^2 \\ &= 0.\end{aligned}$$

Since $I_B H^{(n)} \to \int_0^{\cdot} H_s \mathrm{d}B_s$, by the above estimate the same must hold for $I_B \tilde{H}^{(n)}$ as well. Linearity of the map $H \mapsto \int_0^{\cdot} H_s \mathrm{d}B_s$ is a consequence of the same linearity on $b\mathcal{E}$, and a passage to the limit procedure. So it only remains to show (6.5).

Notice that, if $H^{(n)} \to H$ in $L^2(B)$, then by Cauchy

$$\begin{aligned}\sup_{t\geqslant 0} \left| \int_0^t H_s^2 \, \mathrm{d}s - \int_0^t (H_s^{(n)})^2 \, \mathrm{d}s \right| &\leqslant \int_0^{+\infty} |H_s^2 - (H_s^{(n)})^2| \mathrm{d}s \\ &= \int_0^{+\infty} |H_s - H_s^{(n)}| \, |H_s + H_s^{(n)}| \mathrm{d}s \\ &\leqslant \|H - H^{(n)}\|_{L^2(\mathbb{R}_+)} (\|H\|_{L^2(\mathbb{R}_+)} + \|H^{(n)}\|_{L^2(\mathbb{R}_+)}).\end{aligned}$$

Taking expectation and applying another Cauchy inequality, we get

$$\begin{aligned}\mathbb{E}\left[\sup_{t\geqslant 0} \left| \int_0^t H_s^2 \, \mathrm{d}s - \int_0^t (H_s^{(n)})^2 \, \mathrm{d}s \right| \right] &\leqslant \mathbb{E}[\|H - H^{(n)}\|_{L^2(\mathbb{R}_+)} (\|H\|_{L^2(\mathbb{R}_+)} + \|H^{(n)}\|_{L^2(\mathbb{R}_+)})] \\ &\leqslant \|H - H^{(n)}\|_{L^2(B)} (\|H\|_{L^2(B)} + \|H^{(n)}\|_{L^2(B)}) \\ &\leqslant \left(\|H\|_{L^2(B)} + \sup_n \|H^{(n)}\|_{L^2(B)} \right) \|H - H^{(n)}\|_{L^2(B)} \to 0\end{aligned}$$

as $n \to \infty$; this is because, if $H^{(n)} \to H$ in $L^2(B)$, then $\|H^{(n)}\|_{L^2(B)} \to \|H\|_{L^2(B)}$ and so $\sup_n \|H^{(n)}\|_{L^2(B)} < \infty$. Therefore the martingales $M^n = I_B H^{(n)}$, $M = I_B H \in \mathcal{H}^{2,c}$ are such that (recall the equivalent norm on $\mathcal{H}^{2,c}$ coming from Proposition 5.14)

$$\lim_{n\to\infty} \left(\mathbb{E}\left[\sup_{t\geqslant 0} |M_t^n - M_t|^2 \right] + \mathbb{E}\left[\sup_{t\geqslant 0} \left| \langle M^n \rangle_t - \int_0^t H_s^2 \, \mathrm{d}s \right| \right] \right) = 0.$$

Since $(M^n)^2 - \langle M^n \rangle$ is a martingale for each n, passing to the limit $M^2 - \int_0^t H_s^2 \, \mathrm{d}s$ is also a martingale, so that $\langle I_B H \rangle_{\cdot} = \int_0^{\cdot} H_s^2 \, \mathrm{d}s$, proving (6.5). □

Exercise. Compute $\mathbb{E}[(\int_0^t B_s \mathrm{d}B_s)^2]$ for a Brownian motion B (i.e. you must get an explicitly number at the end).

6.3 Stochastic integration w.r.t. $M \in \mathcal{H}^{2,c}$

Our next aim is to define the stochastic integral $\int_0^t H_s \mathrm{d}M_s$ for $M \in \mathcal{H}^{2,c}$ and for suitable integrands H. This can be accomplished very similarly to the Brownian case. Recall from Proposition 6.3 that, for any $H = \sum_{k=0}^{n-1} h_k \mathbb{1}_{(t_k, t_{k+1}]} \in b\mathcal{E}$ and $M \in \mathcal{H}^{2,c}$, the process $\int_0^{\cdot} H_s \, \mathrm{d}M_s$ is also in $\mathcal{H}^{2,c}$ and

$$\left\| \int_0^{\cdot} H_s \mathrm{d}M_s \right\|_{\mathcal{H}^{2,c}}^2 = \mathbb{E}\left[\int_0^{\infty} H_s^2 \, \mathrm{d}\langle M \rangle_s \right]. \tag{6.6}$$

Let now again $\bar{\Omega} := \Omega \times \mathbb{R}_+$, Prog be the progressive σ-algebra, and set

$$\mathbb{P}_M(\mathrm{d}\omega, \mathrm{d}t) := \langle M \rangle(\omega, \mathrm{d}t) \mathbb{P}(\mathrm{d}\omega),$$

By the above we mean that, for any (ω, t)-measurable and bounded function $F: \bar{\Omega} \to \mathbb{R}$,

$$\int_{\bar{\Omega}} F(\omega, t) \, \mathbb{P}_M(\mathrm{d}\omega, \mathrm{d}t) = \int_{\Omega} \int_0^{+\infty} F(\omega, t) \, \langle M \rangle(\omega, \mathrm{d}t) \, \mathbb{P}(\mathrm{d}\omega) = \mathbb{E}\left[\int_0^{+\infty} F_t \, \mathrm{d}\langle M \rangle_t \right]$$

where the last identity comes from interpreting F as a stochastic process with measurable trajectories (which is why we omit the $\omega \in \Omega$ inside $\mathbb{E}$ as usual). Notice that, for bounded F, the above quantity is finite since

$$\mathbb{E}\left[\int_0^{+\infty} |F_t| \, \mathrm{d}\langle M \rangle_t \right] \leqslant \|F\|_{L^{\infty}(\bar{\Omega})} \, \mathbb{E}\left[\int_0^{+\infty} 1 \, \mathrm{d}\langle M \rangle_t \right] = \|F\|_{L^{\infty}(\bar{\Omega})} \, \mathbb{E}[\langle M \rangle_{\infty}] < \infty$$

since $\mathbb{E}[\langle M \rangle_{\infty}] = \|M\|_{\mathcal{H}^{2,c}} < \infty$ by Corollary 5.25. Moreover set $L^2(M) := L^2(\bar{\Omega}, \mathrm{Prog}, \mathbb{P}_M)$, namely

$$L^2(M) = \left\{ H: \Omega \times \mathbb{R}_+ \to \mathbb{R} \,\middle|\, H \text{ is progressive and } \|H\|_{L^2(M)}^2 := \mathbb{E}\left[\int_0^{\infty} H_t^2 \, \mathrm{d}\langle M \rangle_t \right] < \infty \right\},$$

which is a Hilbert space with inner product

$$(H, K)_{L^2(M)} = \mathbb{E}\left[\int_0^{\infty} H_t K_t \, \mathrm{d}\langle M \rangle_t \right].$$

Remark 6.8. As before, we identify processes $H, \tilde{H}$ as the same element in $L^2(M)$ if $\|H - \tilde{H}\|_{L^2(M)} = 0$. This condition is now slightly more subtle than in the Brownian case: if $u \mapsto \langle M \rangle_u$ is constant on some interval $[s, t]$, then H can take any value therein and yet it will be identified with 0 on $[s, t]$. In the extreme case where M is constant and $\langle M \rangle \equiv 0$, any process will be identified with 0. The point is exactly that we only care about the result stochastic integral $\int_0^{\cdot} H_s \, \mathrm{d}M_s$, and if M is constant then formally "$\mathrm{d}M \equiv 0$".

With these notations in mind, the Itô isometry (6.6) shows that the map

$$I_M: b\mathcal{E} \ni H \mapsto I_M H := \int_0^{\cdot} H_s \, \mathrm{d}M_s \in \mathcal{H}^{2,c}$$

is an isometry between $(b\mathcal{E}, \|\cdot\|_{L^2(M)})$ and $\mathcal{H}^{2,c}$, and thus it can be uniquely extended to an isometry on the closure of $b\mathcal{E}$ in $L^2(M)$.

Going through the same proof line-by-line as in Lemma 6.6, one can then shows that $b\mathcal{E}$ is dense in $L^2(M)$; we only have to redefine $X := \int_0^{\cdot} K_r \, \mathrm{d}\langle M \rangle_r$ in that proof (before we took $X = \int_0^{\cdot} K_r \, \mathrm{d}r = \int_0^{\cdot} K_r \, \mathrm{d}\langle B \rangle_r$).

Exercise. Prove by yourself density of $b\mathcal{E}$ in $L^2(M)$ by following the above guidelines.

This leads to the following result:

Theorem 6.9. (Itô integral for $M \in \mathcal{H}^{2,c}$) *For $M \in \mathcal{H}^{2,c}$, there is a unique linear isometry*

$$L^2(M) \ni H \mapsto \int_0^{\cdot} H_s \, \mathrm{d}M_s \in \mathcal{H}^{2,c}$$

which is an extension of I_M. We call $\int_0^{\cdot} H_s \mathrm{d}M_s$ the stochastic integral *or* Itô integral *of H with respect to M.*

In other words, for any $H \in L^2(M)$, there exists a sequence $(H^n) \subset b\mathcal{E}$ such that $H^n \to H$ in $L^2(M)$, and for any approximating sequence $H^n \to H$ in $L^2(M)$, we have $\int_0^{\cdot} H_s^n \, \mathrm{d}M_s \to \int_0^{\cdot} H_s \, \mathrm{d}M_s$ in $\mathcal{H}^{2,c}$.

The quadratic variation of $\int_0^{\cdot} H_s \, \mathrm{d}M_s$ is given by

$$\left\langle \int_0^{\cdot} H_s \, \mathrm{d}M_s \right\rangle_t = \int_0^t H_s^2 \, \mathrm{d}\langle M \rangle_s \quad \forall t \in [0, +\infty] \tag{6.7}$$

and we have the Itô isometry

$$\mathbb{E}\left[\left(\int_0^{\tau} H_s \, \mathrm{d}M_s\right)^2\right] = \mathbb{E}\left[\int_0^{\tau} H_s^2 \, \mathrm{d}\langle M \rangle_s\right] \tag{6.8}$$

for all stopping times τ (including e.g. $\tau \equiv +\infty$).

Proof. The first part follows immediately from the fact that I_M is an isometry, $b\mathcal{E}$ is dense in $L^2(M)$, and $\mathcal{H}^{2,c}$ is complete; the second part follows arguing like in Theorem 6.7. Finally, (6.8) follows from Corollary 5.39. □

Exercise. Find another formula for $\mathbb{E}[(\int_0^t M_s \, \mathrm{d}M_s)^2]$ when $M \in \mathcal{H}^{2,c}$; compare this to the case $M = B$ by a previous exercise.

Next, we want to discuss a characterizing property of the stochastic integral (which in fact can be used to provide an alternative, more functional analytic way of constructing it). As a preparation, we need to upgrade Lemma 5.34 to a version which is more suitable for stochastic integrals; as therein, one should think of it as a version of the Cauchy-Schwarz inequality for stochastic integrals.

Lemma 6.10. (Kunita–Watanabe inequality, v2) *Let $M, N \in \mathcal{M}_{\mathrm{loc}}^c$ and let H, K be measurable processes such that almost surely $\int_0^{\infty} |H_t K_t| \mathrm{d}V(\langle M, N \rangle)_t < \infty$. Then almost surely*

$$\begin{aligned} \left| \int_0^{\infty} H_t K_t \, \mathrm{d}\langle M, N \rangle_t \right| &\leqslant \int_0^{\infty} |H_t K_t| \mathrm{d}V(\langle M, N \rangle)_t \\ &\leqslant \left(\int_0^{\infty} |H_t|^2 \mathrm{d}\langle M \rangle_t \right)^{1/2} \left(\int_0^{\infty} |K_t|^2 \mathrm{d}\langle N \rangle_t \right)^{1/2}. \end{aligned}$$

Proof.
We skipped the proof in the lectures for lack of time and because it's mostly a technical extension of Lemma 5.34 which was already shown; the full proof is included anyway in the lecture notes for completeness.

Recall from Lemma 5.34 that we already showed the inequality for $H = K = 1$, i.e.

$$|\langle M, N \rangle_{s,t}| \leqslant V(\langle M, N \rangle)_{s,t} \leqslant \langle M \rangle_{s,t}^{1/2} \langle N \rangle_{s,t}^{1/2}.$$

If now H, K are elementary processes (not necessarily $\mathbb{F}$-progressive), in the sense that

$$H = \sum_{j=0}^{n-1} h_j \mathbb{1}_{(t_j, t_{j+1}]}, \quad K = \sum_{j=0}^{n-1} k_j \mathbb{1}_{(t_j, t_{j+1}]},$$

then by definition of Lebesgue–Stjeltes integrals and Cauchy inequality we find

$$\begin{aligned}\left|\int_0^\infty H_t K_t \,\mathrm{d}\langle M,N\rangle_t\right| &= \left|\sum_{j=0}^{n-1} h_j k_j \langle M,N\rangle_{t_j,t_{j+1}}\right|\\ &\leqslant \sum_{j=0}^{n-1} |h_j|\,|k_j||\langle M,N\rangle|_{t_j,t_{j+1}}\\ &\leqslant \sum_{j=0}^{n-1} |h_j|\,|k_j|\langle M\rangle_{s,t}^{1/2}\langle N\rangle_{s,t}^{1/2}\\ &\leqslant \left(\sum_{j=0}^{n-1} |h_j|^2\langle M\rangle_{s,t}\right)^{1/2}\left(\sum_{j=0}^{n-1} |k_j|^2\langle N\rangle_{s,t}\right)^{1/2}\\ &= \left(\int_0^\infty |H_t|^2\,\mathrm{d}\langle M\rangle_t\right)^{1/2}\left(\int_0^\infty |K_t|^2\,\mathrm{d}\langle N\rangle_t\right)^{1/2}.\end{aligned}$$

By the usual approximation arguments, we can extend the inequalities to be valid e.g. for all bounded measurable processes H, K. As usual, we can then relax the boundedness assumption to almost sure finiteness of $\int_0^\infty |H_t K_t| \mathrm{d}V(\langle M,N\rangle)_t$. □

——— End of the lecture on December 19 ———

Theorem 6.11. (Characterization of the stochastic integral) *Let $M\in\mathcal{H}^{2,c}$ and $H\in L^2(M)$. The stochastic integral $\int_0^\cdot H_s\,\mathrm{d}M_s$ is the unique element in $\mathcal{H}^{2,c}$ starting at 0 such that*

$$\left\langle \int_0^\cdot H_s\,\mathrm{d}M_s, N\right\rangle = \int_0^\cdot H_s\,\mathrm{d}\langle M,N\rangle_s \quad \text{for all } N\in\mathcal{H}^{2,c}. \tag{6.9}$$

Proof. Let us first show that there can be at most one process starting at 0 satisfying (6.9): if $X,Y\in\mathcal{H}^{2,c}$ satisfy (6.9) with $\int_0^\cdot H_s\,\mathrm{d}M_s$ replaced by X or Y, then for all $N\in\mathcal{H}^{2,c}$ it holds

$$\langle X-Y,N\rangle = \langle X,N\rangle - \langle Y,N\rangle = \int_0^\cdot H_s\,\mathrm{d}\langle M,N\rangle_s - \int_0^\cdot H_s \mathrm{d}\langle M,N\rangle_s = 0;$$

taking $N = X-Y$, we deduce that $X-Y=0$.

Notice that, in the existence part, we may assume wlog that $M_0=0$. Indeed, otherwise we may replace M by $M-M_0$, and it holds $\langle M-M_0,N\rangle = \langle M,N\rangle$; therefore the uniqueness part shows that the stochastic integrals with respect to M and $M-M_0$ have to agree (loosely speaking, "$\mathrm{d}M = \mathrm{d}(M-M_0)$").

To see that $\int_0^\cdot H_s\,\mathrm{d}M_s$ satisfies (6.9), we divide the proof in two steps. Notice that we can assume wlog $M_0=0$, since M_0 does not play any role in the definition of $\int_0^\cdot H_s\,\mathrm{d}M_s$.

Step 1: $H\in b\mathcal{E}$. Let $H=\sum_{k=0}^{n-1} h_k \mathbb{1}_{(t_k,t_{k+1}]}\in b\mathcal{E}$ and let $N\in\mathcal{H}^{2,c}$. We claim that $(\int_0^\cdot H_s\,\mathrm{d}M_s)N - \int_0^\cdot H_s\,\mathrm{d}\langle M,N\rangle_s$ is a martingale, which by definition of $\langle\cdot,\cdot\rangle$ yields (6.9). Since $\langle\cdot,\cdot\rangle$ is linear in each entry, $H\mapsto\int_0^\cdot H\mathrm{d}M$ is linear and so is relation (6.9), it suffices to verify the claim for $h_k\mathbb{1}_{(t_k,t_{k+1}]}$; therefore we only need to show that

$$\left(\int_0^t H_s\,\mathrm{d}M_s\right)N_t - \int_0^t H_s\,\mathrm{d}\langle M,N\rangle_s = h_k\,(M_{t_k\wedge t,t_{k+1}\wedge t}N_t - \langle M,N\rangle_{t_k\wedge t,t_{k+1}\wedge t})$$

is a martingale. Let us set

$$J_t := M_{t_k\wedge t,t_{k+1}\wedge t}N_t - \langle M,N\rangle_{t_k\wedge t,t_{k+1}\wedge t};$$

By Proposition 5.24-c), we have

$$M_{t_{k+1}\wedge t} N_t - \langle M, N\rangle_{t_{k+1}\wedge t} = M_t^{t_{k+1}} N_t - \langle M^{t_{k+1}}, N\rangle_t$$

which is a martingale by the definition of $\langle\cdot,\cdot\rangle$; the same argument holds for t_{k+1} replaced by t_k, and so by linearity J is a martingale.

We now want to show that $h_k J$ is a martingale as well, where $h_k \in \mathcal{F}_{t_k}$ is bounded. We argue as in the proof of Proposition 6.3: we only need to consider the case $t_k \leqslant s \leqslant t$, as the case $s \leqslant t_k \leqslant t$ reduces to this one by the tower property of conditional expectation, and the case $s \leqslant t \leqslant t_k$ is trivial (everything is 0). When $t_k \leqslant s \leqslant t$, since h_k is $\mathcal{F}_{t_k}$-measurable and J is a martingale, we get

$$\mathbb{E}[h_k J_t | \mathcal{F}_s] = h_k \mathbb{E}[J_t | \mathcal{F}_s] = h_k J_s.$$

Overall, this concludes the proof of (6.9) when $H \in b\mathcal{E}$.

Step 2: extension to $H \in L^2(M)$ by density. Let $H \in L^2(M)$, then by density there exists a sequence $\{H^{(n)}\}_n \subset b\mathcal{E}$ such that $H^{(n)} \to H$ in $L^2(M)$; by Step 1, for each n,

$$\left(\int_0^\cdot H_s^{(n)} \,\mathrm{d}M_s\right) N - \int_0^\cdot H_s^{(n)} \,\mathrm{d}\langle M, N\rangle_s$$

is a martingale; to conclude, it suffices to show that, for any fixed $t \geqslant 0$,

$$\left(\int_0^t H_s^{(n)} \,\mathrm{d}M_s\right) N_t - \int_0^t H_s^{(n)} \,\mathrm{d}\langle M, N\rangle_s \to \left(\int_0^t H_s \,\mathrm{d}M_s\right) N - \int_0^t H_s \,\mathrm{d}\langle M, N\rangle_s \quad \text{in } L^1(\Omega).$$

Convergence of the first term is immediate, since by construction of the stochastic integral we have $\int_0^\cdot H^{(n)} \,\mathrm{d}M \to \int_0^\cdot H \,\mathrm{d}M$ in $\mathcal{H}^{2,c}$. For the second term, by linearity and the Kunita–Watanabe inequality (with $\tilde{H} := (H^{(n)} - H)\,\mathbb{1}_{[0,t]}$, $K = 1$) we have

$$\begin{aligned}
\mathbb{E}\left[\left|\int_0^t H_s^{(n)} \,\mathrm{d}\langle M, N\rangle_s - \int_0^t H_s \,\mathrm{d}\langle M, N\rangle_s\right|\right] &= \mathbb{E}\left[\left|\int_0^{+\infty} \tilde{H}_s \,\mathrm{d}\langle M, N\rangle_s\right|\right]\\
&\leqslant \mathbb{E}\left[\left(\int_0^t |H_s^{(n)} - H_s|^2 \mathrm{d}\langle M\rangle_s\right)^{1/2} \langle N\rangle_\infty^{1/2}\right]\\
&\leqslant \mathbb{E}\left[\int_0^t |H_s^{(n)} - H_s|^2 \mathrm{d}\langle M\rangle_s\right] \mathbb{E}[\langle N\rangle_\infty]^{1/2}\\
&\leqslant \|H^n - H\|_{L^2(M)} \|N\|_{\mathcal{H}^{2,c}} \to 0
\end{aligned}$$

as $n \to \infty$, where in the intermediate passage we applied Cauchy's inequality. □

Exercise. Find a formula for $\langle \int_0^\cdot B_s \,\mathrm{d}B_s, B\rangle_t$, where B is a Brownian motion B, and compute $\mathbb{E}[\langle \int_0^\cdot B_s \mathrm{d}B_s, B\rangle_t]$.

Corollary 6.12. *Let $M \in \mathcal{H}^{2,c}$ and $H \in L^2(M)$.*

i. For any stopping time τ, it holds that

$$\left(\int_0^\cdot H_s \,\mathrm{d}M_s\right)^\tau = \int_0^\cdot \mathbb{1}_{[0,\tau]}(s) H_s \,\mathrm{d}M_s = \int_0^\cdot H_s \,\mathrm{d}M_s^\tau.$$

ii. For any other $N \in \mathcal{H}^{2,c}$ and $K \in L^2(N)$, we have

$$\left\langle \int_0^\cdot H_s \,\mathrm{d}M_s, \int_0^\cdot K_s \,\mathrm{d}N_s \right\rangle = \int_0^\cdot H_s K_s \,\mathrm{d}\langle M, N\rangle_s. \tag{6.10}$$

Notice that, for $M = N$, $H = K$ we recover formula (6.7).

iii. Associativity of the stochastic integral*: Let K be progressively measurable. Then $KH \in L^2(M)$ if and only if $K \in L^2(\int_0^\cdot H_s\,\mathrm{d}M_s)$; in that case, we have*

$$\int_0^\cdot K_s H_s\,\mathrm{d}M_s = \int_0^\cdot K_s\,\mathrm{d}\Big(\int_0^\cdot H_r\,\mathrm{d}M_r\Big)_s.$$

Proof.

i. We apply our characterization of stochastic integrals, Theorem 6.11: for all $N \in \mathcal{H}^{2,c}$

$$\begin{aligned}\Big\langle\Big(\int_0^\cdot H_s\,\mathrm{d}M_s\Big)^\tau, N\Big\rangle &= \Big\langle\int_0^\cdot H_s\,\mathrm{d}M_s, N\Big\rangle^\tau = \Big(\int_0^\cdot H_s\,\mathrm{d}\langle M, N\rangle_s\Big)^\tau \\ &= \int_0^\cdot \mathbb{1}_{[0,\tau]}(s) H_s\,\mathrm{d}\langle M, N\rangle_s \\ &= \int_0^\cdot H_s\,\mathrm{d}\langle M, N\rangle_s^\tau \\ &= \int_0^\cdot H_s\,\mathrm{d}\langle M^\tau, N\rangle_s.\end{aligned}$$

The second line shows that $(\int_0^\cdot H_s\,\mathrm{d}M_s)^\tau = \int_0^\cdot \mathbb{1}_{[0,\tau]}(s) H_s\,\mathrm{d}M_s$ and the last line shows that $(\int_0^\cdot H_s\,\mathrm{d}M_s)^\tau = \int_0^\cdot H_s\,\mathrm{d}M_s^\tau$.

ii. The Lebesgue-Stieltjes integral on the r.h.s. of (6.10) is well defined by the Kunita-Watanabe inequality. To prove (6.10), we apply Theorem 6.11 twice, together with the associativity of the Lebesgue-Stieltjes integral (Remark 5.9):

$$\begin{aligned}\Big\langle\int_0^\cdot H_s\,\mathrm{d}M_s, \int_0^\cdot K_s\,\mathrm{d}N_s\Big\rangle &= \int_0^\cdot H_s\,\mathrm{d}\Big\langle M, \int_0^\cdot K_r\,\mathrm{d}N_r\Big\rangle_s \\ &= \int_0^\cdot H_s\,\mathrm{d}\Big(\int_0^\cdot K_r\,\mathrm{d}\langle M, N\rangle_r\Big)_s \\ &= \int_0^\cdot H_s K_s\,\mathrm{d}\langle M, N\rangle_s.\end{aligned}$$

iii. We have $\langle\int_0^\cdot H_s\,\mathrm{d}M_s\rangle = \int_0^\cdot H_s^2\,\mathrm{d}\langle M\rangle_s$, so again by Remark 5.9

$$\int_0^\cdot K_s^2\,\mathrm{d}\Big(\int_0^\cdot H_r^2\,\mathrm{d}\langle M\rangle_r\Big)_s = \int_0^\cdot (K_s H_s)^2\mathrm{d}\langle M\rangle_s;$$

so we see that $K \in L^2(\int_0^\cdot H_s\,\mathrm{d}M_s)$ if and only if $KH \in L^2(M)$. Moreover, for any $N \in \mathcal{H}^{2,c}$, similarly as in Part ii. we have:

$$\begin{aligned}\Big\langle\int_0^\cdot K_s H_s\,\mathrm{d}M_s, N\Big\rangle &= \int_0^\cdot K_s H_s\,\mathrm{d}\langle M, N\rangle_s = \int_0^\cdot K_s\,\mathrm{d}\Big(\int_0^\cdot H_r\,\mathrm{d}\langle M, N\rangle_r\Big)_s \\ &= \int_0^\cdot K_s\,\mathrm{d}\Big\langle\int_0^\cdot H_r\,\mathrm{d}M_r, N\Big\rangle_s = \Big\langle\int_0^\cdot K_s\,\mathrm{d}\Big(\int_0^\cdot H_r\,\mathrm{d}M_r\Big)_s, N\Big\rangle,\end{aligned}$$

which by Theorem 6.11 implies the conclusion. □

Exercise. Let B, $\tilde{B}$ be independent Brownian motions; show that, for any $t \geqslant 0$, $B \in L^2(\tilde{B}^t)$ and $\tilde{B} \in L^2(B^t)$, and compute

$$\Big\langle\int_0^\cdot B_s\,\mathrm{d}\tilde{B}_s, \int_0^\cdot \tilde{B}_s\,\mathrm{d}B_s\Big\rangle_t.$$

Exercise. Let B be a Brownian motion, let $T>0$ and let $H\in L^2(B^T)$. Find a formula for

$$\left\langle \int_0^\cdot H_r\,\mathrm{d}B_r^T \right\rangle_t \quad \text{for } t\geqslant 0.$$

Alternative construction of stochastic integrals (not examinable): The arguments from Theorem 6.11 and Corollary 6.12, and in particular formula (6.9), can be actually used to directly *construct* stochastic integrals in a more functional analytic manner, completely bypassing the approximations by bounded elementary processes. The argument goes as follows.

Given $M\in\mathcal{H}^{2,c}$ and $H\in L^2(M)$, we can define a linear operator $\mathcal{I}:\mathcal{H}^{2,c}\to\mathbb{R}$ by

$$\mathcal{I}(N):=\mathbb{E}\left[\int_0^\infty H_s\,\mathrm{d}\langle M,N\rangle_s\right]. \tag{6.11}$$

This is a bounded operator, since by the Yamada–Watanabe and Cauchy inequalities it holds

$$|\mathcal{I}(N)|\leqslant\mathbb{E}\left[\left(\int_0^{+\infty} H_s^2\,\mathrm{d}\langle M\rangle_s\right)^{1/2}\langle N\rangle_\infty^{1/2}\right]\leqslant\|H\|_{L^2(M)}\|N\|_{\mathcal{H}^{2,c}}.$$

Since $\mathcal{H}^{2,c}$ is a Hilbert space, by the Riesz representation theorem there exists a unique element $\tilde{M}\in\mathcal{H}^{2,c}$ such that $\mathcal{I}(N)=\langle M,N\rangle_{\mathcal{H}^{2,c}}$ for all $N\in\mathcal{H}^{2,c}$, namely

$$\mathbb{E}[\tilde{M}_\infty N_\infty]=\mathbb{E}\left[\int_0^\infty H_s\,\mathrm{d}\langle M,N\rangle_s\right]\quad\forall N\in\mathcal{H}^{2,c}.$$

Let τ be any bounded stopping time; applying the above property to N replaced by N^τ, arguing as in Corollary 6.12 and using the optional sampling theorem, one finds

$$\begin{aligned}\mathbb{E}[\tilde{M}_\tau N_\tau] &= \mathbb{E}[\tilde{M}_\infty N_\infty^\tau]\\ &= \mathbb{E}\left[\int_0^\infty H_s\,\mathrm{d}\langle M,N^\tau\rangle_s\right]\\ &= \mathbb{E}\left[\int_0^\tau H_s\,\mathrm{d}\langle M,N\rangle_s\right].\end{aligned}$$

By Exercise Sheet 6, this implies that $\tilde{M}N-\int_0^\cdot H_s\,\mathrm{d}\langle M,N\rangle_s$ is a (continuous) martingale; moreover one can show that $\tilde{M}_0=0$ by taking $N_t(\omega)=\mathbb{1}_A(\omega)$ for $A\in\mathcal{F}_0$. Therefore

$$\langle\tilde{M},N\rangle=\int_0^\cdot H_s\,\mathrm{d}\langle M,N\rangle_s\quad\forall N\in\mathcal{H}^{2,c},$$

which as we just saw characterizes the stochastic integral. In other words, $\tilde{M}$ coming from the application of Riesz theorem to $\mathcal{I}$ defined by (6.11) coincides with $\int_0^\cdot H_s\,\mathrm{d}M_s$.

6.4 Stochastic integration w.r.t. $M\in\mathcal{M}_{\mathrm{loc}}^c$

We can now use localization arguments to extend our definition of stochastic integrals to integrands M which are only continuous local martingales, not necessarily in $\mathcal{H}^{2,c}$. And even for $M\in\mathcal{H}^{2,c}$, the following considerations allow us to consider more general integrands than $H\in L^2(M)$.

Definition 6.13. **($L^2_{\mathrm{loc}}(M)$)** *For $M\in\mathcal{M}_{\mathrm{loc}}^c$, we denote by $L^2_{\mathrm{loc}}(M)$ the space of progressively measurable processes H which satisfy*

$$\mathbb{P}\left(\int_0^T H_s^2\,\mathrm{d}\langle M\rangle_s<\infty\right)=1,\qquad \textit{for all } T\geqslant 0.$$

Exercise. Show that any càdlàg adapted process H is in $L^2_{\mathrm{loc}}(M)$, for any $M\in\mathcal{M}^c_{\mathrm{loc}}$.

Theorem 6.14. (Localization of the stochastic integral) *Let $M\in\mathcal{M}^c_{\mathrm{loc}}$ and let $H\in L^2_{\mathrm{loc}}(M)$. Then:*

i. There exists a unique process $X\in\mathcal{M}^c_{\mathrm{loc}}$ such that $X_0=0$ and

$$\langle X,N\rangle=\int_0^\cdot H_s\,\mathrm{d}\langle M,N\rangle_s\quad\forall N\in\mathcal{M}^c_{\mathrm{loc}}.$$

We write $\int_0^\cdot H_s\,\mathrm{d}M_s:=X$ and call this process the Itô integral *or* stochastic integral *of H against M.*

ii. For any stopping time τ, we have

$$\int_0^\cdot H_s\mathbb{1}_{[0,\tau]}(s)\mathrm{d}M_s=\left(\int_0^\cdot H_s\,\mathrm{d}M_s\right)^\tau=\int_0^\cdot H_s\,\mathrm{d}M^\tau_s.$$

iii. For any other $N\in\mathcal{M}^c_{\mathrm{loc}}$ and $K\in L^2_{\mathrm{loc}}(N)$, we have

$$\left\langle\int_0^\cdot H_s\,\mathrm{d}M_s,\int_0^\cdot K_s\,\mathrm{d}N_s\right\rangle=\int_0^\cdot H_sK_s\,\mathrm{d}\langle M,N\rangle_s.\tag{6.12}$$

iv. For progressive K, we have $K\in L^2_{\mathrm{loc}}(\int_0^\cdot H_s\,\mathrm{d}M_s)$ if and only $KH\in L^2_{\mathrm{loc}}(M)$; in that case

$$\int_0^\cdot K_sH_s\,\mathrm{d}M_s=\int_0^\cdot K_s\,\mathrm{d}\left(\int_0^\cdot H_r\,\mathrm{d}M_r\right)_s.$$

v. If $M\in\mathcal{H}^{2,c}$ and $H\in L^2(M)$, then $\int_0^\cdot H_s\,\mathrm{d}M_s$ is the same process that we constructed in Theorem 6.9; in other words, this notion of stochastic integral is a consistent extension of the previous one.

Proof. i.: Uniqueness is clear by the usual argument: if $X,Y\in\mathcal{M}^c_{\mathrm{loc}}$ satisfy $X_0=Y_0=0$ and $\langle X,N\rangle=\langle Y,N\rangle$ for all $N\in\mathcal{M}^c_{\mathrm{loc}}$, then $X-Y\in\mathcal{M}^c_{\mathrm{loc}}$ with $\langle X-Y\rangle=0$, and therefore $X-Y\equiv0$.

For the construction of $\int_0^\cdot H_s\,\mathrm{d}M_s$, as before we may assume that $M_0=0$; otherwise, we can consider $M-M_0$, which satisfies $\langle M,N\rangle=\langle M-M_0,N\rangle$.

Let

$$\tau_n=\inf\left\{t\geqslant0\colon\int_0^t(1+H_s^2)\,\mathrm{d}\langle M\rangle_s\geqslant n\right\},$$

which is a localizing sequence by the definition of $H\in L^2_{\mathrm{loc}}(M)$. Then $M^{\tau_n}\in\mathcal{H}^{2,c}$ (by Corollary 5.40) and $H\in L^2(M^{\tau_n})$; therefore we can define the stochastic integral $\int_0^\cdot H_s\,\mathrm{d}M^{\tau_n}_s\in\mathcal{H}^{2,c}$ using Theorem 6.9. For $m>n$, we get

$$\left(\int_0^\cdot H_s\,\mathrm{d}M^{\tau_m}_s\right)^{\tau_n}=\int_0^\cdot H_s\,\mathrm{d}(M^{\tau_m})^{\tau_n}_s=\int_0^\cdot H_s\,\mathrm{d}M^{\tau_n}_s,$$

so that we can define without ambiguity

$$\left(\int_0^t H_s\,\mathrm{d}M_s\right)\mathbb{1}_{\{t\leqslant\tau_n\}}:=\left(\int_0^t H_s\,\mathrm{d}M^{\tau_n}_s\right)\mathbb{1}_{\{t\leqslant\tau_n\}}$$

to obtain a unique process $\int_0^\cdot H_s\,\mathrm{d}M_s$ such that $(\int_0^\cdot H_s\,\mathrm{d}M_s)^{\tau_n}=\int_0^\cdot H_s\,\mathrm{d}M^{\tau_n}_s$ for all n. In particular, $\int_0^\cdot H_s\,\mathrm{d}M_s\in\mathcal{M}^c_{\mathrm{loc}}$ by definition, with localizing sequence being given exactly by $\{\tau_n\}_n$; we also have $\int_0^0 H_s\,\mathrm{d}M_s=0$ by definition.

Given $N\in\mathcal{M}^{c}_{\mathrm{loc}}$, similarly we may assume without loss of generality that $N_0=0$, since the quadratic covariation does not depend on N_0. Let $\tau_n'=\inf\{t\geqslant 0\colon \langle N\rangle_t\geqslant n\}$ and $\sigma_n=\tau_n\wedge\tau_n'$. Then $N^{\sigma_n}\in\mathcal{H}^{2,c}$ and

$$\left\langle \int_0^\cdot H_s\,\mathrm{d}M_s, N\right\rangle^{\sigma_n} = \left\langle \left(\int_0^\cdot H_s\,\mathrm{d}M_s\right)^{\sigma_n}, N^{\sigma_n}\right\rangle = \left\langle \int_0^\cdot H_s\mathrm{d}M_s^{\sigma_n}, N^{\sigma_n}\right\rangle$$
$$= \int_0^\cdot H_s\,\mathrm{d}\langle M^{\sigma_n}, N^{\sigma_n}\rangle_s = \int_0^\cdot H_s\,\mathrm{d}\langle M,N\rangle_s^{\sigma_n} = \left(\int_0^\cdot H_s\,\mathrm{d}\langle M,N\rangle_s\right)^{\sigma_n},$$

so for $n\to\infty$ we get $\langle \int_0^\cdot H_s\,\mathrm{d}M_s, N\rangle = \int_0^\cdot H_s\,\mathrm{d}\langle M,N\rangle_s$.

Properties ii. and iv. of the integral $\int_0^\cdot H_s\,\mathrm{d}M_s$ then follow via localization from Corollary 6.12; iii. follows from applying part i. twice together with the associativity of Lebesgue-Stjeltes integrals. Finally, v. comes from the characterizing property of Theorem 6.11. □

Exercise. Let B be a Brownian motion, $H\in L^2_{\mathrm{loc}}(B)$. What can we say about $\mathbb{E}[\int_0^1 H_s\,\mathrm{d}B_s]$?

Remark 6.15. Let $M\in\mathcal{M}^{c}_{\mathrm{loc}}$ and $H\in L^2_{\mathrm{loc}}(M)$. Then $\int_0^\cdot H_s\,\mathrm{d}M_s\in\mathcal{M}^{c}_{\mathrm{loc}}$ is a continuous local martingale starting from 0 and by (6.12) (with $M=N$, $H=K$), its quadratic variation is given by

$$\left\langle \int_0^\cdot H_s\,\mathrm{d}M_s\right\rangle_t = \int_0^t H_s^2\,\mathrm{d}\langle M\rangle_s. \tag{6.13}$$

By Corollary 5.40, we deduce the following: if

$$\mathbb{E}\left[\int_0^t H_s^2\,\mathrm{d}\langle M\rangle_s\right]<\infty \quad \forall t\geqslant 0,$$

then $\int_0^\cdot H_s\,\mathrm{d}M_s\in\mathcal{M}^{2,c}$ (and not just $\mathcal{M}^{c}_{\mathrm{loc}}$); being a genuine martingale, it satisfies

$$\mathbb{E}\left[\int_0^t H_s\,\mathrm{d}M_s\right]=0, \qquad \mathbb{E}\left[\left(\int_0^t H_s\,\mathrm{d}M_s\right)^2\right]=\mathbb{E}\left[\int_0^t H_s^2\,\mathrm{d}\langle M\rangle_s\right] \qquad \forall t\geqslant 0.$$

If additionally

$$\mathbb{E}\left[\int_0^\infty H_s^2\,\mathrm{d}\langle M\rangle_s\right]<\infty,$$

then $\int_0^\cdot H_s\,\mathrm{d}M_s\in\mathcal{H}^{2,c}$. In that case, we may write $H\in L^2(M)$, even though $M\notin\mathcal{H}^{2,c}$.

——— End of the lecture on January 8 ———

Exercise. Strengthen the above result as follows: if $M\in\mathcal{M}^{c}_{\mathrm{loc}}$, $H\in L^2_{\mathrm{loc}}(M)$ are such that

$$\mathbb{E}\left[\left(\int_0^t H_s^2\,\mathrm{d}\langle M\rangle_s\right)^{1/2}\right] \quad \forall t\geqslant 0$$

then $\int_0^\cdot H_s\,\mathrm{d}M_s$ is a genuine martingale, and in particular $\mathbb{E}[\int_0^t H_s\,\mathrm{d}M_s]=0$ for all $t\geqslant 0$.

Example 6.16.

i. If B is a Brownian motion and $G,H\in L^2_{\mathrm{loc}}(B)$, then

$$\left\langle \int_0^\cdot G_s\,\mathrm{d}B_s, \int_0^\cdot H_s\,\mathrm{d}B_s\right\rangle_t = \int_0^t G_sH_s\,\mathrm{d}s;$$

in particular, we get a natural extension of formula (6.7):

$$\left\langle \int_0^{\cdot} H_s \,\mathrm{d}B_s \right\rangle_t = \int_0^t H_s^2 \,\mathrm{d}s.$$

ii. Every càdlàg adapted process H belongs $L^2_{\mathrm{loc}}(M)$, for all $M \in \mathcal{M}^c_{\mathrm{loc}}$. But in general we do not have $\mathbb{E}[\int_0^t H_s^2 \,\mathrm{d}\langle M\rangle_s] < \infty$. Consider for example a Brownian motion B and the integrand $H_s = e^{B_s^4}$. Then

$$\mathbb{E}\left[\int_0^t H_s^2 \,\mathrm{d}s\right] = \mathbb{E}\left[\int_0^t e^{2B_s^4} \,\mathrm{d}s\right] = \int_0^t \left(\frac{1}{\sqrt{2\pi s}} \int_{\mathbb{R}} e^{2x^4} e^{-\frac{x^2}{2s}} \,\mathrm{d}x\right) \mathrm{d}s,$$

and the inner integral is infinite for all $s > 0$.

6.5 Stochastic integration w.r.t. continuous semimartingales

Recall that a continuous semimartingale X is an adapted process

$$X = X_0 + M + A,$$

where $M \in \mathcal{M}^c_{\mathrm{loc}}$ with $M_0 = 0$ and $A \in \mathcal{A}$ with $A_0 = 0$, and that this decomposition is unique because $\mathcal{A} \cap \mathcal{M}^c_{\mathrm{loc}} = \{0\}$ up to indistinguishability.

Definition 6.17. *Let $X = X_0 + M + A$ be a continuous semimartingale. We define*

$$\mathbb{L}(X) := \left\{ H \in L^2_{\mathrm{loc}}(M) \colon \int_0^t |H_s| \mathrm{d}V(A)_s < \infty \text{ almost surely for all } t \geqslant 0 \right\},$$

or equivalently

$$\mathbb{L}(X) := \left\{ H \text{ progressive: } \int_0^t |H_s| \mathrm{d}V(A)_s + \int_0^t H_s^2 \,\mathrm{d}\langle M\rangle < \infty \text{ almost surely for all } t \geqslant 0 \right\}.$$

For $H \in \mathbb{L}(X)$, we define

$$\int_0^{\cdot} H_s \,\mathrm{d}X_s := \int_0^{\cdot} H_s \,\mathrm{d}M_s + \int_0^{\cdot} H_s \,\mathrm{d}A_s$$

where the first term is interpreted as in the stochastic sense coming from Theorem 6.14, while the second term is interpreted in the Lebesgue-Stjeltes sense.

Remark 6.18. Let H be progressively measurable and *locally bounded*, in the sense that

$$\sup_{t \in [0,T]} |H_t(\omega)| < +\infty \quad \forall T \in (0, +\infty)$$

for $\mathbb{P}$-a.e. ω. Then $H \in \mathbb{L}(X)$ for every continuous semimartingale X.

Lemma 6.19. *The following hold:*

i. Let $H_s(\omega) = \sum_{k=0}^{n-1} h_k(\omega) \mathbb{1}_{(t_k, t_{k+1}]}(s)$, for some real-valued $\mathcal{F}_{t_k}$–measurable random variables h_k. Then $H \in \mathbb{L}(X)$ for any continuous semimartingale X and

$$\int_0^t H_s \,\mathrm{d}X_s = \sum_{k=0}^{n-1} h_k (X_{t_{k+1} \wedge t} - X_{t_k \wedge t}). \tag{6.14}$$

ii. *Let τ be a stopping time and let h a real-valued, $\mathcal{F}_\tau$-measurable random variable; let $H := h\mathbb{1}_{[\tau,\infty)}$ (with the convention that $\mathbb{1}_{[\tau,\infty)} \equiv 0$ when $\tau = +\infty$). Then $H \in \mathbb{L}(X)$ for any continuous semimartingale X and*

$$\int_0^t H_s \,\mathrm{d}X_s = h(X_t - X_{\tau\wedge t}).$$

Proof. Since h_k are real-valued and the sum is finite, it is clear that $\sup_{t\geqslant 0}|H_t(\omega)| < \infty$ for every ω, so that $H \in \mathbb{L}(X)$; similarly for $H = h\mathbb{1}_{[\tau,\infty)}$ in *ii.* Let $X = X_0 + M + A$; both identities are true for integration w.r.t. A by (ω-wise) properties of Lebesgue-Stjeltes integral, and X_0 does not play any role, therefore wlog we may assume $X = M$ with $M \in \mathcal{M}^c_{\text{loc}}$, $M_0 = 0$. By localization, we may further reduce ourselves to the case of $M \in \mathcal{H}^{2,c}$.

In part *ii.*, up to another localization/approximation procedure (cf. Corollary 5.41, or the upcoming Section 6.6), we may further assume that $h \in L^\infty(\Omega)$.

The rest of the proof is left as an exercise in Exercise Sheet 10. □

Overall in this chapter we have constructed various stochastic integrals, under different assumptions on the integrand H and the integrator M as "input variables", obtaining an integral process $\int_0^\cdot H_s \,\mathrm{d}M_s$ in different classes of processes as an "output". In the following table, we summarize these results; we write $\mathcal{S}^c$ for the space of continuous semimartingales.

$H \in \cdot$	$M \in \cdot$	$\int_0^\cdot H_s \,\mathrm{d}M_s \in \cdot$
$L^2(M)$	$\mathcal{H}^{2,c}$	$\mathcal{H}^{2,c}$
$L^2_{\text{loc}}(M)$	$\mathcal{M}^c_{\text{loc}}$	$\mathcal{M}^c_{\text{loc}}$
$L^2(M)$	$\mathcal{M}^c_{\text{loc}}$	$\mathcal{H}^{2,c}$
$L^2_{\text{loc}}(M)$ + $\mathbb{E}\left[\int_0^t H_s^2 \mathrm{d}\langle M\rangle_s\right] < \infty \ \forall t \geqslant 0$	$\mathcal{M}^c_{\text{loc}}$	$\mathcal{M}^{2,c}$
$\mathbb{L}(M)$	$\mathcal{S}^c$	$\mathcal{S}^c$

6.6 Approximations of stochastic integrals

Recall the ucp-convergence from Definition 5.15; it is a natural notion of convergence for stochastic integrals, as the next results show.

Proposition 6.20. ("Dominated convergence" for stochastic integrals) *Let X be a continuous semimartingale; let $\{H^{(n)}\}_n, H$ be progressively measurable and such that $\mathbb{P}$-almost surely*

$$H_t^{(n)} \to H_t \qquad \forall t \geqslant 0.$$

Further assume that there exists $K \in \mathbb{L}(X)$ with $K \geqslant 0$, such that $\mathbb{P}$-almost surely

$$\left|H_t^{(n)}\right| \leqslant K_t \qquad \forall t \geqslant 0. \tag{6.15}$$

Then $(H^{(n)}), H \subset \mathbb{L}(X)$ and $\int_0^\cdot H_s^{(n)} \mathrm{d}X_s \to \int_0^\cdot H_s \,\mathrm{d}X_s$ in ucp.

Proof. Since $\left|H_s^{(n)}\right| \leqslant K_s$ and $|H_s| \leqslant K_s$ with $K \in \mathbb{L}(X)$, we clearly have $H^{(n)}, H \in \mathbb{L}(X)$. Let $X = X_0 + M + A$ be the semimartingale decomposition of X; to show the ucp convergence, we similarly decompose $\int_0^\cdot H_s^{(n)} \mathrm{d}X_s = \int_0^\cdot H_s^{(n)} \mathrm{d}M_s + \int_0^\cdot H_s^{(n)} \mathrm{d}A_s$ and treat the two terms separately.

For the finite variation terms $\int_0^{\cdot} H_s^{(n)} \mathrm{d}A_s$, the ucp convergence follows by the (usual) dominated convergence theorem applied ω-wise (so that one gets $\mathbb{P}$-a.s. convergence uniformly on compact sets, which is stronger than ucp convergence).

For the local martingale part, note that by (6.13) we have

$$\left\langle \int_0^{\cdot} H_s^{(n)} \mathrm{d}M_s - \int_0^{\cdot} H_s \,\mathrm{d}M_s \right\rangle = \left\langle \int_0^{\cdot} \left(H_s^{(n)} - H_s\right) \mathrm{d}M_s \right\rangle = \int_0^{\cdot} \left|H_s^{(n)} - H_s\right|^2 \mathrm{d}\langle M\rangle_s;$$

as before, by the usual dominated convergence theorem applied ω-wise, we deduce that the above quadratic variation is converging to 0 in ucp as $n \to \infty$. It then follows from Corollary 5.41 that $\int_0^{\cdot} H_s^{(n)} \mathrm{d}M_s \to \int_0^{\cdot} H_s \,\mathrm{d}M_s$ in ucp. □

Remark 6.21. A closer look at the proof reveals that we never need K to be progressive, as long as it is nonnegative, with measurable trajectories and such that $\mathbb{P}$-a.s.

$$\int_0^T K_s \,\mathrm{d}V(A)_s + \int_0^T |K_s|^2 \,\mathrm{d}\langle M\rangle_s < \infty \quad \forall T \geqslant 0.$$

Indeed, we never integrate K w.r.t. M, but we only use K at the level of ω-wise defined Lebesgue-Stjeltes integrals (e.g. when checking that $H \in \mathbb{L}(X)$, or when applying classical dominated convergence). In particular, assumption (6.15) holds if for almost all ω and all $T \geqslant 0$ there exists $C_T(\omega)$ with $|K_t(\omega)| \leqslant C_T(\omega)$ for all $t \in [0, T]$. Indeed, in this case one may take

$$K_t(\omega) := \sum_{n=1}^{\infty} C_n(\omega) \,\mathbb{1}_{[n-1,\infty)}(t).$$

Corollary 6.22. ("Stochastic integrals respect ucp convergence") *Let X be a continuous semimartingale; let $\{H^{(n)}\}_n, H$ be continuous, adapted processes such that*

$$H^{(n)} \to H \quad \text{in ucp.}$$

Then $\int_0^{\cdot} H_s^{(n)} \mathrm{d}X_s \to \int_0^{\cdot} H_s \,\mathrm{d}X_s$ in ucp.

Proof. Since $H^{(n)}$ is continuous and adapted, it is progressive and therefore (by continuity) $H^{(n)} \in \mathbb{L}(X)$; similarly for H.

Since $H^{(n)} \to H$ in ucp, by Lemma 5.16 we can extract a subsequence $\{H^{(n_k)}\}_k$ such that $\mathbb{P}$-a.s. $H^{(n_k)}(\omega) \to H(\omega)$ uniformly on compact sets. As a consequence of the properties of uniform convergence on compact sets (see exercise below), we deduce that the process

$$K_t := \sup_{n \in \mathbb{N}} \left|H_t^{(n_k)}\right| + |H_t|$$

is continuous and adapted, thus in $\mathbb{L}(X)$, and assumption (6.15) is satisfies. We deduce from Proposition 6.20 that $\int_0^{\cdot} H_s^{(n_k)} \,\mathrm{d}X_s \to \int_0^{\cdot} H_s \,\mathrm{d}X_s$ in ucp.

General fact from metric spaces: given a metric space (E, d), a sequence $\{x^n\}_n \subset E$ and $x \in E$, the following are equivalent:

i. $x^n \to x$ in E, namely $d(x^n, x) \to 0$ as $n \to \infty$.

ii. Any subsequence $\{x^{n_j}\}_j$ of $\{x^n\}_n$ admits a further subsequence $\{x^{n_{j_k}}\}_k$ such that $x^{n_{j_k}} \to x$ in E, namely $d(x^{n_{j_k}}, x) \to 0$ as $n \to \infty$.

By the same argument as above, for any other given subsequence $\{H^{(n_j)}\}_j$, we can extract a further subsequence $\{H^{(n_{j_k})}\}_k$ such that $\int_0^\cdot H_s^{(\tilde{n}_{j_k})}\,\mathrm{d}X_s \to \int_0^\cdot H_s\,\mathrm{d}X_s$ in ucp; since the upc topology is induced by a distance (Lemma 5.16), the above fact implies that the whole sequence $\{\int_0^\cdot H_s^{(n)}\,\mathrm{d}X_s\}_n$ converges. □

Exercise. Let $\{f^n\}_n$, f be deterministic continuous functions from $\mathbb{R}_+$ to $\mathbb{R}$ such that $f^n \to f$ uniformly on compact sets. Show that

$$t \mapsto \sup_{n \geqslant 0} |f_t^n|$$

is a continuous function, bounded on compact sets.

Corollary 6.23. *Let X be a continuous semimartingale and let H be a continuous, adapted process. Let $\pi^n = \{t_k^n\}_{k \geqslant 0}$ be a sequence of deterministic, locally finite partitions with infinitesimal mesh. Then*

$$\left(\sum_{k=0}^{\infty} H_{t_k^n}(X_{t_{k+1}^n \wedge t} - X_{t_k^n \wedge t})\right)_{t \geq 0} \longrightarrow \int_0^\cdot H_s\,\mathrm{d}X_s \qquad \textit{in ucp.}$$

Proof. This is a special case of Proposition 6.20, with $K_t := \sup_{s \in [0,t]} |H_s|$ and

$$H^n = \sum_{k=0}^{\infty} H_{t_k^n} \mathbb{1}_{(t_k^n, t_{k+1}^n]}$$

(cf. identity (6.14)). □

The next result already provides one important "rule of calculus" for stochastic integrals; we will see later a far reaching generalization.

Theorem 6.24. (Integration by parts formula for stochastic integrals) *Let X, Y be continuous semimartingales. Then, up to indistinguishability, the following* integration by parts *formula holds:*

$$X_t Y_t = X_0 Y_0 + \int_0^t X_s\,\mathrm{d}Y_s + \int_0^t Y_s\,\mathrm{d}X_s + \langle X, Y \rangle_t \qquad \forall t \geqslant 0. \tag{6.16}$$

In particular, for $X = Y$ we find

$$X_t^2 = X_0^2 + 2\int_0^t X_s\,\mathrm{d}X_s + \langle X \rangle_t \qquad \forall t \geqslant 0. \tag{6.17}$$

Proof. Exercise Sheet 10. □

Formula (6.16) may be formally written in differential form as

$$\text{“}\mathrm{d}(XY) = X\mathrm{d}Y + Y\mathrm{d}X + \mathrm{d}\langle X, Y \rangle.\text{”}$$

Exercise. Using formula (6.17), are you now able to show that if X is a continuous semimartingale, then X^2 is a continuous semimartingale?

Corollary 6.25. *Let X be a continuous semimartingale. Then for any $n \in \mathbb{N}$, up to indistinguishability, it holds that*

$$X_t^n = X_0^n + \int_0^t n X_s^{n-1}\mathrm{d}X_s + \frac{1}{2}\int_0^t n(n-1) X_s^{n-2}\mathrm{d}\langle X \rangle_s \qquad \forall t \geqslant 0.$$

With $f(x) = x^n$, we can also write the above formula as

$$f(X_t) = f(X_0) + \int_0^t f'(X_s) \mathrm{d}X_s + \frac{1}{2} \int_0^t f''(X_s) \mathrm{d}\langle X \rangle_s.$$

Proof. Exercise Sheet 10. □

For practical purposes, Corollary 6.22 is often very useful; but when $X = M \in \mathcal{M}_{\mathrm{loc}}^c$, ucp convergence of the integrands can be drastically relaxed.

Definition 6.26. *Let $M \in \mathcal{M}_{\mathrm{loc}}^c$ and let $H^{(n)}, H \in L_{\mathrm{loc}}^2(M)$. We say that $H^{(n)} \to H$ in $L_{\mathrm{loc}}^2(M)$ if*

$$\lim_{n \to \infty} \mathbb{P}\left(\int_0^T \left| H_s^{(n)} - H_s \right|^2 \mathrm{d}\langle M \rangle_s > \varepsilon \right) = 0 \quad \forall \varepsilon > 0, T \in (0, +\infty).$$

Exercise. Show that, if $H^{(n)} \to H$ in ucp, then $H^{(n)} \to H$ in $L_{\mathrm{loc}}^2(M)$.

Lemma 6.27. *Let $M \in \mathcal{M}_{\mathrm{loc}}^c$. Then the following are equivalent:*

a) $H^{(n)} \to H$ *in* $L_{\mathrm{loc}}^2(M)$*;*

b) $\int_0^\cdot H_s^{(n)} \mathrm{d}M_s \to \int_0^\cdot H_s \mathrm{d}M_s$ *in ucp.*

Proof.

Proof skipped in the lectures, included here for completeness.

Set $\tilde{M}^{(n)} := \int_0^\cdot H_s^{(n)} \mathrm{d}M_s$, $\tilde{M} := \int_0^\cdot H_s \mathrm{d}M_s$. By Corollary 5.41, b) is equivalent to

$$\langle \tilde{M}^{(n)} - \tilde{M} \rangle = \left\langle \int_0^\cdot (H_s^{(n)} - H_s) \mathrm{d}M_s \right\rangle = \int_0^\cdot \left| H_s^{(n)} - H_s \right|^2 \mathrm{d}\langle M \rangle_s \to 0 \quad \text{in ucp}.$$

But by definition, $\int_0^\cdot \left| H_s^{(n)} - H_s \right|^2 \mathrm{d}\langle M \rangle_s \to 0$ in ucp coincides with $H^{(n)} \to H$ in $L_{\mathrm{loc}}^2(M)$. □

——— End of the lecture on January 9 ———

Definition 6.28. (Stratonovich integral) *If X, Y are continuous semimartingales, then we define the* Stratonovich integral *of Y w.r.t. X as*

$$\int_0^\cdot Y_s \circ \mathrm{d}X_s := \int_0^\cdot Y_s \mathrm{d}X_s + \frac{1}{2} \langle X, Y \rangle.$$

The motivation for considering this at first weird looking integral will become clear later. For now let us just observe a couple of properties coming from the definition; the first one may be interpreted as the fact that the Stratonovich integral arises from the limit of Riemann-type sums obtained by using the *trapezoidal rule*.

Proposition 6.29. *Let X, Y be continuous semimartingales. Let $\pi^n = \{t_k^n\}_{k \geqslant 0}$ be a sequence of deterministic, locally finite partitions with infinitesimal mesh. Then*

$$\left(\sum_{k=0}^{\infty} \frac{1}{2} (Y_{t_{k+1}^n \wedge t} + Y_{t_k^n \wedge t})(X_{t_{k+1}^n \wedge t} - X_{t_k^n \wedge t}) \right)_{t \geq 0} \longrightarrow \int_0^\cdot Y_s \circ \mathrm{d}X_s \qquad \textit{in ucp}.$$

Moreover the following integration by parts formula for Stratonovich integrals *holds:*

$$X_t Y_t = X_0 Y_0 + \int_0^t X_s \circ \mathrm{d}Y_s + \int_0^t Y_s \circ \mathrm{d}X_s. \qquad (6.18)$$

Proof. Exercise Sheet 11. □

Note that, differently from (6.17), formula (6.18) now does not contain any quadratic covariation term and resembles the usual integration by parts rule from classical calculus. In the same way in which standard integration by parts comes from integrating the product rule for derivatives $(fg)' = fg' + f'g$, here (6.18) may be formally written as

$$\text{“}\circ\mathrm{d}(XY) = X \circ \mathrm{d}Y + Y \circ \mathrm{d}X.\text{”}$$

7 Main theorems of stochastic analysis

We are now ready to present the main “rules of stochastic calculus”. In this section, we will constantly work with vector-valued processes, therefore we need to extend the definition of stochastic processes considered so far to this setting.

Definition 7.1. *We say that a $\mathbb{R}^d$-valued process $X = (X^1, \ldots, X^d)$ is a d-dimensional continuous semimartingale if each of its coordinates X^i is a real-valued continuous semimartingale. Similarly for d-dimensional martingales and continuous local martingales.*

Exercise. Show that X is a d-dimensional continuous semimartingale (resp. martingale, resp. continuous local martingale) if and only if, for all $\lambda \in \mathbb{R}^d$, $\lambda \cdot X = \sum_{i=1}^d \lambda^i X^i$ is a real-valued continuous semimartingale (resp. martingale, resp. continuous local martingale).

7.1 Itô's formula

Given a function $f \in C^2(\mathbb{R}^d; \mathbb{R})$, $f(x) = f(x_1, \ldots, x_d)$, we will use the following notations for its partial derivatives:

$$\partial_i f = \frac{\partial}{\partial x_i} f, \qquad \partial_{ij} f = \frac{\partial^2}{\partial x_i \partial x_j} f.$$

We denote the gradient, Hessian and Laplacian of f respectively by

$$\nabla f(x) = \begin{pmatrix} \partial_1 f(x) \\ \vdots \\ \partial_d f(x) \end{pmatrix}, \qquad D^2 f(x) = (\partial_{ij} f(x))_{i,j=1}^d \qquad \Delta f(x) = \sum_{i=1}^d \partial_{ii} f(x),$$

so that $\nabla f: \mathbb{R}^d \to \mathbb{R}^d$, $D^2 f: \mathbb{R}^d \to \mathbb{R}^{d \times d}$, $\Delta f: \mathbb{R}^d \to \mathbb{R}$. Recall that by Schwartz's theorem $\partial_{ij} f = \partial_{ji} f$, so that $D^2 f(x)$ is a symmetric matrix.

Theorem 7.2. (Itô's formula) *Let $X = (X^1, \ldots, X^d)$ be a d-dimensional continuous semimartingale and $f \in C^2(\mathbb{R}^d, \mathbb{R})$. Then $f(X)$ is a continuous semimartingale and (up to indistinguishability)*

$$f(X_t) = f(X_0) + \sum_{i=1}^d \int_0^t \partial_i f(X_s) \mathrm{d}X_s^i + \frac{1}{2} \sum_{i,j=1}^d \int_0^t \partial_{ij} f(X_s) \mathrm{d}\langle X^i, X^j \rangle_s \qquad \forall t \geqslant 0. \qquad (7.1)$$

Before delving into the proof, some comments are in order.

i. If a is a function of finite variation, then we have seen that for $f \in C^1(\mathbb{R}, \mathbb{R})$ it holds

$$f(a_t) - f(a_0) = \int_0^t f'(a_s)\, \mathrm{d}a_s;$$

this can be shown by considering telescopic sums. On the other hand, we have already seen in Corollary 6.25 that, at least for polynomial $f: \mathbb{R} \to \mathbb{R}$ and continous semimartingales X, it must hold

$$f(X_t) - f(X_0) = \int_0^t f'(X_s)\, \mathrm{d}X_s + \frac{1}{2}\int_0^t f''(X_s)\, \mathrm{d}\langle X\rangle_s;$$

formula (7.1) is a natural multidimensional generalization of this one.

The additional second order term $\frac{1}{2}\sum_{i,j}\int_0^t \partial_{ij} f(X_s) \mathrm{d}\langle X^i, X^j\rangle_s$ in (7.1), sometimes called "Itô corrector", marks the transition from standard analysis calculus (valid for finite variation functions) to stochastic calculus (which involves less regular processes, like martingales).

ii. Itô's formula shows that semimartingales remain semimartingales under composition with C^2 functions; moreover it provides an explicit formula for the decomposition of $f(X_t)$ into its bounded variation and local martingales components (find it as an Exercise). Recall that for $M \in \mathcal{M}_{\mathrm{loc}}^c$ and $f \in C^2$, we do not have $f(M) \in \mathcal{M}_{\mathrm{loc}}^c$ in general, i.e. local martingales are not closed under the composition with C^2 functions: just think of $X = B$ Brownian motion and $f(x) = x^2$.

iii. It is common and often convenient to write Itô's formula (7.1) in "differential" form as

$$\mathrm{d}f(X_t) = \sum_{i=1}^d \partial_i f(X_t) \mathrm{d}X_t^i + \frac{1}{2}\sum_{i,j=1}^d \partial_{ij} f(X_t) \mathrm{d}\langle X^i, X^j\rangle_t. \tag{7.2}$$

Introducing the notation $\langle X\rangle = (\langle X^i, X^j\rangle)_{i,j=1}^d$ (so that $\langle X\rangle$ is a $\mathbb{R}^{d\times d}$-valued process of finite variation) and the Frobenius product for matrices

$$A: B := \sum_{i,j=1}^d A_{ij} B_{ij},$$

eq. (7.2) can be written compactly as

$$\mathrm{d}f(X_t) = \nabla f(X_t) \cdot \mathrm{d}X_t + \frac{1}{2} D^2 f(X_t): \mathrm{d}\langle X\rangle_t.$$

In order to prove Itô formula, we need some preliminaries.

Fact. *Let $\{X^n\}_n$, X be continuous processes, then the following are equivalent:*

1. *$X^n \to X$ in ucp.*
2. *Every subsequence $\{X^{n_k}\}_k$ admits a subsubsequence $\{X^{n_{k_j}}\}_j$ such that for $\mathbb{P}$-a.e. ω, $X^{n_{k_j}}(\omega) \to X(\omega)$ uniformly on compacts sets as $j \to \infty$.*

Exercise. With the help of Lemma 5.16, prove the above Fact.

Lemma 7.3. *Let X, Y be continuous semimartingales and let H be a continuous adapted process. Let $\pi^n = \{t_k^n\}_{k\in\mathbb{N}}$ be a sequence of deterministic, locally finite partitions with infinitesimal mesh. Then*

$$\sum_k H_{t_k^n} X_{t\wedge t_k^n, t\wedge t_{k+1}^n} Y_{t\wedge t_k^n, t\wedge t_{k+1}^n} \to \int_0^t H_s \, \mathrm{d}\langle X, Y\rangle_s \quad \text{in ucp.}$$

Proof. By polarization, it suffices to show the statement for $X = Y$.

By Lemma 5.45 and Lemma 5.16, we can find a (not relabelled for simplicity) subsequence such that for $\mathbb{P}$-a.e. ω

$$\sum_k (X_{t\wedge t_k^n, t\wedge t_{k+1}^n}(\omega))^2 \to \langle X\rangle_t(\omega) \quad \text{uniformly on compact sets.} \tag{7.3}$$

From now on, we fix any ω on which the above convergence holds; the argument is completely pathwise. For fixed $t \geqslant 0$, we define a nonnegative (random) measure $\mu^{n,t}$ by

$$\mu^{n,t} = \sum_{k=0}^{n-1} \delta_{t_k^n} (X_{t\wedge t_k^n, t\wedge t_{k+1}^n})^2 = \sum_{k=0}^{n-1} \delta_{t_k^n} \mathbb{1}_{t_k^n \leqslant t} (X_{t\wedge t_k^n, t\wedge t_{k+1}^n})^2.$$

Here δ_x denotes a Dirac measure centered at x, namely such that $\int g(s)\delta_x(\mathrm{d}s) = g(x)$ for all continuous $g: \mathbb{R}_+ \to \mathbb{R}$. So $\mu^{n,t}$ is a linear combination of multiple of Diracs and so a nonnegative measure.

Notice that by its definition

$$Z_t^n := \sum_k H_{t_k^n} (X_{t\wedge t_k^n, t\wedge t_{k+1}^n})^2 = \int_0^{+\infty} H_s \, \mu^{n,t}(\mathrm{d}s).$$

If we show that $\mu^{n,t}$ converge as measures to a limit μ^t, then Z_t^n will converge as well (because H is continuous and actually bounded on $[0,t]$, which is the interval on which all these measures are supported).

By Stochastics I (cf. Portmanteu's theorem), in order to verify convergence in the sense of measures, it suffices to show that their cumulative distribution functions converge:

$$\mu^{n,t}([0,s]) \to \mu^t([0,s]) \quad \forall s \geqslant 0.$$

Notice that

$$\mu^{n,t}([0,s]) = \sum_{k=0}^{\infty} \mathbb{1}_{t_k^n \leqslant t} \mathbb{1}_{t_k^n \leqslant s} (X_{t\wedge t_k^n, t\wedge t_{k+1}^n})^2 = \sum_{k=0}^{\infty} \mathbb{1}_{t_k^n \leqslant s\wedge t} (X_{s\wedge t\wedge t_k^n, t\wedge t_{k+1}^n})^2;$$

in particular, for fixed s, by considering j s.t. $s \in [t_j^n, t_{j+1}^n)$, we have

$$\begin{aligned}\left|\mu^{n,t}([0,s]) - \sum_{k=0}^{\infty} (X_{s\wedge t\wedge t_k^n, s\wedge t\wedge t_{k+1}^n})^2\right| &= |(X_{t_j^n, t\wedge t_{j+1}^n})^2 - (X_{t_j^n, s})^2| \\ &\leqslant |X_{t_j^n, t\wedge t_{j+1}^n}|^2 + |X_{t_j^n, s}|^2 \\ &\leqslant 2 \sup_{\substack{0\leqslant u\leqslant v\leqslant T \\ |u-v|\leqslant |\pi^n|}} |X_{u,v}|^2.\end{aligned}$$

Since $|\pi^n| \to 0$ and X is continuous, the right hand side goes to 0; combined with (7.3), we deduce that

$$\lim_{n\to\infty} \mu^{n,t}([0,s]) = \lim_{n\to\infty} \sum_{k=0}^{\infty} (X_{s\wedge t\wedge t_k^n, s\wedge t\wedge t_{k+1}^n})^2 = \langle X\rangle_{s\wedge t} = \langle X\rangle_s^t \quad \forall s \geqslant 0$$

so that

$$Z_t^n = \int_0^{+\infty} H_s\, \mu^{n,t}(\mathrm{d}s) \to \int_0^{+\infty} H_s\, \mathrm{d}\langle X\rangle_s^t = \int_0^t H_s\, \mathrm{d}\langle X\rangle_s \quad \forall t \geqslant 0.$$

This shows pointwise convergence of $Z^n \to \int_0^\cdot H_s\, \mathrm{d}\langle X\rangle_s$. With a bit more effort, one can improve it to uniform convergence on compact sets, but we omit the details here.

All of the above was true on any fixed ω such that (7.3) holds, which overall shows $\mathbb{P}$-a.s. convergence, uniformly on compact sets, for the extracted subsequence.

We ran argument for the full sequence $\{Z^n\}_n$, but thanks to Lemmas 5.16 and 5.45, the same applies to any other subsequence $\{Z^{n_k}\}_k$ we could start from. By the Fact above, we conclude that $Z^n \to \int_0^\cdot H_s\, \mathrm{d}\langle X\rangle_s$ in ucp. □

Proof of Theorem 7.2. We use a pathwise argument, which is due to Föllmer.

Let π^n be any sequence of locally finite partitions of infinitesimal mesh; to fix ideas, we can take $t_k^n := tk/n$, but it's not really relevant.

Recall Taylor's formula up to second order: given $f \in C^2(\mathbb{R}^d; \mathbb{R})$, $x, h \in \mathbb{R}^d$, it holds

$$\begin{aligned} f(x+h) &= f(x) + \sum_{i=1}^d \partial_i f(x) h^i + \frac{1}{2} \sum_{i,j=1}^d \partial_{ij} f(x) h^i h^j \\ &\quad + \sum_{i,j=1}^d \left[\int_0^1 (1-\lambda)(\partial_{ij} f(x+\lambda h) - \partial_{ij} f(x))\mathrm{d}\lambda\right] h^i h^j. \end{aligned}$$

(Here we gave the explicit integral expression for the remainder, but other expressions would equally work in the proof below, as long as they allow to prove that the remainder R^n defined below goes to 0).

By writing $f(X_t) - f(X_0)$ as a telescopic sum and using Taylor's formula for f up to second order (with $x = X_{t\wedge t_k^n}$, $h = X_{t\wedge t_k^n, t\wedge t_{k+1}^n}$), we find

$$\begin{aligned} & f(X_t) - f(X_0) \\ &= \sum_{k=0}^\infty (f(X_{t\wedge t_{k+1}^n}) - f(X_{t\wedge t_k^n})) \\ &= \sum_{i=1}^d \sum_{k=0}^\infty \partial_i f(X_{t_k^n}) X^i_{t\wedge t_k^n, t\wedge t_{k+1}^n} + \frac{1}{2} \sum_{i,j=1}^d \sum_{k=0}^\infty \partial_{ij} f(X_{t_k^n}) X^i_{t\wedge t_k^n, t\wedge t_{k+1}^n} X^j_{t\wedge t_k^n, t\wedge t_{k+1}^n} \\ &\quad + \sum_{i,j=1}^d \sum_{k=0}^\infty \left[\int_0^1 (1-\lambda)(\partial_{ij} f(X_{t_k^n} + \lambda X_{t\wedge t_k^n, t\wedge t_{k+1}^n}) - \partial_{ij} f(X_{t_k^n}))\mathrm{d}\lambda\right] X^i_{t\wedge t_k^n, t\wedge t_{k+1}^n} X^j_{t\wedge t_k^n, t\wedge t_{k+1}^n} \\ &=: I_t^{1,n} + I_t^{2,n} + R_t^n \end{aligned}$$

where we used several times the fact that $g(X_{t\wedge t_k^n}) X^i_{t\wedge t_k^n, t\wedge t_{k+1}^n} = g(X_{t_k^n}) X^i_{t\wedge t_k^n, t\wedge t_{k+1}^n}$. By Corollary 6.23 and Lemma 7.3, we know that

$$I_\cdot^{1,n} \to \sum_{i=1}^d \int_0^\cdot \partial_i f(X_s) \mathrm{d}X_s^i, \quad I_\cdot^{2,n} \to \frac{1}{2} \sum_{i,j=1}^d \int_0^\cdot \partial_{ij} f(X_s) \mathrm{d}\langle X^i, X^j\rangle_s \quad \text{in ucp}$$

therefore in order to conclude it suffices to show that $R_\cdot^n \to 0$ in ucp as well. Using the Fact about ucp convergence, arguing as in Lemma 7.3, without loss of generality we can extract a (not relabelled) subsequence of the partitions $\{\pi^n\}_n$ and assume that $\mathbb{P}$-a.s

$$\sum_k X^i_{t\wedge t_k^n, t\wedge t_{k+1}^n} X^j_{t\wedge t_k^n, t\wedge t_{k+1}^n} \to \langle X^i, X^j\rangle_t \quad \text{uniformly on compact sets.}$$

The rest of the argument is completely pathwise, i.e. we fix $\omega\in\Omega$ such that the above holds. Let $[0,T]$ be a finite interval and let $K\subset\mathbb{R}^d$ be a convex compact set in which $X(\omega)|_{[0,T]}$ takes its values (for instance the closed ball of radius $R=\sup_{t\in[0,T]}|X_t(\omega)|$). Let

$$\varphi(h):=\sup\{|\partial_{ij}f(x)-\partial_{ij}f(x')|: x,x'\in K, |x-x'|\leqslant h\}\quad \text{for } h\geqslant 0,$$

namely φ is the modulus of continuity of $D^2 f$ restricted to the set K (since K is compact and $f\in C^2$, $D^2 f|_K$ is uniformly continuous). Then we can bound

$$\left|\int_0^1 (1-\lambda)(\partial_{ij}f(X_{t_k^n}+\lambda X_{t\wedge t_k^n,t\wedge t_{k+1}^n})-\partial_{ij}f(X_{t_k^n}))\mathrm{d}\lambda\right|\leqslant\varphi(|X_{t\wedge t_k^n,t\wedge t_{k+1}^n}|);$$

with the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}&\sum_{k=0}^{n-1}\varphi(|X_{t\wedge t_k^n,t\wedge t_{k+1}^n}|)\left|X^i_{t\wedge t_k^n,t\wedge t_{k+1}^n}X^j_{t\wedge t_k^n,t\wedge t_{k+1}^n}\right|\\ &\leqslant\varphi\Big(\max_{k\in\mathbb{N}}|X_{t\wedge t_k^n,t\wedge t_{k+1}^n}|\Big)\left(\sum_{k=0}^{\infty}(X^i_{t\wedge t_k^n,t\wedge t_{k+1}^n})^2\right)^{1/2}\left(\sum_{k=0}^{\infty}(X^j_{t\wedge t_k^n,t\wedge t_{k+1}^n})^2\right)^{1/2}\\ &\longrightarrow 0\cdot\langle X^i\rangle_t^{1/2}\langle X^j\rangle_t^{1/2}=0,\end{aligned}$$

since $\varphi(0)=0$, X is uniformly continuous on $[0,T]$ and the partitions have mesh $|\pi^n|\to 0$. In fact, the above convergence is uniform in $t\in[0,T]$; as the argument holds for any $T\in(0,+\infty)$, we conclude that $\mathbb{P}$-a.s. $R^n_\cdot\to 0$ uniformly on compact sets.

As the above argument works for any subsequence we can extract, by the Fact about ucp convergence we get the conclusion. □

Comment on the proof: the above proof not only shows the Itô formula, but also that it is well approximated by the corresponding Riemann-Stjeltes approximations uniformly on compact sets. However if one only wants to prove (7.1), the argument may be simplified as follows:

- We fix $t\geqslant 0$ at the beginning, and prove (7.1) at fixed t. Using the fact that both left and right hand sides of (7.1) are continuous processes, we can then derive equality up to indistinguishability by the usual arguments (take any $t\in\mathbb{Q}_+$ and continuity).
- For fixed t, since we have the freedom to pick any partition we want, we can assume that $t\in\pi^n$ for all n. Notationwise, it means we do not need to carry around $t\wedge t_k^n$ in the computations.
- Similarly, once we work with fixed $t\in\pi^n$, at the level of Lemma 7.3 we do not need ucp convergence, but only convergence in probability at fixed $t\geqslant 0$, which slightly simplifies the argument in the proof therein as well.

——— End of the lecture on January 15 ———

Exercise. Let B be a Brownian motion and $n\in\mathbb{N}$. What is the semimartingale decomposition of B^n?

Exercise. Let X, Y be continuous semimartingales. Recover the integration by parts formula by applying Itô formula to $f(x,y)=xy$.

Remark 7.4. Going through the same proof, one can see that if some coordinates $A^1, \ldots,$ A^e with $e \leqslant d$ of $X = (A, Y)$ are continuous and of finite variation, then the regularity conditions on f can be relaxed to requiring $f \in C^{1,2}(\mathbb{R}^e \times \mathbb{R}^{d-e})$. By this notation we mean that $\partial_i f$ are required to exist and be continuous for all $i = 1, \ldots, d$, but $\partial_{ij} f$ must exist and be continuous only when $i, j \geqslant e + 1$. In other words, only C^1 differentiability is required in the coordinates i where the process $X^i = A^i$ is of finite variation. In this case, the formula becomes

$$f(X_t) = f(X_0) + \sum_{i=1}^{d} \int_0^t \partial_i f(X_s) \mathrm{d}X_s^i + \frac{1}{2} \sum_{i,j=e+1}^{d} \int_0^t \partial_{ij} f(X_s) \mathrm{d}\langle X^i, X^j \rangle_s$$

which is consistent with the fact that $\langle X^i, X^j \rangle \equiv 0$ whenever $i \leqslant e$ or $j \leqslant e$.

Corollary 7.5. (Itô formula for Brownian motion) *Let B be a d-dimensional Brownian motion, $f \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^d)$, $f = f(t, x)$. Then (up to indistinguishability)*

$$f(t, B_t) = f(0, B_0) + \sum_{i=1}^{d} \int_0^t \partial_i f(s, B_s) \mathrm{d}B_s^i + \int_0^t \left(\partial_t f(s, B_s) + \frac{1}{2} \Delta f(s, B_s) \right) \mathrm{d}s. \tag{7.4}$$

The formula may be written a bit formally in differential form as

$$\mathrm{d}f(t, B_t) = \nabla f(t, B_t) \cdot \mathrm{d}B_t + \left[\partial_t f(t, B_t) + \frac{1}{2} \Delta f(t, B_t) \right] \mathrm{d}t.$$

Proof. Since $t \mapsto t$ is of finite variation and $f \in C^{1,2}(\mathbb{R}_+ \times \mathbb{R}^d)$, by the above remark Itô formula in this case becomes

$$\begin{aligned} f(t, X_t) = & f(0, X_0) + \int_0^t \partial_t f(s, X_s) \mathrm{d}s + \sum_{i=1}^{d} \int_0^t \partial_i f(s, X_s) \mathrm{d}X_s^i \\ & + \frac{1}{2} \sum_{i,j=1}^{d} \int_0^t \partial_{ij} f(s, X_s) \mathrm{d}\langle X^i, X^j \rangle_s. \end{aligned}$$

Since B is a d-dimensional Brownian motion $\langle B^i, B^j \rangle_t = \delta_{ij}\, t$ (equivalently $\langle B \rangle_t = I_d\, t$); inserting this fact in the above formula, using that $\Delta f = \sum_{i=1}^d \partial_{ii} f$ and rearranging the terms yields the conclusion. □

Lemma 7.6. *Let $X = (X^1, \ldots, X^d)$ be a d-dimensional continuous semimartingale and let $U \subset \mathbb{R}^d$ be open and such that, for $\mathbb{P}$-a.e. ω,*

$$X_t(\omega) \in U \quad \textit{for all } t \geqslant 0.$$

Then Itô's formula holds for $f \in C^2(U, \mathbb{R})$.

Sketch of proof.

Proof skipped in the lectures in the interst of time, here is a sketch for completeness.

We consider the stopping times $\tau_n = \inf \{t \geqslant 0 \colon d(X_t, U^c) \leqslant 1/n\}$ and first prove Itô's formula for X^{τ_n} before letting $n \to \infty$ at the end. To prove the formula for X^{τ_n} we have two options:

- Either we redo the proof that we gave above. Here we have to be a bit careful, because we need some convexity to control the error term in Taylor's formula (in the proof above we chose a compact convex set K that contains the image of X) and U is not necessarily convex. But in fact we just need $\lambda X_{t_k^m}^{\tau_n} + (1 - \lambda) X_{t_{k+1}^m}^{\tau_n}$ to be bounded away from the boundary of U, and we can achieve this by making the step size small enough.

- Alternatively, we find an approximation f_m to f such that $f_m \in C^2(\mathbb{R}^d, \mathbb{R})$ and f_m converges to f uniformly on compact subsets of U. For example, one can extend f by zero to U^c and then convolve with a smooth approximation of the identity, or multiply f by a smooth function which is identically 0 on U^c and 1 on

$$U_n := \left\{ x \in U : d(x, \partial U) \leqslant \frac{1}{n} \right\}.$$

Then we can apply Itô's formula to $f_m(X)$ and pass to the limit on both sides. □

The extension from the above lemma will be very useful in order to apply Itô's formula to processes such as $\frac{1}{X}$ or $\log X$, provided e.g. that X is a strictly positive semimartingale.

Recall the Stratonovich integral from Definition 6.28. Similarly to the case of the integration by parts formula (6.18), for regular enough f this notion of integral recovers the classical rules of calculus.

Corollary 7.7. (Chain rule for Stratonovich integral) *Let X be a d-dimensional continuous semimartingale and let $f \in C^3(\mathbb{R}^d; \mathbb{R})$. Then up to indistinguishability*

$$f(X_t) = f(X_0) + \sum_{i=1}^{d} \int_0^t \partial_i f(X_t) \circ \mathrm{d}X_t^i.$$

Proof. Exercise Sheet 11. □

7.2 First applications of Itô's formula

We say that a *complex-valued* continuous stochastic process is a (local, semi-) martingale if both its real and imaginary parts are (local, semi-) martingales. In other words, $X = X^1 + \iota X^2$ is a $\mathbb{C}$-valued (local, semi-) martingale if and only if (X^1, X^2) is a $\mathbb{R}^2$-valued (local, semi-) martingale. Here $\iota = \sqrt{-1}$ is the imaginary unit.

By "enforcing" bilinearity of $\langle \cdot, \cdot \rangle$, we may extend its definition to $\mathbb{C}$-valued continous semimartingales: if $M = M^1 + \iota M^2$, $N = N^1 + \iota N^2$, we set

$$\langle M, N \rangle := \langle M^1, N^1 \rangle - \langle M^2, N^2 \rangle + \iota \langle M^1, N^2 \rangle + \iota \langle M^2, N^1 \rangle$$

According to this definition, one can check that $\langle \cdot, \cdot \rangle$ is still bilinear and symmetric, namely

i. $\langle M, N \rangle = \langle N, M \rangle$;

ii. $\langle zM, N \rangle = z \langle M, N \rangle$ for all $z \in \mathbb{C}$.

Moreover, if M, N are $\mathbb{C}$-valued continuous local martingales, then so is $MN - M_0 N_0 - \langle M, N \rangle$.

Exercise. Prove all the above statements. Note that in Point ii. complex products appear: if $z^j = x^j + \iota y^j$ for $j = 1, 2$, then

$$z^1 z^2 = (x^1 x^2 - y^1 y^2) + \iota (x^1 y^2 + x^2 y^1).$$

Moreover, show that if M, N are $\mathbb{C}$-valued local martingales, then so is

$$MN - M_0 N_0 - \langle M, N \rangle.$$

When needed, we may always identity $\mathbb{C}$ with $\mathbb{R}^2$, by mapping $z = x + \iota y$ to (x, y). In this way, $f : \mathbb{C} \to \mathbb{C}$ corresponds to $\tilde{f} : \mathbb{R}^2 \to \mathbb{R}^2$; so $f = u + \iota v$ where $u, v : \mathbb{R}^2 \to \mathbb{R}$.

Small complex analysis recap: If $f:\mathbb{C}\to\mathbb{C}$ is *complex differentiable*, then u, v are smooth functions satisfying the Cauchy-Riemann equations

$$\partial_1 u = \partial_2 v, \quad \partial_1 v = -\partial_2 u.$$

In this case, we have $f'(z) = \partial_1 u(z) + \iota \partial_1 v(z)$, $f''(z) = \partial_{11} u(z) + \iota \partial_{11} v(z)$ and

$$\partial_{11} u = \partial_{12} v = -\partial_{22} u, \quad \partial_{22} v = \partial_{12} u = -\partial_{11} v.$$

Moreover, if $f:\mathbb{C}\to\mathbb{C}$ is complex differentiable (equivalently, f is *holomorphic*), then it is actually infinitely differentiable (in fact, u and v must be analytic).

Given the above definitions, one can still show that if X is a $\mathbb{C}$-valued semimartingale and $f:\mathbb{C}\to\mathbb{C}$ is complex differentiable, then we have the $\mathbb{C}$-valued Itô formula

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)\mathrm{d}X_s + \frac{1}{2}\int_0^t f''(X_s)\mathrm{d}\langle X\rangle_s. \tag{7.5}$$

In (7.5), the stochastic integral $\int_0^\cdot H_s\,\mathrm{d}X_s$ is defined similarly to $\langle M, N\rangle$, by enforcing bilinearity of the map $(H, X)\mapsto \int_0^\cdot H_s\,\mathrm{d}X_s$, splitting real and imaginary parts of H and X, and reducing to (four) real-valued stochastic integrals.

Exercise. Derive formula (7.5) from its $\mathbb{R}^2$-valued analogue and the above relations (as before, note that complex products appear in (7.5)).

We have seen before that for $\lambda\in\mathbb{R}$, the process $t\mapsto e^{\lambda B_t - \lambda^2 t/2}$ is a martingale. Using Itô's formula, we obtain the following generalization.

Recall: The complex exponential of $z = x + \iota y$ is defined as

$$e^z = e^x\, e^{\iota y} = e^x(\cos y + \iota \sin y).$$

$z\mapsto e^z$ is complex differentiable and $(e^z)' = e^z$.

Proposition 7.8. (Stochastic exponential) *Let $M\in\mathcal{M}_{\text{loc}}^c$ and $\lambda\in\mathbb{C}$. We set*

$$\mathcal{E}(\lambda M)_t := \exp\left(\lambda M_t - \frac{\lambda^2}{2}\langle M\rangle_t\right) \qquad \forall t\geqslant 0.$$

Then $\mathcal{E}(\lambda M)$ is a local martingale and it solves the stochastic differential equation

$$\mathcal{E}(\lambda M)_t = e^{\lambda M_0} + \lambda\int_0^t \mathcal{E}(\lambda M)_s \mathrm{d}M_s.$$

For $\lambda = 1$, we call $\mathcal{E}(M)$ the stochastic exponential (or Doléans-Dade exponential) *of M.*

Proof. Notice that, by relabelling $\tilde{M} = \lambda M$, we may assume without loss of generality that $\lambda = 1$ and M is a $\mathbb{C}$-valued continuous local martingale.

The process $X = M - \frac{1}{2}\langle M\rangle$ is a $\mathbb{C}$-valued semimartingale, $\langle X\rangle = \langle M\rangle$ and applying the complex Itô formula (7.5) to $f(z) = e^z$ we find

$$\begin{aligned}
e^{X_t} =& e^{X_0} + \int_0^t e^{X_s}\,\mathrm{d}X_s + \frac{1}{2}\int_0^t e^{X_s}\,\mathrm{d}\langle X\rangle_s \\
=& e^{M_0} + \int_0^t e^{X_s}\,\mathrm{d}M_s - \frac{1}{2}\int_0^t e^{X_s}\,\mathrm{d}\langle X\rangle_s + \frac{1}{2}\int_0^t e^{X_s}\,\mathrm{d}\langle X\rangle_s \\
=& e^{M_0} + \int_0^t e^{X_s}\,\mathrm{d}M_s
\end{aligned}$$

which is exactly the stochastic differential equation above. In particular, e^{X_t} is a local martingale since $\int_0^t e^{X_s}\mathrm{d}M_s$ is so, being a stochastic integral w.r.t. a cts local martingale. □

Remark 7.9. The above provides our first example of a solution to a *stochastic differential equation* (SDE): given $M \in \mathcal{M}_{\mathrm{loc}}^c$, we say that a continuous, adapted process N solves

$$\mathrm{d}N_t = N_t \,\mathrm{d}M_t \tag{7.6}$$

if the same relation holds in integral form, namely (up to indistinguishability)

$$N_t = N_0 + \int_0^t N_s \,\mathrm{d}M_s \quad \forall t \geqslant 0.$$

By Proposition 7.8, a solution to the SDE (7.6) is given by

$$N_t = \mathcal{E}(M)_t = e^{M_t - \frac{1}{2}\langle M \rangle_t}.$$

In turn, Proposition 7.8 gives a very useful and often easy to check characterization of the Brownian motion:

Theorem 7.10. (Lévy's characterization of Brownian motion) *Let $X = (X^1, \ldots, X^d)$ be a d-dimensional $\mathbb{F}$-adapted continuous process with $X_0 = 0$. Then X is a d-dimensional $\mathbb{F}$-Brownian motion if and only if the following two conditions hold:*

i. All components X^j are local martingales, and

ii. $\langle X^j, X^k \rangle_t = \delta_{j,k} t$ for all $t \geqslant 0$, equivalently $\langle X \rangle_t = t I_d$.

In particular, a real-valued continuous local martingale M with $M_0 = 0$ is a Brownian motion if and only if $\langle M \rangle_t = t$.

Proof. We already know that the conditions are necessary. Let us show that they are sufficient. Let $\xi \in \mathbb{R}^d$, then the process $\xi \cdot X := \sum_{j=1}^d \xi^j X^j$ is a continuous local martingale with

$$\langle \xi \cdot X, \xi \cdot X \rangle_t = \sum_{j,k=1}^d \xi^j \xi^k \langle X^j, X^k \rangle_t = \sum_{j,k=1}^d \xi^j \xi^k \delta_{j,k} t = |\xi|^2 t.$$

By Proposition 7.8 (with $\lambda = \iota$), the process

$$\exp\left(\iota \xi \cdot X_t - \frac{\iota^2 |\xi|^2}{2} t\right) = \exp\left(\iota \xi \cdot X_t + \frac{|\xi|^2}{2} t\right), \qquad t \geqslant 0,$$

is therefore a local martingale. But this process is also bounded on every compact interval (because $|e^{\iota y}| = 1$ for all $y \in \mathbb{R}$), and therefore it is a martingale. Hence, we get

$$\mathbb{E}\left[\exp\left(\iota \xi \cdot X_t + \frac{|\xi|^2}{2} t\right) \middle| \mathcal{F}_s\right] = \exp\left(\iota \xi \cdot X_s + \frac{|\xi|^2}{2} s\right) \quad \forall s < t.$$

But then for every $\mathcal{F}_s$-measurable random variable Y, the characteristic function of $(Y, X_{s,t}^1, \ldots, X_{s,t}^d)$ is given by

$$\begin{aligned}
\mathbb{E}[\exp(\iota u Y + \iota \xi \cdot (X_{s,t}))] &= \mathbb{E}[\exp(\iota u Y) \mathbb{E}[\exp(\iota \xi \cdot (X_{s,t})) | \mathcal{F}_s]] \\
&= \mathbb{E}\left[\exp(\iota u Y) \exp\left(-\frac{|\xi|^2}{2}(t-s)\right)\right] \\
&= \mathbb{E}[\exp(\iota u Y)] \exp\left(-\frac{|\xi|^2}{2}(t-s)\right) \\
&= \mathbb{E}[\exp(\iota u Y)] \mathbb{E}[\exp(\iota \xi \cdot (X_{s,t}))],
\end{aligned}$$

which proves that $X_{s,t}$ is independent of Y and thus of $\mathcal{F}_s$ (since Y was an arbitrary $\mathcal{F}_s$-measurable random variable). Moreover, taking $Y=0$ we get that $X_{s,t}\sim\mathcal{N}(0,(t-s)\mathbb{I})$, where $\mathbb{I}$ is the identity matrix in $\mathbb{R}^{d\times d}$. Combining these facts implies that X must be a (d-dimensional) $\mathbb{F}$-Brownian motion. □

—— End of the lecture on January 16 ——

Exercise. Use Levy's characterization theorem to give another proof of the reflection principle (Proposition 3.24): if B is a one-dimensional $\mathbb{F}$-BM and τ is a $\mathbb{F}$-stopping time, then $B_t^*:=B_t\mathbb{1}_{\{t\leqslant\tau\}}+(2B_\tau-B_t)\mathbb{1}_{\{t>\tau\}}$ is a $\mathbb{F}$-BM (you might want to find H such that $B_\cdot^*=\int_0^\cdot H_s\,\mathrm{d}B_s$).

Remark 7.11. The result is false if M is not assumed to be continuous and allowed to have jumps. Indeed, recall Example 4.4-iii): we can construct a suitable compound Poisson process X_t (by taking $m=0$, $\lambda=a=1$) such that X and $|X_t|^2-t$ are both martingales.

Technically, we didn't define what is the quadratic variation $\langle M\rangle$ for local martingales with jumps, but the above would suggest $\langle X\rangle=t$, yielding a counterexample. In fact, in the presence of jumps, things are more subtle and there exist two distinct definitions of $\langle M\rangle$ in the literature, which are both useful for different reasons.

Some more (not examinable) details on the above: Given a càdlàg martingale M, whose paths possibly exhibit jumps, one must distinguish between the *optional quadratic variation* $[M]$ and the *predictable quadratic variation* $\langle M\rangle$. Both are increasing processes such that $M_t^2-M_0^2-A$ is a martingale; $[M]$ is the process obtained by looking at the sum of squares of increments along partitions of infinitesimal mesh, which exhibits jumps whenever M does, while $\langle M\rangle$ is uniquely characterized by the property of being *predictable*, in relation to the *Doob-Meyer decomposition theorem*. Under mild assumptions, one can show that $\langle M\rangle$ is continuous.

An example to clarify the above: by Example 4.4 (for $m=a=1$, $Y_k\equiv1$), $\tilde{N}_t=N_t-\lambda t$ is a martingale; noticing that N is increasing and only moves upwards by istantaneous jumps of size 1, it's not difficult to show that

$$\lim_{n\to\infty}\sum_k(\tilde{N}_{t_k^n\wedge t,t_{k+1}^n\wedge t})^2=\lim_{n\to\infty}\sum_k(N_{t_k^n\wedge t,t_{k+1}^n\wedge t})^2=N_t$$

so that $[\tilde{N}]_t=N_t$. On the other hand, by Example 4.4 we know that $(N_t-\lambda t)^2-\lambda t$ is a martingale, which in fact amounts to $\langle\tilde{N}\rangle_t=\lambda t$. There is no contradiction in this, as it only implies that $[\tilde{N}]_t-\langle N\rangle_t=N_t-\lambda t=\tilde{N}_t$ is again a martingale, as well as a process of bounded variation.

Lévy's theorem admits a natural generalization to the complex-valued case. We say that a $\mathbb{C}$-valued process $B=B^1+\iota B^2$ is a *complex Brownian motion* if (B^1,B^2) is a $\mathbb{R}^2$-valued Brownian motion; similarly for the definition of a complex $\mathbb{F}$-Brownian motion. In the following, given $X=X^1+\iota X^2$, $\bar{X}$ denotes its complex conjugate, namely $\bar{X}=X^1-\iota X^2$.

Corollary 7.12. *Let X be a $\mathbb{C}$-valued, $\mathbb{F}$-adapted continuous process with $X_0=0$. Then X is a complex $\mathbb{F}$-Brownian motion if and only if the following two conditions hold:*

i. X is a $\mathbb{C}$-valued local martingale, and

ii. $\langle X,X\rangle_t=0$ and $\langle X,\bar{X}\rangle_t=2t$ for all $t\geqslant0$.

Proof. Exercise Sheet 12. □

Exercise. Arguing as in the proof of Theorem 7.10, show the following: let $M \in \mathcal{M}_{\mathrm{loc}}^c$ with $M_0 = 0$ be such that $M_0 = 0$ and such that $\langle M \rangle_t = f(t)$, where $f: \mathbb{R}_+ \to \mathbb{R}_+$ is given a (deterministic) continuous increasing function with $f(0) = 0$. Show that M has the same distribution as $B_{f(t)}$, where B is a Brownian motion.

By the result of a previous Exercise Sheet, we already know the converse result: $\tilde{M}_t := B_{f(t)}$ defines a martingale (w.r.t. a suitable filtration) and that $\langle \tilde{M} \rangle_t = f(t)$.

Thanks to Lévy's characterization theorem, we can prove a far reaching generalization: every continuous local martingale started at 0 can be transformed into a Brownian motion through a *random* time-change. To do this, we will use the following key lemma, which says that the quadratic variation of a local martingale acts as a "clock" that runs exactly when the local martingale is moving.

Lemma 7.13. *Let $M \in \mathcal{M}_{\mathrm{loc}}^c$. Then M and $\langle M \rangle$ almost surely have the same intervals of being constant. That is, for almost all $\omega \in \Omega$, we have that for all $0 \leqslant s < t$:*

$$M_r(\omega) = M_s(\omega) \quad \text{for all } r \in [s, t] \qquad \Leftrightarrow \qquad \langle M \rangle_t(\omega) = \langle M \rangle_s(\omega).$$

Proof.

Proof skipped in the lectures and not examinable, included here for completeness.

Define for $t \geqslant 0$ the times

$$\tau_t := \inf \{s \geqslant t : M_s - M_t \neq 0\}, \qquad \sigma_t := \inf \{s \geqslant t : \langle M \rangle_s - \langle M \rangle_t \neq 0\}.$$

Then τ_t and σ_t are stopping times because they are entry times of continuous processes into open sets and because our filtration is right-continuous. By the density of $\mathbb{Q}_+$ in $\mathbb{R}_+$ it suffices to show that almost surely $\sigma_q = \tau_q$ for all $q \in \mathbb{Q}_+$, and as usual this follows if we can show it for fixed $q \in \mathbb{Q}_+$.

So let $q \in \mathbb{Q}_+$. Then τ_q and σ_q are stopping times with $\tau_q, \sigma_q \geqslant q$, and for any stopping time $\rho \geqslant q$ we get

$$\langle M^\rho - M^q \rangle = \langle M \rangle^\rho + \langle M \rangle^q - 2\langle M^\rho, M^q \rangle = \langle M \rangle^\rho - \langle M \rangle^q. \tag{7.7}$$

With $\rho = \tau_q$ we get

$$M^{\tau_q} - M^q \equiv 0 \qquad \Rightarrow \qquad 0 \equiv \langle M^{\tau_q} - M^q \rangle = \langle M \rangle^{\tau_q} - \langle M \rangle^q,$$

and therefore $\sigma_q \geqslant \tau_q$. For $\rho = \sigma_q$ we read (7.7) from right to left and obtain

$$0 \equiv \langle M \rangle^{\sigma_q} - \langle M \rangle^q = \langle M^{\sigma_q} - M^q \rangle \qquad \Rightarrow \qquad M^{\sigma_q} - M^q \equiv 0,$$

and therefore $\tau_q \geqslant \sigma_q$. This concludes the proof. □

Theorem 7.14. (Dambis, Dubins-Schwarz[7.1]) *Let $M \in \mathcal{M}_{\mathrm{loc}}^c$ with $M_0 = 0$ and such that almost surely $\langle M \rangle_\infty = \infty$. Then, there exists a Brownian motion β such that almost surely*

$$M_t = \beta_{\langle M \rangle_t}, \qquad t \geqslant 0.$$

In other words, M is a time-changed Brownian motion.

Remark 7.15.

i. The assumption $\langle M \rangle_\infty = \infty$ is not necessary, but without it we might have to enlarge the probability space (think of $|\Omega| = 1$, $\mathcal{F} = \{\emptyset, \Omega\}$, $M = 0$). See Theorem V.1.7 of [23] for a proof.

7.1. **This result was proved in 1965 in two papers written independently, one by Dubins and Schwarz, and another one by Dambis.**

ii. We will see in the proof that the Brownian motion β is in general not adapted to our original filtration, but rather to a "time-changed filtration".

Proof. By Lemma 7.13, by changing M and $\langle M\rangle$ on a null set, we may suppose that $\langle M\rangle_\infty(\omega)=\infty$ for all $\omega\in\Omega$ and that $M(\omega)$ and $\langle M\rangle(\omega)$ have the same intervals of being constant for all $\omega\in\Omega$. Define the finite stopping times

$$\tau_t := \inf\{s\geqslant 0: \langle M\rangle_s > t\} \qquad \text{for } t\geqslant 0.$$

Then for all $t\geqslant 0$ the random variable $\beta_t := M_{\tau_t}$ is $\mathcal{F}_{\tau_t}$–measurable by Lemma 3.16. In other words, the process $(\beta_t)_{t\geqslant 0}$ is adapted to $(\mathcal{G}_t)_{t\geqslant 0}$ for $\mathcal{G}_t := \mathcal{F}_{\tau_t}$. We will apply Lévy's characterization to show that β is a Brownian motion.

First we show that β is continuous with $\beta_0=0$. We claim that the function

$$\mathbb{R}_+ \ni t \mapsto \tau_t \in \mathbb{R}_+$$

is increasing and right-continuous (so in particular càdlàg): indeed, for any $s>\tau_t$ we have $\langle M\rangle_s > t$, so

$$\tau_t \leqslant \tau_{t+} = \lim_{t_n\downarrow t} \tau_{t_n} < s.$$

Since $s>\tau_t$ was arbitrary, we get $\tau_t=\tau_{t+}$. Since $t\mapsto\tau_t$ is càdlàg and M is continuous, β is also càdlàg. But we can do better: if τ is continuous in t, then also β is continuous in t. If τ is discontinuous in t, then there exist $s<r$ with $\tau_{t-}=s<r=\tau_t$, and therefore $\langle M\rangle_r = \langle M\rangle_s$ (the process $\langle M\rangle$ has a constant stretch). But then also $M_s=M_r$, and therefore $\beta_{t-}=M_s=M_r=\beta_t$. So β is continuous. The same argument shows that $\beta_0(\omega)=0$ for all ω.

Next, we show that β and β_t^2-t, $t\geqslant 0$, are $(\mathcal{G}_t)_{t\geqslant 0}$-(local) martingales, so that β is a $(\mathcal{G}_t)_{t\geqslant 0}$-Brownian motion by Lévy's characterization. Note that for $n\in\mathbb{N}$ the processes M^{τ_n} and $(M^{\tau_n})^2-\langle M^{\tau_n}\rangle$ are uniformly integrable martingales because

$$\langle M^{\tau_n}\rangle_\infty = \langle M\rangle_{\tau_n} = n$$

is integrable. So by the stopping theorem we get for $s<t\leqslant n$

$$\mathbb{E}[\beta_t|\mathcal{G}_s] = \mathbb{E}[M^{\tau_n}_{\tau_t}|\mathcal{F}_{\tau_s}] = M^{\tau_n}_{\tau_s} = \beta_s$$

and

$$\mathbb{E}\,[\beta_t^2 - t|\mathcal{G}_s] = \mathbb{E}\,[(M^{\tau_n})^2_{\tau_t} - \langle M^{\tau_n}\rangle_{\tau_t}|\mathcal{F}_{\tau_s}] = (M^{\tau_n})^2_{\tau_s} - \langle M^{\tau_n}\rangle_{\tau_s} = \beta_s^2 - s.$$

Therefore, β is a Brownian motion.

To conclude the proof it suffices to show that $\beta_{\langle M\rangle_t} = M_t$ for all $t\geqslant 0$. If $\tau_{\langle M\rangle_t}=t$, then this is obvious. It may happen that $\tau_{\langle M\rangle_t}>t$ because $\langle M\rangle$ may have constant stretches, but as above we then use that M and $\langle M\rangle$ are constant on the same intervals to obtain the result also in that case. □

Exercise. Let $M\in\mathcal{M}^c_{\mathrm{loc}}$. Show that almost surely the set

$$\{t\geqslant 0: M \text{ is differentiable in } t \text{ and } M|_{(t-\varepsilon,t+\varepsilon)} \text{ is not constant for any } \varepsilon>0\}$$

is empty.

Example. Here is a simple application of the Dambis, Dubins-Schwarz theorem: let B be a Brownian motion and let $\lambda\in\mathbb{R}\setminus\{0\}$. Then $\lambda B\in\mathcal{M}^c_{\mathrm{loc}}$ and $\langle\lambda B\rangle_t=\lambda^2 t$. Therefore, there exists a Brownian motion β such that $B_t=\beta_{\lambda^2 t}$. In other words, λB has the same distribution as $(B_{\lambda^2 t})_{t\geqslant 0}$ – we have recovered the scaling invariance of Brownian motion!

Similarly to Lévy's theorem, we have a version of Theorem 7.14 for the complex-valued case, however under more stringent conditions.

Corollary 7.16. *Let M be a $\mathbb{C}$-valued continuous local martingale with $M_0=0$; assume that almost surely $\langle M,M\rangle=0$ and $\langle M,\bar{M}\rangle_\infty=\infty$. Then, there exists a $\mathbb{C}$-valued Brownian motion β such that almost surely*

$$M_t=\beta_{\frac{\langle M,\bar{M}\rangle_t}{2}}\qquad\forall t\geqslant 0.$$

Sketch of proof. Note that, if $M=M^1+\iota M^2$, then by linearity

$$\langle M,\bar{M}\rangle=\langle M^1,M^1\rangle+\langle M^2,M^2\rangle$$

is a real-valued increasing process. We can now go through the same proof as above, for stopping times $\tilde{\tau}_t$ defined by

$$\tilde{\tau}_t:=\inf\{s\geqslant 0: \langle M,\bar{M}\rangle_s>2t\}\qquad\text{for } t\geqslant 0$$

and go through similar passages as therein (roughly speaking using that "quadratic covariation behaves well under time change") to find that

$$\langle\beta,\beta\rangle_t=\langle M,M\rangle_{\tilde{\tau}_t}=0,\quad\langle\beta,\bar{\beta}\rangle=\langle M,\bar{M}\rangle_{\tilde{\tau}_t}=2t$$

so that β is a $\mathbb{C}$-valued Brownian motion. The rest of the proof proceeds identically. □

Among nice applications of Itô's formula, let us mention its consequence in terms of recurrence and transience properties of Brownian motion.

Proposition 7.17. (Recurrence/transience of BM in $\mathbb{R}^d$) *Let B be a d-dimensional Brownian motion. Then:*

a) *For $d=1$, BM is* point-recurrent*: it holds*

$$\mathbb{P}(\textit{for every } x, \textit{ there exists a sequence } t_n\to\infty \textit{ such that } B_{t_n}=x \textit{ for every } n)=1.$$

b) *For $d=2$, BM $\mathbb{P}$-a.s. does not pass through any given point: for any $x\in\mathbb{R}^2\setminus\{0\}$,*

$$\mathbb{P}(\textit{there exists some } t\geqslant 0 \textit{ such that } B_t=x)=0.$$

c) *For $d\geqslant 3$, BM is* transient*: $\mathbb{P}-a.s.$*

$$\lim_{t\to\infty}|B_t|=+\infty.$$

Proof. Part a) immediately follows from the already seen fact that $\mathbb{P}$-a.s.

$$\liminf_{t\to\infty}B_t=-\infty,\quad\limsup_{t\to\infty}B_t=+\infty.$$

Parts b) and c) are part of Exercise Sheet 12. The proof of b) is based on the *conformal invariance* of planar Brownian motion; c) is based on the fact that in $d\geqslant 3$, $x\mapsto|x|^{2-d}$ is *harmonic* on $\mathbb{R}^d\setminus\{0\}$ (in fact, up to multiplicative constant, it is the *Green function*). □

Remark 7.18. The proof of part c) provides an important example of an L^2-bounded continuous local martingale which is not a martingale: for $d=3$, given $x\in\mathbb{R}^3\setminus\{0\}$, one can take $M_t:=\frac{1}{|x-B_t|}$.

More details about recurrence: Dimension $d=2$ is somewhat "critical" for Brownian motion: even though in this case B $\mathbb{P}$-a.s. does not hit a given point, it can be shown that it is *neighbourhood recurrent*:

$$\mathbb{P}(\text{for all } x\in\mathbb{R}^2 \text{ and } \varepsilon>0, \text{there exists a sequence } t_n\to\infty \text{ such that } B_{t_n}\in B(x,\varepsilon))=1;$$

see for instance Theorem 7.17 from [16].
It is also interesting to compare Proposition 7.17 to the corresponding results for the d-dimensional simple random walk $(X_n)_{n\in\mathbb{N}}$ in $\mathbb{Z}^d$ (see Examples 1.6.2-1.6.3 from [19]): in this case X is recurrent for $d\leqslant 2$ and transient for $d\geqslant 3$.

——— End of the lecture on January 22 ———

7.3 Girsanov's theorem

So far, we have mostly considered a fixed reference filtered probability space $(\Omega,\mathcal{F},\mathbb{F},\mathbb{P})$ and developed all our calculus tools with respect to it. We might wonder what happens if we change our reference probability $\mathbb{P}$, by introducing another probability $\mathbb{Q}$: how will the (semi)martingale property of a process X, and the notion of stochastic integration w.r.t. X, be modified accordingly?

We can see that something nontrivial can happen already by considering finite dimensional Gaussian measures: let $Z\sim\mathcal{N}(0,I_d)$ and let $h\in\mathbb{R}^d$. Then, for any bounded f: $\mathbb{R}^d\to\mathbb{R}$, we have

$$\begin{aligned}
\mathbb{E}[f(h+Z)] &= (2\pi)^{-d/2}\int_{\mathbb{R}^d} f(h+x)e^{-\frac{|x|^2}{2}}\,\mathrm{d}x\\
&= (2\pi)^{-d/2}\int_{\mathbb{R}^d} f(y)e^{-\frac{|y-h|^2}{2}}\,\mathrm{d}y\\
&= (2\pi)^{-d/2}\int_{\mathbb{R}^d} f(y)e^{y\cdot h-\frac{|h|^2}{2}}e^{-\frac{|y^2}{2}}\,\mathrm{d}y\\
&= \mathbb{E}\left[f(Z)e^{Z\cdot h-\frac{|h|^2}{2}}\right].
\end{aligned}$$

So in this particular case, *shifting* the mean of our Gaussian Z by a factor h is equivalent (in expectation) to changing the underlying measure by multiplying by another random variable, namely considering a new probability measure $\mathbb{Q}$ via the Radon-Nikodym density

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}=e^{Z\cdot h-\frac{|h|^2}{2}}, \quad \text{so that} \quad \mathbb{E}_{\mathbb{Q}}[f(Z)]=\mathbb{E}_{\mathbb{P}}\left[f(Z)e^{Z\cdot h-\frac{|h|^2}{2}}\right].$$

In general, if Z has a possibly degenerate covariance matrix Σ, the above might work only when shifting Z under certain directions.

Exercise. Let $Z\sim\mathcal{N}(0,\Sigma)$, where $\Sigma\in\mathbb{R}^{d\times d}$. Show that, for any $h\in\operatorname{Im}(\Sigma^{1/2})$, namely such that $h=\Sigma^{1/2}\tilde{h}$ for some $\tilde{h}\in\mathbb{R}^d$, it holds

$$\mathbb{E}[f(Z+h)]=\mathbb{E}\left[f(Z)\,e^{Z\cdot(\Sigma^{-1}h)-\frac{|\Sigma^{-1/2}h|^2}{2}}\right].$$

If we think of Brownian motion B as an infinite dimensional version of a Gaussian variable Z, we might wonder if a similar result still holds in this case. And since we have now understood that continuous martingales are time-changed versions of Brownian motions, we can ask what happens in that case as well.

To address these questions, we work in a more abstract setting:

Definition 7.19. (absolutely continuous, equivalent, mutually singular) *Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on a measurable space $(\Omega, \mathcal{F})$. We say that:*

i. $\mathbb{Q}$ *is* absolutely continuous with respect to $\mathbb{P}$, $\mathbb{Q} \ll \mathbb{P}$, *if for all $A \in \mathcal{F}$ with $\mathbb{P}(A) = 0$ we also have $\mathbb{Q}(A) = 0$.*

ii. $\mathbb{Q}$ *is* equivalent to $\mathbb{P}$, $\mathbb{Q} \sim \mathbb{P}$, *if for all $A \in \mathcal{F}$ we have $\mathbb{P}(A) = 0$ if and only if $\mathbb{Q}(A) = 0$; namely, $\mathbb{Q} \sim \mathbb{P}$ if $\mathbb{Q} \ll \mathbb{P}$ and $\mathbb{P} \ll \mathbb{Q}$.*

iii. $\mathbb{Q}$ *and* $\mathbb{P}$ *are* mutually singular *if there exists $A \in \mathcal{F}$ with $\mathbb{P}(A) = 0$ and $\mathbb{Q}(A) = 1$ (note that then $\mathbb{P}(A^c) = 1$ and $\mathbb{Q}(A^c) = 0$).*

In finite dimensional spaces, there are many equivalent/absolutely continuous probability measures. For example, any probability measure on $\mathcal{B}(\mathbb{R}^d)$ that has a density with respect to Lebesgue measure is absolutely continuous with respect to the distribution of a Gaussian variable $Z \sim \mathcal{N}(0, I_d)$. In particular, all Gaussians $\mathcal{N}(\mu, \Sigma)$ with invertible covariance matrix Σ are equivalent. But in infinite-dimensional spaces there exists no Lebesgue measure (→Wikipedia), and "most" probability measures tend to be mutually singular.

Loosely speaking we will see that, if B is a Brownian motion under $\mathbb{P}$, then for any other measure $\mathbb{Q} \sim \mathbb{P}$, it holds $B = \tilde{B} + A$ where $\tilde{B}$ is a Brownian motion under $\mathbb{Q}$ and A is an absolutely continuous process, acting as a shift on B. Such transformations are widely applied in finance, e.g. in the Black-Scholes model of option pricing, as they allow to work under a so called *risk-neutral measure*.

At the same time, this means that the class of probabilities $\mathbb{Q} \sim \mathbb{P}$ is quite rigid; for example, we cannot transform the law of B into that of $2B$ by an equivalent change of measure.

Exercise. Let B be a Brownian motion and let $\lambda \in \mathbb{R} \setminus \{-1, 1\}$. Show that $\mathbb{P}_B$ and $\mathbb{P}_{\lambda B}$ are mutually singular (both as measures on $C(\mathbb{R}_+, \mathbb{R})$).

Hint: Think of the quadratic variation.

As usual, we work on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, $\mathbb{F}$ satisfying the usual assumption. Suppose we are given $\mathbb{Q} \ll \mathbb{P}$ (on $(\Omega, \mathcal{F})$).

Radon-Nikodym theorem: Let μ, ν be σ-finite measures on a measurable space $(S, \mathcal{S})$ such that $\nu \ll \mu$, i.e. for all $A \in \mathcal{S}$ with $\mu(A) = 0$ we have $\nu(A) = 0$. Then there exists a measurable, non-negative function f such that

$$\nu(A) = \int_S \mathbb{1}_A(x) \nu(\mathrm{d}x) = \int_S \mathbb{1}_A(x) f(x) \mu(\mathrm{d}x) \quad \forall A \in \mathcal{S}.$$

The function f is μ-almost surely unique and we write $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} := f$; f is often referred to as the *Radon-Nykodym derivative* of ν with respect to μ.

Since $\mathbb{Q} \ll \mathbb{P}$, by the Radon-Nikodym theorem there exists an $\mathcal{F}$–measurable random variable $Z \in L^1(\mathbb{P})$ (the *Radon-Nikodym derivative*) with $Z \geqslant 0$ and $\mathbb{E}_{\mathbb{P}}[Z] = 1$, such that

$$\mathbb{Q}(A) = \mathbb{E}_{\mathbb{P}}[\mathbb{1}_A Z] \quad \forall A \in \mathcal{F}.$$

We write

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} := Z.$$

If $\mathcal{G} \subset \mathcal{F}$ is a sub σ-algebra, then clearly we also have $\mathbb{Q}|_{\mathcal{G}} \ll \mathbb{P}|_{\mathcal{G}}$, so there exists a $\mathcal{G}$–measurable Radon-Nikodym derivative $\frac{\mathrm{d}\mathbb{Q}|_{\mathcal{G}}}{\mathrm{d}\mathbb{P}|_{\mathcal{G}}}$. We write

$$\left.\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\right|_{\mathcal{G}} := \frac{\mathrm{d}\mathbb{Q}|_{\mathcal{G}}}{\mathrm{d}\mathbb{P}|_{\mathcal{G}}}.$$

Given the filtration $\mathbb{F}$, we then set

$$Z_t := \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\bigg|_{\mathcal{F}_t} \qquad \forall t \in [0, \infty].$$

Lemma 7.20. *$Z = (Z_t)_{t \geqslant 0}$ is a uniformly integrable $\mathbb{P}$-martingale and $Z_t = \mathbb{E}[Z_\infty | \mathcal{F}_t]$.*

Proof. Z is adapted and integrable by construction. For any $t \geqslant 0$ and $A \in \mathcal{F}_t$, by construction we have

$$\mathbb{E}_{\mathbb{P}}[\mathbb{1}_A Z_\infty] = \mathbb{Q}(A) = \mathbb{Q}|_{\mathcal{F}_t}(A) = \mathbb{E}_{\mathbb{P}|_{\mathcal{F}_t}}[\mathbb{1}_A Z_t] = \mathbb{E}_{\mathbb{P}}[\mathbb{1}_A Z_t]$$

which proves that $Z_t = \mathbb{E}[Z_\infty | \mathcal{F}_t]$; uniform integrability follows since $Z_\infty \in L^1$. □

Since $\mathbb{F}$ satisfies the usual conditions, Z has a càdlàg modification; from now on, we work with this modification.

Lemma 7.21. *Let $\mathbb{Q} \ll \mathbb{P}$ and let Z be the càdlàg modification of $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\big|_{\mathcal{F}_t}$, $t \geqslant 0$.*

i. If τ is a stopping time, then $\mathbb{P}$-a.s.

$$Z_\tau = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\bigg|_{\mathcal{F}_\tau}.$$

ii. We have

$$\mathbb{Q}\Big(\inf_{t \geqslant 0} Z_t > 0\Big) = 1. \tag{7.8}$$

Proof.

i. By the stopping theorem, since Z is a uniformly integrable càdlàg martingale, for any $A \in \mathcal{F}_\tau$ we have

$$\mathbb{Q}(A) = \mathbb{E}_{\mathbb{P}}[\mathbb{1}_A Z_\infty] = \mathbb{E}_{\mathbb{P}}[\mathbb{1}_A Z_\tau].$$

ii. Let $\tau^\varepsilon = \inf\{t \geqslant 0 \colon Z_t < \varepsilon\}$; since Z is càdlàg and $\mathbb{F}$ is right-cts, it is a stopping time. By right continuity, it holds $Z_{\tau^\varepsilon} \leqslant \varepsilon$, and so by Point *i.* we find

$$\mathbb{Q}(\tau^\varepsilon < \infty) = \mathbb{E}_{\mathbb{P}}[\mathbb{1}_{\{\tau^\varepsilon < \infty\}} Z_{\tau^\varepsilon}] \leqslant \varepsilon,$$

where we used the fact that that $\{\tau^\varepsilon < \infty\} \in \mathcal{F}_{\tau^\varepsilon}$. But then

$$\mathbb{Q}\Big(\inf_{t \geqslant 0} Z_t > 0\Big) = \lim_{\varepsilon \to 0} \mathbb{Q}\Big(\inf_{t \geqslant 0} Z_t \geqslant \varepsilon\Big) = \lim_{\varepsilon \to 0} \mathbb{Q}(\tau^\varepsilon = +\infty) \geqslant \lim_{\varepsilon \to 0} (1 - \varepsilon) = 1$$

yielding the conclusion. □

Under the assumption that $\mathbb{Q} \sim \mathbb{P}$, from (7.8) we also deduce that for $\mathbb{P}$-a.e. ω

$$\inf_{t \geqslant 0} Z_t(\omega) > 0$$

Proposition 7.22. (Bayes' formula) *Let $\mathbb{Q} \sim \mathbb{P}$ and let $Z_t = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\big|_{\mathcal{F}_t}$, $t \geqslant 0$. Then for any $\mathcal{F}_t$–measurable random variable X, it holds $X \in L^1(\mathbb{Q})$ if and only if $X Z_t \in L^1(\mathbb{P})$. Moreover in this case $\mathbb{Q}$-a.s. (equivalently $\mathbb{P}$-a.s.) we have*

$$\mathbb{E}_{\mathbb{Q}}[X | \mathcal{F}_s] = \frac{\mathbb{E}_{\mathbb{P}}[X Z_t | \mathcal{F}_s]}{Z_s}. \tag{7.9}$$

Proof. Given X as above, since Z is non-negative, we have

$$\mathbb{E}_{\mathbb{P}}[|X Z_t|] = \mathbb{E}_{\mathbb{P}}[|X|\, Z_t] = \mathbb{E}_{\mathbb{Q}}[|X|]$$

so that one term is finite if and only if the other is, proving the first claim.

Assume now $X \in L^1(\mathbb{Q})$; then the right hand side of (7.9) is well-defined, since $X Z_t \in L^1(\mathbb{P})$ and there is no problem to divide by Z_s, since we just showed that $Z_s > 0$ a.s. Next, we show that $\frac{\mathbb{E}_{\mathbb{P}}[X Z_t|\mathcal{F}_s]}{Z_s} \in L^1(\mathbb{Q})$; since this is a $\mathcal{F}_s$–measurable random variable, we have

$$\mathbb{E}_{\mathbb{Q}}\left[\left|\frac{\mathbb{E}_{\mathbb{P}}[X Z_t|\mathcal{F}_s]}{Z_s}\right|\right] \leqslant \mathbb{E}_{\mathbb{P}}\left[\frac{\mathbb{E}_{\mathbb{P}}[|X Z_t||\mathcal{F}_s]}{Z_s} Z_s\right] = \mathbb{E}_{\mathbb{P}}[|X Z_t|] < \infty.$$

It remains to verify the defining property of the conditional expectation. For any $A \in \mathcal{F}_s$, we get:

$$\begin{aligned}\mathbb{E}_{\mathbb{Q}}[\mathbb{1}_A X] &= \mathbb{E}_{\mathbb{P}}[\mathbb{1}_A X Z_t] = \mathbb{E}_{\mathbb{P}}[\mathbb{1}_A \mathbb{E}_{\mathbb{P}}[X Z_t|\mathcal{F}_s]] \\ &= \mathbb{E}_{\mathbb{P}}\left[\mathbb{1}_A \mathbb{E}_{\mathbb{P}}[X Z_t|\mathcal{F}_s]\frac{Z_s}{Z_s}\right] = \mathbb{E}_{\mathbb{Q}}\left[\mathbb{1}_A \frac{\mathbb{E}_{\mathbb{P}}[X Z_t|\mathcal{F}_s]}{Z_s}\right]\end{aligned}$$

which proves the claim. □

Exercise. What can we say if we only know that $\mathbb{Q} \ll \mathbb{P}$ but not necessarily $\mathbb{Q} \sim \mathbb{P}$? In this case, how should we interpret the identity

$$\mathbb{E}_{\mathbb{Q}}[X|\mathcal{F}_s] = \frac{\mathbb{E}_{\mathbb{P}}[X Z_t|\mathcal{F}_s]}{Z_s}?$$

We want to understand what happens with (local) martingales if we change the probability measure. We can achieve this with the help of Bayes' formula and Itô's formula. We start with the following auxiliary result:

Corollary 7.23. *Let $\mathbb{Q} \sim \mathbb{P}$ and let $Z_t = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big|_{\mathcal{F}_t}$, $t \geqslant 0$ (càdlàg modification). For any adapted càdlàg M we have:*

$$\begin{aligned} M \text{ is a } \mathbb{Q}\text{–martingale} &\iff MZ \text{ is a } \mathbb{P}\text{–martingale} \\ M \text{ is a } \mathbb{Q}\text{–local martingale} &\iff MZ \text{ is a } \mathbb{P}\text{–local martingale.}\end{aligned}$$

Proof. We first show the "non-local" version. By Proposition 7.22, M is $\mathbb{Q}$-integrable if and only if MZ is $\mathbb{P}$-integrable; by assumption, M and MZ are adapted. For any $s < t$, Bayes' formula yields:

$$M_s = \mathbb{E}_{\mathbb{Q}}[M_t|\mathcal{F}_s] = \frac{\mathbb{E}_{\mathbb{P}}[M_t Z_t|\mathcal{F}_s]}{Z_s} \qquad \iff \qquad M_s Z_s = \mathbb{E}_{\mathbb{P}}[M_t Z_t|\mathcal{F}_s],$$

where the identities hold a.s. (under both $\mathbb{P}$ and $\mathbb{Q}$, which have the same null sets).

For the claim about the local martingale property, let (τ_n) be a localizing sequence (under one and then under both measures). We just saw that

$$M^{\tau_n} \text{ is a } \mathbb{Q}\text{–martingale} \iff M^{\tau_n} Z \text{ is a } \mathbb{P}\text{–martingale.}$$

We would like to see $M^{\tau_n} Z^{\tau_n}$ on the right hand side, so we need to show that $M^{\tau_n} Z$ is a $\mathbb{P}$–martingale if and only if $M^{\tau_n} Z^{\tau_n}$ is a $\mathbb{P}$-martingale. Since $M_t^{\tau_n}$ is $\mathcal{F}_{\tau_n \wedge t}$–measurable, we get

$$\mathbb{E}_{\mathbb{P}}[|M_t^{\tau_n} Z_t|] = \mathbb{E}_{\mathbb{Q}}[|M_t^{\tau_n}|] = \mathbb{E}_{\mathbb{P}}[|M_t^{\tau_n} Z_{t\wedge\tau_n}|] = \mathbb{E}_{\mathbb{P}}[|M_t^{\tau_n} Z_t^{\tau_n}|],$$

so $M^{\tau_n}Z$ is integrable if and only if $M^{\tau_n}Z^{\tau_n}$ is integrable. To conclude, it suffices to check that

$$M_t^{\tau_n}Z_t - M_t^{\tau_n}Z_t^{\tau_n} = M_{\tau_n}Z_{\tau_n\wedge t,t} \text{ is a martingale.} \tag{7.10}$$

From (7.10) we can then deduce that

$$M^{\tau_n} \text{ is a } \mathbb{Q}\text{–mart.} \quad \Leftrightarrow \quad M^{\tau_n}Z \text{ is a } \mathbb{P}\text{–mart.} \quad \Leftrightarrow \quad M^{\tau_n}Z^{\tau_n} \text{ is a } \mathbb{P}\text{–mart.}$$

which concludes the proof.

Verification of (7.10) is elementary but tedious, skipped in the lectures but given here for completeness:

Let $s \leqslant t$, then

$$\mathbb{E}[M_{\tau_n}Z_{\tau_n\wedge t,t}|\mathcal{F}_s] = \mathbb{E}[M_{\tau_n}Z_{\tau_n\wedge t,t}\,\mathbb{1}_{\tau_n\leqslant s}|\mathcal{F}_s] + \mathbb{E}[M_{\tau_n}Z_{\tau_n\wedge t,t}\,\mathbb{1}_{\tau_n>s}|\mathcal{F}_s].$$

For the first term, noticing that $M_{\tau_n}\mathbb{1}_{\tau_n\leqslant s}$ is $\mathcal{F}_s$-measurable, by the stopping theorem for Z we find

$$\mathbb{E}[M_{\tau_n}Z_{\tau_n\wedge t,t}\,\mathbb{1}_{\tau_n\leqslant s}|\mathcal{F}_s] = M_{\tau_n}Z_{\tau_n\wedge s,s}\,\mathbb{1}_{\tau_n\leqslant s} = M_{\tau_n}Z_{\tau_n\wedge s,s}.$$

To conclude, it then suffices to show that for any $A \in \mathcal{F}_s$ it holds

$$\mathbb{E}[M_{\tau_n}Z_{\tau_n\wedge t,t}\,\mathbb{1}_{\tau_n>s}\,\mathbb{1}_A] = 0.$$

Notice that $\mathbb{1}_{\tau_n>s}\,\mathbb{1}_A \in \mathcal{F}_{\tau_n}$, therefore by conditional expectation and stopping theorem we have

$$\mathbb{E}[M_{\tau_n}Z_{\tau_n\wedge t,t}\,\mathbb{1}_{\tau_n>s}\,\mathbb{1}_A] = \mathbb{E}[M_{\tau_n}\mathbb{1}_{\tau_n>s}\,\mathbb{1}_A\,\mathbb{E}[Z_t - Z_{t\wedge\tau_n}|\mathcal{F}_{\tau_n}]] = 0.$$

Combining everything, this shows (7.10). □

Exercise. What do we get if we only know that $\mathbb{Q} \ll \mathbb{P}$?

——— End of the lecture on January 23 ———

To apply the previous result, we need to understand the density process Z better. For that purpose, we need the following result.

Proposition 7.24. (Stochastic logarithm) *Let Z be a continuous $\mathbb{P}$–local martingale such that $\mathbb{P}$–almost surely $Z_t > 0$ for all $t \geqslant 0$. Then there exists a unique continuous $\mathbb{P}$–local martingale $L = \mathcal{L}(Z)$, called the* stochastic logarithm *of Z, such that*

$$Z_t = \mathcal{E}(L)_t = \exp\left(L_t - \frac{1}{2}\langle L\rangle_t\right), \qquad t \geqslant 0.$$

If moreover $Z_0 \in L^1(\mathbb{P})$, then Z is a supermartingale; in this case, given $T \in (0, +\infty]$, Z is a martingale on $[0, T]$ if and only if $\mathbb{E}[Z_T] = \mathbb{E}[Z_0]$.

Proof. See Exercise Sheet 13 (for $Z_0 = 1$, the general case is analogous). □

We are now ready to state and prove Girsanov's theorem, which tells us exactly how continuous local martingales are affected by an equivalent change of the underlying probability measure.

Theorem 7.25. (Girsanov) *Let $\mathbb{Q}\sim\mathbb{P}$ and assume that the density process $Z_t = \frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}}\Big|_{\mathcal{F}_t}$, $t\geqslant 0$, is continuous. Let L be the stochastic logarithm of Z. If M is a continuous $\mathbb{P}$–local martingale, then*

$$\tilde{M} = M - \langle M, L\rangle$$

is a continuous $\mathbb{Q}$–local martingale. Here $\langle M, L\rangle$ and $\langle M, Z\rangle$ are the quadratic covariation processes under $\mathbb{P}$.

Remark 7.26. In hindsight, we then deduce that $M = \tilde{M} + \langle M, L\rangle$ is still a $\mathbb{Q}$-semimartingale, and therefore we can also define $\langle M, Z\rangle$ and $\langle M, L\rangle$ under $\mathbb{Q}$ and we have

$$\langle M, Z\rangle = \langle \tilde{M}, Z\rangle, \quad \langle M, L\rangle = \langle \tilde{M}, L\rangle.$$

Thanks to the approximation result in Lemma 5.45, and the fact that $\mathbb{Q}\sim\mathbb{P}$, the quadratic covariations under $\mathbb{Q}$ and under $\mathbb{P}$ are the same, since they can be obtained as the ucp limit (with respect to both $\mathbb{P}$ and $\mathbb{Q}$) of sums of squared increments obtained by deterministic partitions of infinitesimal mesh.

Proof of Theorem 7.25. We know from Corollary 7.23 that $\tilde{M}$ is a $\mathbb{Q}$–local martingale if and only if $\tilde{M}Z$ is a $\mathbb{P}$–local martingale. Recall from Remark 7.9 that (w.r.t. $\mathbb{P}$) Z satisfies

$$\mathrm{d}Z = Z\mathrm{d}L.$$

Integration by parts (manipulated in differential form for simplicity) then yields

$$\begin{aligned}
\mathrm{d}(\tilde{M}Z)_t &= \tilde{M}_t\,\mathrm{d}Z_t + Z_t\,\mathrm{d}\tilde{M}_t + \mathrm{d}\langle \tilde{M}, Z\rangle_t\\
&= \tilde{M}_t\,\mathrm{d}Z_t + Z_t\,\mathrm{d}M_t - Z_t\,\mathrm{d}\langle M, L\rangle_t + \mathrm{d}\langle M, Z\rangle_t\\
&= \tilde{M}_t\,\mathrm{d}Z_t + Z_t\,\mathrm{d}M_t - Z_t\,\mathrm{d}\langle M, L\rangle_t + Z_t\,\mathrm{d}\langle M, L\rangle_t\\
&= \tilde{M}_t\,\mathrm{d}Z_t + Z_t\,\mathrm{d}M_t.
\end{aligned}$$

Since Z and M are both $\mathbb{P}$–local martingales, the right hand side defines a $\mathbb{P}$–local martingale. □

Remark 7.27. The same statement and proof also work on finite time horizons: for fixed $T\in(0,+\infty)$, if $\mathbb{Q}|_{\mathcal{F}_T}\sim\mathbb{P}|_{\mathcal{F}_T}$ and M is a $\mathbb{P}$-local martingale on $[0,T]$, then

$$\tilde{M}_t = M_t - \langle M, L\rangle_t \qquad t\in[0,T],$$

is a $\mathbb{Q}$–local martingale on $[0,T]$; by this we mean that the stopped processes $(\tilde{M}_t^{T_n})_{t\in[0,T]}$ are martingales, where stopping after T has no effect.

More generally, Girsanov's theorem tells us that continuous $\mathbb{P}$-semimartingales are still continuous $\mathbb{Q}$-semimartingales, under an equivalent change of measure $\mathbb{Q}\sim\mathbb{P}$.

Corollary 7.28. *In the setting of Girsanov's theorem, any continuous $\mathbb{P}$–semimartingale $X = X_0 + M + A$ is also a continuous $\mathbb{Q}$–semimartingale, with decomposition under $\mathbb{Q}$ given by*

$$X = X_0 + (M - \langle M, L\rangle) + (A + \langle M, L\rangle) =: X_0 + \tilde{M} + \tilde{A}.$$

Theorem 7.29. *Let B be a d-dimensional $\mathbb{P}$-Brownian motion and let H be a d-dimensional progressively measurable process such that $\mathbb{P}$-almost surely*

$$\int_0^T |H_s|^2\,\mathrm{d}s = \sum_{i=1}^d \int_0^T |H_s^i|^2\,\mathrm{d}s < \infty \quad \forall T\geqslant 0,$$

so that in particular $H^i \in L^2_{\mathrm{loc}}(B^i)$ *for* $i = 1, \ldots, d$. *Assume that the process*

$$Z_t = \exp\left(\int_0^t H_s \cdot \mathrm{d}B_s - \frac{1}{2}\int_0^t |H_s|^2 \mathrm{d}s\right) = \exp\left(\sum_{i=1}^d \int_0^t H_s^i \, \mathrm{d}B_s^i - \frac{1}{2}\int_0^t |H_s|^2 \mathrm{d}s\right), \qquad t \geqslant 0,$$

is a uniformly integrable martingale such that $Z_\infty > 0$ $\mathbb{P}$*-a.s. Define the probability measure* $\mathrm{d}\mathbb{Q} = Z_\infty \mathrm{d}\mathbb{P}$. *Then under* $\mathbb{Q}$ *the process*

$$\tilde{B} = B - \int_0^\cdot H_s \, \mathrm{d}s$$

is a d*-dimensional Brownian motion.*

Proof. Since $Z_\infty > 0$ we have $\mathbb{P} \ll \mathbb{Q}$ and thus $\mathbb{Q} \sim \mathbb{P}$: indeed, for any $A \in \mathcal{F}$ such that $\mathbb{Q}(A) = 0$, we must have

$$\mathbb{E}_{\mathbb{P}}[\mathbb{1}_A Z_\infty] = \mathbb{Q}(A) = 0,$$

which is only possible if $\mathbb{P}(A) = 0$. Therefore, we can apply Girsanov's theorem: noting that $Z = \mathcal{E}(L)$ for $L_t = \int_0^t H_s \cdot \mathrm{d}B_s$, and

$$\langle L, B^i \rangle_t = Z_t = \sum_{j=1}^d \left\langle \int_0^\cdot H_s^j \mathrm{d}B_s^j, B^i \right\rangle_t = \int_0^t H_s^i \, \mathrm{d}s,$$

we deduce that $\tilde{B} = B - \int_0^\cdot H_s \mathrm{d}s$ is a d-dimensional $\mathbb{Q}$-local martingale. Moreover, since the process $\int_0^\cdot H_s \, \mathrm{d}s$ is of finite variation, by Lemmas 5.16-5.45 we can find a deterministic sequence of partitions of infinitesimal mesh such that $\mathbb{P}$-a.s.

$$\lim_{n \to \infty} \sum_{k=0}^{n-1} \tilde{B}^i_{t \wedge t_k^n, t \wedge t_{k+1}^n} \tilde{B}^j_{t_k^n, t_{k+1}^n} = \delta_{ij} t \quad \text{uniformly on compact sets;}$$

since $\mathbb{Q} \sim \mathbb{P}$, the same convergence holds $\mathbb{Q}$-a.s. and therefore

$$\mathbb{Q}(\langle \tilde{B}^i, \tilde{B}^j \rangle_t = \delta_{ij} t) = 1.$$

By Lévy's characterization it then follows that $\tilde{B}$ is a d-dimensional Brownian motion under $\mathbb{Q}$. □

Exercise. Let $d = 1$. Find an example of an H that works for the previous theorem.

As an consequence of Theorem 7.29, we get the following result. One may loosely think of it as a generalization of our initial example about translations of d-dimensional standard Gaussians Z by $h \in \mathbb{R}^d$, upon enforcing the correspondence $B \leftrightarrow Z$, $H \leftrightarrow h$, $\int_0^\cdot H_s \mathrm{d}B_s \leftrightarrow h \cdot Z$ and $\int_0^\cdot |H_s|^2 \mathrm{d}s \leftrightarrow |h|^2$.

Corollary 7.30. (Cameron-Martin formula) *Let* B *be a* d*-dimensional Brownian motion and let* $h \in L^2(\mathbb{R}_+, \mathbb{R}^d)$. *Let* $\Phi\colon C(\mathbb{R}_+, \mathbb{R}^d) \to \mathbb{R}$ *be a measurable function and assume that* Φ *is either bounded or nonnegative. Then*

$$\mathbb{E}\left[\Phi\left(B + \int_0^\cdot h_s \, \mathrm{d}s\right)\right] = \mathbb{E}\left[\Phi(B) \exp\left(\int_0^\infty h_s \cdot \mathrm{d}B_s - \frac{1}{2}\int_0^\infty |h_s|^2 \, \mathrm{d}s\right)\right].$$

Proof. This follows from the previous theorem, provided we can show that

$$Z_t = \exp\left(\int_0^t h_s \cdot \mathrm{d}B_s - \frac{1}{2}\int_0^t |h_s|^2 \mathrm{d}s\right), \qquad t \geqslant 0,$$

is a uniformly integrable martingale with $Z_\infty > 0$. But $\int_s^t h_r \cdot \mathrm{d}B_r$ is independent of $\mathcal{F}_s$ and it has the distribution $\mathcal{N}(0, \int_s^t |h_r|^2 \mathrm{d}r)$ because the integrand is deterministic (using that limits of Gaussian random variables are Gaussian, and that we can approximate the integral $\int_s^t h_r \, \mathrm{d}B_r$ by step functions; alternatively, recall that for determinstic integrands, $\int_0^t h_s \cdot \mathrm{d}B_s$ coincides with the Wiener integral we constructed at the beginning of the course). Therefore,

$$\mathbb{E}[Z_t|\mathcal{F}_s] = Z_s \, \mathbb{E}\bigg[\exp\bigg(\int_s^t h_r \cdot \mathrm{d}B_r - \frac{1}{2}\int_s^t |h_r|^2 \mathrm{d}r\bigg)\bigg|\mathcal{F}_s\bigg] = Z_s,$$

and thus Z is a martingale. Moreover we can explicitly compute higher moments since we know that $\int_0^t h_s \cdot \mathrm{d}B_s \sim \mathcal{N}(0, \int_s^t |h_r|^2 \mathrm{d}r)$:

$$\begin{aligned} \mathbb{E}[Z_t^2] &= \mathbb{E}\bigg[\exp\bigg(2\int_0^t h_s \cdot \mathrm{d}B_s - \int_0^t |h_s|^2 \mathrm{d}s\bigg)\bigg] \\ &= \exp\bigg(\frac{2^2}{2}\int_0^t |h_s|^2 \mathrm{d}s - \int_0^t |h_s|^2 \mathrm{d}s\bigg) \\ &\leqslant \exp\bigg(\int_0^\infty |h_s|^2 \mathrm{d}s\bigg). \end{aligned}$$

This shows uniform integrability of Z. Finally, Z_∞ is an exponential and therefore strictly positive. □

Example 7.31. (Brownian motion with drift) Let B be a 1-dimensional Brownian motion, $\theta \in \mathbb{R}$ with $\theta \neq 0$, and consider $\tilde{B}_t = B_t + \theta t = B_t + \int_0^t \theta \, \mathrm{d}s$. Here we cannot apply directly the previous result, as $\theta \notin L^2(\mathbb{R}_+)$. However, since $\theta \in L^2([0,T])$ for any finite T, we can apply Girsanov's theorem on $[0,T]$ (equivalently, we may apply it to B^T, or replace θ by $h_s = \theta \, \mathbb{1}_{[0,T]}(s)$) to deduce that $(\tilde{B}_t)_{t\in[0,T]}$ and $(B_t)_{t\in[0,T]}$ are equivalent probability measures on $C([0,T];\mathbb{R})$, and for any bounded $\Phi: C([0,T];\mathbb{R}) \to \mathbb{R}$ it holds

$$\mathbb{E}[\Phi(\tilde{B})] = \mathbb{E}\bigg[\Phi(B)\exp\bigg(\theta B_T - \frac{\theta^2}{2}T\bigg)\bigg].$$

The restriction to finite T here is necessary, as the laws of $(\tilde{B}_t)_{t\in\mathbb{R}_+}$ and $(B_t)_{t\in\mathbb{R}_+}$ are singular with respect to each other: for instance if $\theta > 0$, as a consequence of the law of iterated logarithm we know that $\mathbb{P}$-a.s.

$$\liminf_{t\to\infty} \tilde{B}_t = \lim_{t\to\infty} \tilde{B} = +\infty, \quad \liminf_{t\to\infty} B_t = -\infty.$$

Indeed, one can see that the Girsanov density $Z_T = \exp\Big(\theta B_T - \frac{\theta^2}{2} T\Big) \to 0$ as $T \to \infty$.

——— End of the lecture on January 29 ———

As the previous results and examples have shown, often in applications (especially when dealing with Brownian motion), we want to "reverse engineer" Girsanov's theorem, in the following sense:

1. As a starting point, we know what is the drift counterterm $-\langle M, L\rangle$ we would like to see appearing as an effect of the change of measure.
2. From this we reconstruct the candidate L and therefore ultimately the candidate density process $Z = \mathcal{E}(L)$.

3. Ultimately, to verify that everything works as planned, we need to know that the assumptions of Theorem 7.25 are satisfied. Namely, we must check that Z is a genuine martingale and that $\mathbb{Q}$ defined via $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = Z_\infty$ is equivalent to $\mathbb{P}$.

In the setting of Corollary 7.30, we were able to do so since the integrand h was deterministic, and so everything was very explicit. But for general local martingales L, Step 3. can be quite hard, as the exponential martingale is only guaranteed to be a local martingale (and a supermartingale, cf. Proposition 7.24). Think of $L = \int_0^\cdot H_s \cdot \mathrm{d}B_s$, for some non-trivial random integrand H; already taking $H = B$ poses a significant challenge.

The following condition is often very useful:

Theorem 7.32. (Novikov's criterion) *Let $L \in \mathcal{M}_{\mathrm{loc}}^{\mathrm{c}}$ with $L_0 = 0$. If*

$$\mathbb{E}\left[e^{\frac{1}{2}\langle L\rangle_\infty}\right] < \infty,$$

then $Z = \mathcal{E}(L) = \exp(L - \frac{1}{2}\langle L\rangle)$ is a uniformly integrable martingale; moreover the probability measure $\mathbb{Q}$ defined by $\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = Z_\infty$ is equivalent to $\mathbb{P}$.

Proof. See Le Gall [16], Theorem 5.23. □

Exercise. Find now an interesting example of a random integrand H that works for Girsanov's theorem for Brownian motion.

Exercise. In the setting of Theorem 7.32, since $\mathbb{P} \ll \mathbb{Q}$, it follows that there must exist another exponential martingale $\bar{Z}$, of the form $\bar{Z} = \mathcal{E}(\bar{L})$ for a unique $\mathbb{Q}$-local martingale $\bar{L}$, such that $\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} = \bar{Z}_\infty$. Find the expressions for $\bar{Z}$ and $\bar{L}$.

8 Stochastic differential equations

The references for this section include most of the previous ones, since most monographs on stochastic analysis also treat stochastic differential equations.

8.1 First examples

In order to properly motivate the usefulness of *stochastic differential equations* (SDEs) as a more refined, random version of their deterministic counterpart given by *ordinary differential equations* (ODEs), we start by providing some examples. They will also justify the need for an abstract solution theory, which applies in situations where deriving an explicit solution formula might be no longer possible.

Example 8.1. (Malthusian growth model) The Malthusian growth model (introduced by Malthus in 1798) is the simplest continuous-time model describing the evolution of a population size. Assume that each individual of the population on average gives birth to a new individual with rate $b > 0$, and dies with rate d. Setting $r = b - d \in \mathbb{R}$, this leads to the ODE

$$\frac{\mathrm{d}}{\mathrm{d}t} X_t = r X_t, \qquad X_0 = x_0 \geqslant 0, \tag{8.1}$$

where X_t is the number of individuals at time t and $x \geqslant 0$ is a given initial condition, namely the population size at the initial time $t = 0$. The unique solution to (8.1) is given by

$$X_t = x_0 \, e^{rt}. \tag{8.2}$$

So we see exponential growth for $r > 0$ and exponential decay for $r < 0$. For very large populations, this might be a reasonable approximation; but for populations of "finite size", we will see random fluctuations, because the rate r only holds on average and not every individual behaves the same. If those fluctuations are independent in time and of finite variance, then by the central limit theorem/Donsker's invariance principle, we expect them to be Gaussian, so time-dependent multiples of Brownian increments. For simplicity, let us assume the fluctuations to be stationary in time, after all we already chose r independent of time. This gives rise to the *stochastic differential equation (SDE)*

$$\mathrm{d}X_t = r X_t \,\mathrm{d}t + \sigma \,\mathrm{d}B_t, \qquad X_0 = x_0, \tag{8.3}$$

where σ is the "size of random fluctuations". Eq. (8.3) should be read in integrated form:

$$X_t = x_0 + \int_0^t r X_s \mathrm{d}s + \sigma B_t.$$

There is however a problem with this equation: even though X models the population size (so it should be a nonnegative number), since the Brownian motion B can become very negative, it can "push X below 0" with positive probability. Therefore, our new equation is maybe not so suitable for modeling a population size. Also, it is not so natural to assume that the fluctuations are independent of the state: If there are 10^{10} individuals, the oscillations should be stronger than if there are 100 individuals. It is therefore more natural to also have the fluctuations depend linearly on X_t, namely to consider instead the SDE

$$\mathrm{d}X_t = r X_t \,\mathrm{d}t + \sigma X_t \,\mathrm{d}B_t, \qquad X_0 = x_0,$$

or in integrated form

$$X_t = x_0 + \int_0^t r X_s \,\mathrm{d}s + \int_0^t \sigma X_s \,\mathrm{d}B_s. \tag{8.4}$$

For $r = 0$, we have already seen in Proposition 7.8 that a solution to (8.4) is given by

$$X_t = x_0 \,\mathcal{E}(\sigma B)_t = x_0 \exp\!\left(\sigma B_t - \frac{1}{2}\sigma^2 t\right). \tag{8.5}$$

What if $r \neq 0$? By extrapolating the formulas (8.2)-(8.5), corresponding to the two special cases $\sigma = 0$ and $r = 0$, we can guess that the general formula solution formula for (8.4) is given by

$$X_t = x_0 \exp\!\left(r t + \sigma B_t - \frac{1}{2}\sigma^2 t\right) = x_0 \exp\!\left(\sigma B_t + \left(r - \frac{\sigma^2}{2}\right) t\right). \tag{8.6}$$

Indeed, by applying Itô's formula, one can see that X given by (8.6) solves (8.4).

Is X the only solution? To verify that this is indeed the case, first notice that since X is an exponential, it is strictly positive, therefore by Itô's formula $Z_t := X_t^{-1}$ solves

$$\begin{aligned}\mathrm{d}Z_t &= -\frac{1}{X_t^2}\,\mathrm{d}X_t + \frac{1}{X_t^3}\,\mathrm{d}\langle X\rangle_t \\ &= -\frac{r}{X_t}\mathrm{d}t - \frac{\sigma}{X_t}\,\mathrm{d}B_t + \frac{\sigma^2}{X_t}\,\mathrm{d}t \\ &= -(r - \sigma^2)\, Z_t \,\mathrm{d}t - \sigma Z_t \,\mathrm{d}B_t\end{aligned}$$

where in the computation we used the SDE (8.4) satisfied by X itself. In particular, we see that Z also solves an SDE. Now assume let us assume that Y is another solution to (8.4) starting at $Y_0 = X_0$; then the integration by parts formula gives

$$\mathrm{d}(YZ) = -YZ(r - \sigma^2)\mathrm{d}t - \sigma YZ\,\mathrm{d}B_t + rYZ\mathrm{d}t + \sigma YZ\mathrm{d}B_t - \sigma^2\, YZ\,\mathrm{d}\langle B\rangle_t = 0$$

so that

$$Y_t Z_t = Y_0 Z_0 = 1 \quad \Longrightarrow \quad Y_t = Z_t^{-1} = X_t \quad \forall t \geqslant 0.$$

So uniqueness of solutions holds holds. Moreover, we know that $\mathcal{E}(\sigma B)$ is a martingale, and therefore

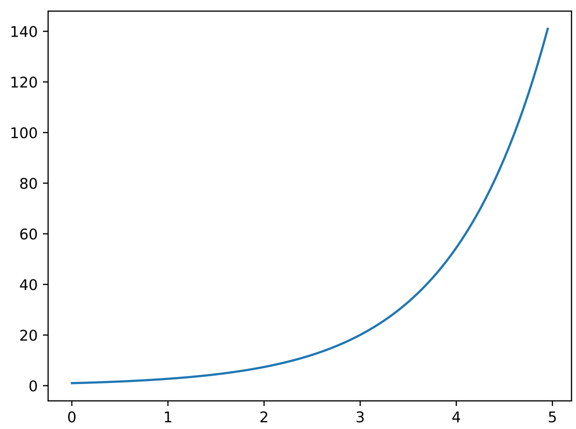
$$\mathbb{E}[X_t] = x_0 \mathbb{E}[\mathcal{E}(\sigma B)_t e^{rt}] = x_0 e^{rt};$$

in expectation, the solution X_t to our stochastic model (8.4) behaves like the deterministic model x_t to the deterministic model (8.1) with $\sigma = 0$.

Let us look at some simulations. We first simulate the deterministic model with a simple Euler scheme, for $x_0 = 1$ and $r = 1$:

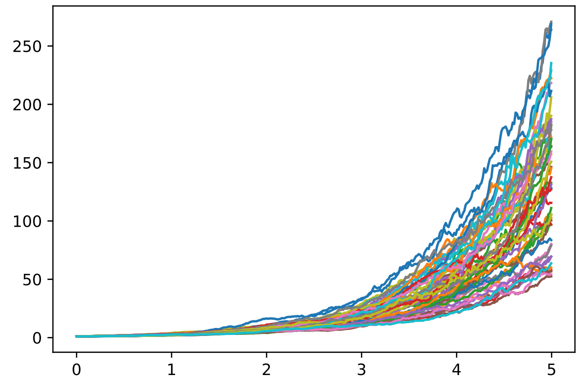
```
Python 3.7.4 (default, Aug 13 2019, 15:17:50)
[Clang 4.0.1 (tags/RELEASE_401/final)]
Python plugin for TeXmacs.
Please see the documentation in Help -> Plugins -> Python
```

```
>>> import numpy as np
    import matplotlib.pyplot as plt

    T, h = 5, 1e-3
    n=int(T/h)

    X_0, r = 1, 1
    time = np.arange(0,T+h,h)
    X = np.zeros(n+1)
    X[0] = X_0

    for i in range(n):
        X[i+1] = X[i] + r*X[i]*h

    plt.clf()
    plt.plot(time[1::50],X[1::50])

    pdf_out(plt.gcf())
```



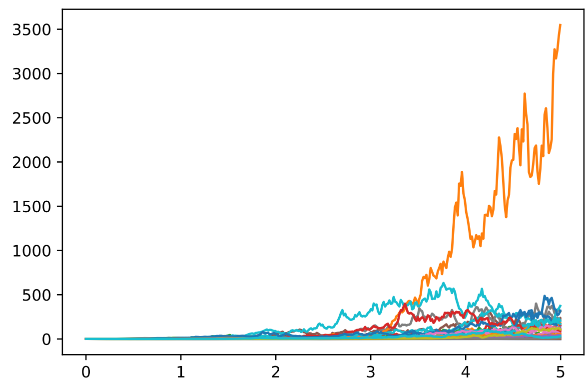
We observe exponential growth. Now let us add noise, so take $\sigma \neq 0$. To make the behavior more transparent, we don't only simulate one trajectory, but 50 trajectories with the same initial conditions and the same parameters, but different realizations of the noise. We take $\sigma = 0.2$.

```
>>> T, h = 5, 1e-3
    n = int(T/h)
    k = 50

    X_0, r, sigma = 1, 1, 0.2
    time = np.arange(0,T+h,h)
    dB = np.sqrt(h)*(np.random.randn(k,n))
    X = np.zeros((k,n+1))
    X[:,0] = X_0


    plt.clf()

    for i in range(n):
        X[:,i+1]=X[:,i] + r*X[:,i]*h + sigma*X[:,i]*dB[:,i]

    for i in range(k):
        plt.plot(time[1::15],X[i,1::15])

    pdf_out(plt.gcf())
```



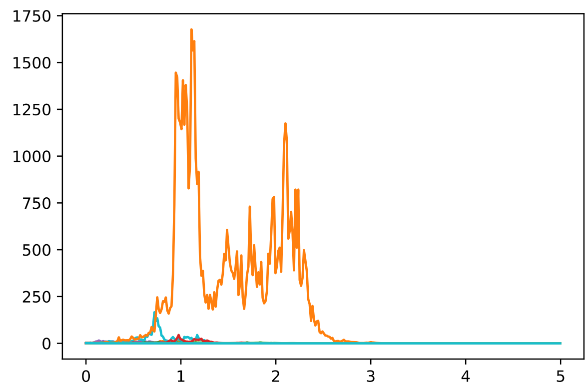
So we see trajectories that qualitatively look like the trajectories of the deterministic model, but they are "dispersed" and the precise behavior depends on the realization of the noise. Now let us crank up the noise and consider $\sigma = 1$:

```
>>> T, h = 5, 1e-3
    n = int(T/h)
    k = 50

    X_0, r, sigma = 1, 1, 1
    time = np.arange(0,T+h,h)
    dB = np.sqrt(h)*(np.random.randn(k,n))
    X = np.zeros((k,n+1))
    X[:,0] = X_0


    plt.clf()

    for i in range(n):
        X[:,i+1]=X[:,i] + r*X[:,i]*h + sigma*X[:,i]*dB[:,i]

    for i in range(k):
        plt.plot(time[1::15],X[i,1::15])

    pdf_out(plt.gcf())
```



The system looks much more oscillatory and erratic now, especially if we take into account that the y-axis scaled and now we see much larger values. Remarkably, now there is no clear growth any more, but the population oscillates up and down. Let us take an even bigger σ, say $\sigma = 3$:

```
>>> T, h = 5, 1e-3
    n = int(T/h)
    k = 50

    X_0, r, sigma = 1, 1, 3
    time = np.arange(0,T+h,h)
    dB = np.sqrt(h)*(np.random.randn(k,n))
    X = np.zeros((k,n+1))
    X[:,0] = X_0


    plt.clf()

    for i in range(n):
        X[:,i+1]=X[:,i] + r*X[:,i]*h + sigma*X[:,i]*dB[:,i]

    for i in range(k):
        plt.plot(time[1::15],X[i,1::15])

    pdf_out(plt.gcf())
```




```
>>>
```

Now the system seems to have a tendency to collapse and to quickly converge to 0. This corresponds to the population eventually becoming extinct, and actually doing so faster as σ gets larger. So in this model more randomness seems to be disadvantageous for the population growth.

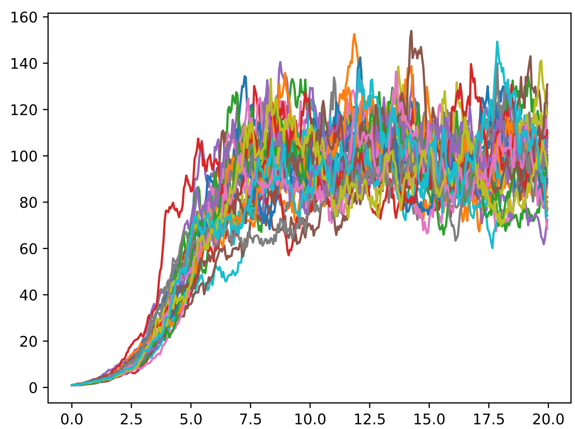
Exercise. The long time behavior of the stochastic model may be very different from that of the deterministic model. We say that the model is *stable* if for initial conditions x “close to 0” (in this simple example it is not necessary to specify what we mean by that) we have $\lim_{t\to\infty} X_t = 0$ almost surely. Show that in the deterministic case $\sigma = 0$ the system is stable if and only if $r < 0$, while in the stochastic case $\sigma > 0$ the system is stable if and only if $r < \frac{1}{2}\sigma^2$. There is a *stabilization by noise* effect.

Example 8.2. (Logistic growth model) The previous discussion strongly depends on the linearity of the equation. However, the principle that our population can grow indefinitely is rather irrealistic and eventually this growth shoudl become unsustainable, for instance because the environmental resources (e.g. food) are limited. We could make the model more realistic by considering a decreasing function $f: \mathbb{R}_+ \to \mathbb{R}_+$ and changing the birth rate to $b\,f(X_t)$, which leads to the equation

$$\mathrm{d}X_t = (b\,f(X_t) - d)X_t\,\mathrm{d}t + \sigma X_t\,\mathrm{d}B_t, \qquad X_0 = x.$$

In this way, for small X, the population grows, but as X the function f has a penalizing effect, to the point where $b\,f(x) - d$ becomes negative for x too large; the population then shrink over time, until it reached the regime where $b\,f(x) - d \geqslant 0$, and so on. The simplest choice of f is given by $\tilde{f}(x) = 1 - \frac{x}{M}$, where M represents a given "maximum size" of individuals that the environment can support. In this case, the resulting ODE/SDE (depending on whether $\sigma = 0$ or $\sigma > 0$) is called the *logistic growth model*.

```
>>> import numpy as np
    import matplotlib.pyplot as plt

    T, h = 20, 1e-3
    n = int(T/h)
    k = 30

    X_0, r, sigma, M = 1, 1, 0.2, 100
    time = np.arange(0,T+h,h)
    dB = np.sqrt(h)*(np.random.randn(k,n))
    X = np.zeros((k,n+1))
    X[:,0] = X_0


    plt.clf()
    for i in range(n):
        X[:,i+1]=X[:,i] + r*(1-X[:,i]/M)*X[:,i]*h + sigma*X[:,i]*dB[:,i]

    for i in range(k):
        plt.plot(time[1::50],X[i,1::50])

    pdf_out(plt.gcf())
```




```
>>>
```

In the numerical simulation of $\mathrm{d}X_t = r\left(1 - \frac{X_t}{M}\right)X_t\,\mathrm{d}t + \sigma X_t\,\mathrm{d}B_t$ for small σ, the population size oscillates around the "saturation size" M.

Example 8.3. So far we focused on one-dimensional models, where X is a number. But in general our model can consist of several "components", which we can represent as vector $X=(X^1,\ldots,X^d)$. This is for instance the case if we want to differentiate between different types of individuals (or species) among our population. Therefore, in general we are interested in solving *multidimensional SDEs*.

Here are some examples which are relevant for applications:

i. *Stochastic Lotka–Volterra equation*: We model a biological system in which two species interact, one as prey and one as predator (say rabbits and foxes). Let X^1 be the number of prey and X^2 the number of predators. Then we postulate the stochastic differential equation

$$\begin{aligned} \mathrm{d}X_t^1 &= (a-bX_t^2)X_t^1\mathrm{d}t+\sigma_1 X_t^1\mathrm{d}B_t^1, \\ \mathrm{d}X_t^2 &= (-c+gX_t^1)X_t^2\mathrm{d}t+\sigma_2 X_t^2\mathrm{d}B_t^2. \end{aligned}$$

Here, $a,b,c,g,\sigma_1,\sigma_2$ are positive constants, and $B=(B^1,B^2)$ is a two-dimensional Brownian motion.

ii. *Stochastic FitzHugh–Nagumo equation*: We model the state of a neuron, where we keep track of the membrane voltage X^1 and a "delay variable" X^2. We postulate the stochastic differential equation

$$\begin{aligned} \mathrm{d}X_t^1 &= (X_t^1-(X_t^1)^3+X_t^2+I_{\mathrm{ext}})\mathrm{d}t+\mathrm{d}B_t, \\ \mathrm{d}X_t^2 &= (X_t^1+1-X_t^2)\mathrm{d}t, \end{aligned}$$

where $I_{\mathrm{ext}}\in\mathbb{R}$ is the "external stimulus" and B is a Brownian motion.

iii. *Stochastic SIR model:* We model the spread of a disease across a population by distinguishing between:

a) healthy individuals who might contract the disease in the future (S for *susceptible)*,

b) sick individuals (I for *infected*),

c) individuals who do not fall into the previous categories, because they either developed immunity or died after contracting the disease (R for *recovered* or *removed*).

The resulting system $X_t=(S_t,I_t,R_t)$ is devised in such a way that the overall amount of individuals $N_t=S_t+I_t+R_t$ stays constant over time:

$$\begin{aligned} \mathrm{d}S_t &= -\beta S_t I_t\,\mathrm{d}t-\sigma S_t I_t\,\mathrm{d}B_t, \\ \mathrm{d}I_t &= (\beta S_t I_t-\gamma I_t)\,\mathrm{d}t+\sigma S_t I_t\,\mathrm{d}B_t, \\ \mathrm{d}R_t &= \gamma I_t\,\mathrm{d}t. \end{aligned}$$

The (stochastic) SIR model is one of the simplest mathematical epidemiology models, and it has also been used as a predictive tool during the Corona pandemic.

Both the general logistic growth model (with nonlinear, possibly complicated f) and the above examples provide nonlinear, complicated systems where we have little hope solving the equation explicitly, like we did in (8.6). Ultimately, the reason why we succeded in the case of (8.4) is because this was a *linear SDE*, and indeed in general linear SDEs can be solved (more or less) explicitly. The same is true for some particular nonlinear SDEs, thanks to some clever tricks, but not in general.

Therefore, we need abstract results guaranteeing that solutions exist and are unique, so that the SDE is *well-posed*.

This is a non-trivial question, already if we consider just the (autonomous) 1-dimensional case and we only focus on ODEs

$$\frac{\mathrm{d}}{\mathrm{d}t}X_t = b(X_t), \quad X|_{t=0} = x_0 \tag{8.7}$$

which can be thought of as a special case of SDEs, in which we have a noise term σB_t appearing but with $\sigma = 0$. As the next example shows, there can be at least three different pathological scenarios we may encounter:

1. Solutions do not exist (*failure of existence*).
2. Solutions exist, but they are not unique (*failure of uniqueness*).
3. Solutions exist only on a finite time interval $[0, T^*)$ and they blow-up as $t \to T^*$ (solutions exist locally but *finite-time blow-up* happens, a more refined scenario amounting to *failure of global existence*).

For ODEs, these case are illustrated by the following classical counterexamples:

Example 8.4. Consider the ODE (8.7) with $b: \mathbb{R} \to \mathbb{R}$.

i. Let $b(x) = \mathbb{1}_{\{0\}}(x)$ and $x_0 = 0$. Then existence fails: indeed, assume that X is a solution. Then X must be increasing ($b \geqslant 0$) and there are two possible cases: either $X_t > 0$ for all $t > 0$, or there exists $\varepsilon > 0$ with $X_t \equiv 0$ for all $t \in [0, \varepsilon]$. In the first case we get the contradiction

$$0 < X_t = 0 + \int_0^t \mathbb{1}_{\{0\}}(X_s)\mathrm{d}s = 0 + \int_0^t 0\mathrm{d}s = 0.$$

In the second case we get the contradiction

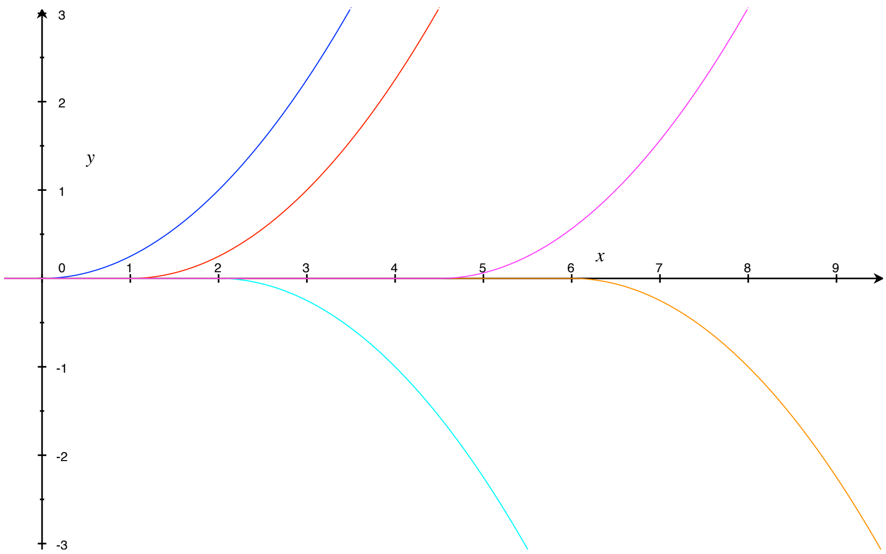
$$0 = X_\varepsilon = 0 + \int_0^\varepsilon \mathbb{1}_{\{0\}}(X_s)\mathrm{d}s = 0 + \int_0^\varepsilon 1\mathrm{d}s = \varepsilon.$$

So the assumption must have been wrong and there cannot exist a solution.

ii. Let $b(x) = \mathrm{sign}(x)\sqrt{|x|}$ and $x_0 = 0$. Then uniqueness fails, because the following functions are all solutions:

$$X_t = \frac{\alpha}{4}(t - t_0)^2 \mathbb{1}_{[t_0, \infty)}(t), \qquad \alpha \in \{-1, 1\}, \qquad t_0 \in [0, \infty].$$

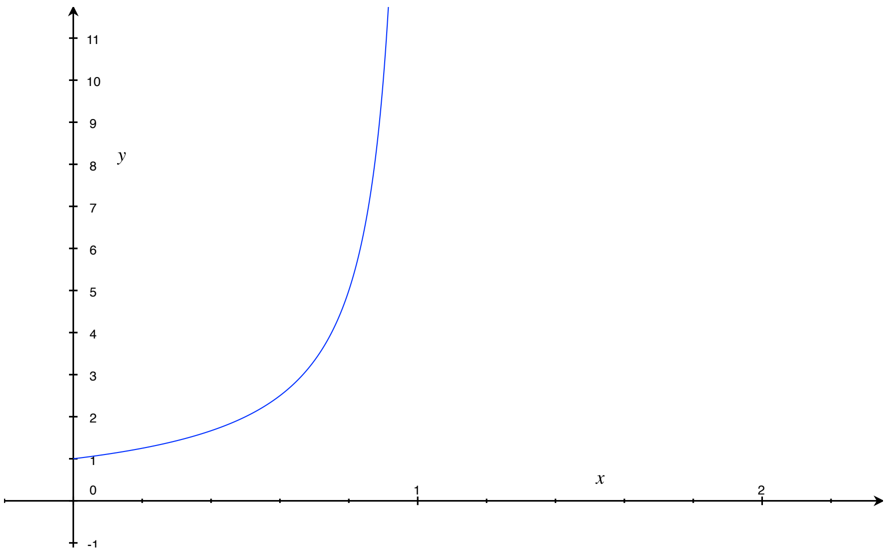
So for this example existence holds, but uniqueness fails.



iii. Let $b(x)=x^2$ and $x_0=1$. Then b is locally Lipschitz continuous and thus we know from the analysis lecture that there is at most one solution (uniqueness holds). However, this unique solution is explicitly given by

$$X_t=\frac{1}{1-t}, \qquad t\in[0,1),$$

and it diverges to $+\infty$ at time 1 (in other words, finite-time blow-up happens). So the solution blows up in finite time and does not exist for all times.



Exercise.

i. Consider the ODE with $b(x)=x^2$ and initial condition x_0. Can you find a range of x_0 for which there is no blow-up?

ii. Let $\beta\in(0,1)$ and $b(x)=\mathrm{sign}(x)|x|^\beta$ and $x_0=0$. Find the (infinitely many) solutions to $\mathrm{d}X_t=b(X_t)\mathrm{d}t$ with initial condition $x_0=0$.

——— **End of the lecture on January 30** ———

8.2 Solution concepts

In the following, we will develop a general existence and uniqueness theory for nonlinear SDEs, under appropriate conditions on the coefficients.

Compared to ODE theory, the situation is more subtle, as there are multiple solution concepts available. The reason for this will become clear as we go on developing the theory and seeing interesting examples where it applies.

For simplicity, whenever discussing SDEs, we will let the *coefficients* b and σ to be Borel measurable functions

$$b\colon\mathbb{R}_+\times\mathbb{R}^d\to\mathbb{R}^d,\quad b(t,x)=(b_i(t,x))_{i\in\{1,\ldots,d\}},$$
$$\sigma\colon\mathbb{R}_+\times\mathbb{R}^d\to\mathbb{R}^{d\times m},\quad \sigma(t,x)=(\sigma_{ij}(t,x))_{i\in\{1,\ldots,d\},j\in\{1,\ldots,m\}}$$

which are assumed to be *locally bounded*, namely such that

$$\sup_{t\in[0,T],|x|\leqslant R}|b(t,x)|+\sup_{t\in[0,T],|x|\leqslant R}|\sigma(t,x)|<\infty \quad \text{for all } T,R\in(0,+\infty)$$

We call b the *drift coefficient* and σ the *diffusion coefficient* of the d-dimensional *stochastic differential equation* (SDE):

$$\mathrm{d}X_t=b(t,X_t)\mathrm{d}\,t+\sigma(t,X_t)\mathrm{d}B_t \tag{8.8}$$

where B is a $\mathbb{R}^m$-valued Brownian motion (so that, since $\sigma(t,X_t)\in\mathbb{R}^{d\times m}$, at least formally, $\sigma(t,X_t)\mathrm{d}B_t\in\mathbb{R}^d$). Sometimes it is more convenient to think of $\sigma_j(t,x)=(\sigma_{ij}(t,x))_{i=1}^d$ to be a velocity field (namely, $\sigma_j\colon\mathbb{R}_+\times\mathbb{R}^d\to\mathbb{R}^d$) "attached" to the one-dimensional Brownian motion B^j, for each $j\in\{1,\dots,m\}$. As a consequence, (8.8) may also be written as

$$\mathrm{d}X_t=b(t,X_t)\mathrm{d}\,t+\sum_{j=1}^m\sigma_j(t,X_t)\mathrm{d}B_t^j. \tag{8.9}$$

In either case, the SDE must be understood in integral form and componentwise, see (8.10) below.

Definition 8.5. *Let b and σ be measurable locally bounded functions as above. A tuple $(\Omega,\mathcal{F},\mathbb{F},\mathbb{P},\xi,B,X)$ is a* weak solution *to the SDE (8.8) if the following hold:*

i. $(\Omega,\mathcal{F},\mathbb{F},\mathbb{P})$ *is a filtered probability space satisfying the usual conditions;*

ii. B *is a d-dimensional $\mathbb{F}$-Brownian motion, ξ be an $\mathcal{F}_0$–measurable $\mathbb{R}^d$–valued random variable and $X=(X_t)_{t\geqslant 0}$ is a d-dimensional continuous $\mathbb{F}$-adapted process;*

iii. $\mathbb{P}$*-almost surely, for all $t\geq 0$ and $i\in\{1,\dots,d\}$ it holds that*

$$X_t^i=\xi+\int_0^t b_i(s,X_s)\mathrm{d}s+\sum_{j=1}^m\int_0^t\sigma_{ij}(s,X_s)\mathrm{d}B_s^j. \tag{8.10}$$

Remark 8.6. Note the the integrals appearing in (8.10) are well-defined. Indeed, since X is continuous and adapted, it is progressive, and therefore so are $t\mapsto b(t,X_t)$, $t\mapsto\sigma(t,X_t)$; since X is continuous and b, σ are locally bounded, then $b(t,X_t)$, $\sigma(t,X_t)$ are locally bounded processes (in the sense of Remark 6.18). Therefore $\int_0^t b_i(s,X_s)\mathrm{d}s$ is meaningful as a Lebesgue integral and $\int_0^t\sigma_{ij}(s,X_s)\mathrm{d}B_s^j$ as a stochastic integral.

Extra comment about the literature, not examinable: The assumption that b and σ are locally bounded is sometimes too restrictive, as in relevant application the coefficients might explode around some critical point, e.g. like $\frac{1}{|x|}$ or $\frac{1}{\sqrt{t}}$. In these cases, one can still define what it means to be a solution to the SDE (8.8), up to additionally enforcing that

$$\mathbb{P}\bigg(\int_0^T[|b(s,X_s)|+|\sigma(s,X_s)|^2]\,\mathrm{d}s<\infty\quad\forall T<\infty\bigg)=1$$

in order to guarantee that all integrals appearing make sense, as either Lebesgue integrals or stochastic integrals.

Remark 8.7. It follows from (8.10) that $\mathbb{P}$-a.s. $X_0=\xi$. ξ is called the *initial condition* of the SDE. Notice that, since B is $\mathbb{F}$-Brownian, by Blumenthal's 0-1 law it is independent of ξ; therefore prescribing the law of ξ is the same as prescribing the joint law of (ξ,B).

In the framework of Definition 8.5, for simplicity one sometimes drops the underlying $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, B)$ and simply says that X is a weak solution of the SDE (8.8) started at ξ. B can be referred to as the *driving noise*, or *driver*, and X is a solution *driven by* B.

Notice however that in Definition 8.5, the underlying probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ and the Brownian motion B are part of the solution itself. This is because, once we are given a filtered probability space and a Brownian motion on it, we have a very rich structure allowing to generate many other Brownian motions (think of time-reversal, reflection principle, Lévy's characterization, time-change) as well as to change the reference probability $\mathbb{P}$ (think of Girsanov). So it is not fully clear that there is a "canonical choice" of $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, B)$ and Definition 8.5 gives us the additional freedom to choose the most convenient one.

At the same time, whenever for instance trying to numerically simulate solutions as we did in the Malthusian growth case, it is clear that we would like to first simulate a Brownian motion B, and then construct the solution X (or its numerical approximation) starting from it. In other words, we would like to specify an order: we first fix the reference setting $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, B)$, and then find a solution on it. This is accomplished by the following solution concept.

Definition 8.8. *We say that X is a* strong solution *if it is a weak solution which is adapted to the (standard augmentation of the) filtration generated by B and ξ.*

Remark 8.9. If $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, B, \xi, X)$ is a weak solution and we let $\mathbb{G}$ be the (standard augmentation of the) filtration generated by X, B and ξ, then it's easy to check that B is a $\mathbb{G}$-Brownian motion, ξ is $\mathcal{G}_0$-adapted and so $(\Omega, \mathcal{F}, \mathbb{G}, \mathbb{P}, B, \xi, X)$ is still a weak solution.

The idea encoded in the definition of strong solution is instead that (up to technical measure theory details, coming from the completion of the measure) X can be expressed as $X = F(\xi, B)$, where $F: \mathbb{R}^d \times C(\mathbb{R}_+; \mathbb{R}^d) \to C(\mathbb{R}_+; \mathbb{R}^d)$ is a measurable function.

This was indeed the case in the SDE (8.4) where we got an explicit solution formula (8.6) corresponding to

$$F(\xi, B)_t = \xi \exp\left(rt + \sigma B_t - \frac{1}{2}\sigma^2 t\right) \quad \forall t \geqslant 0.$$

Having defined what it means to be a solution to the SDE, we want a uniqueness concept as well. Notice that uniqueness should be a property of the equation (8.8), not of the solution (8.10).

Definition 8.10. *We say that* pathwise uniqueness *holds for the SDE (8.8) if for any tuple $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, B, \xi)$ as in Definition 8.5, and any two solutions X and Y solving the SDE wiwht respect to this tuple, it holds*

$$\mathbb{P}(X_t = Y_t \quad \textit{for all } t \geqslant 0) = 1.$$

The concept of "pathwise uniqueness" should be naturally paired with that of "strong solution". We will see later that there is another notion of uniqueness ("uniqueness in law") which is more naturally paired with the larger class of weak solutions.

8.3 Existence and uniqueness under global Lipschitz conditions

In analysis lectures it is typically shown that ODEs with Lipschitz continuous coefficients have unique solutions (the result has several names, typically either *Cauchy–Lipschitz* or *Picard–Lindelöf* theorem). Here we extend this result to the stochastic setting.

Convention. In the following, we will consider norms of the matrix $\sigma \in \mathbb{R}^{d\times m}$. Since $\mathbb{R}^{d\times m}$ is finite dimensional, any two norms are equivalent, and we may just think of σ as a "huge vector" with dm entries. It will be convenient to work with the *Frobenius norm*, so we always take

$$|\sigma| := \left(\sum_{i=1}^{d} \sum_{j=1}^{m} \sigma_{ij}^2 \right)^{1/2} = (\mathrm{trace}(\sigma\sigma^*))^{1/2}.$$

Exercise. Show that, for any $\sigma \in \mathbb{R}^{d\times m}$ and any $v \in \mathbb{R}^m$, $|\sigma v| \leqslant |\sigma|\,|v|$.

Theorem 8.11. (Strong wellposedness for globally Lipschitz coefficients) *Assume that there exists a constant $K > 0$ such that b and σ satisfy the following:*

$$|b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leqslant K|x-y| \qquad \forall t \geqslant 0, x, y \in \mathbb{R}^d, \tag{8.11}$$

$$|b(t,x)| + |\sigma(t,x)| \leqslant K(1+|x|) \qquad \forall t \geqslant 0, x \in \mathbb{R}^d. \tag{8.12}$$

Then pathwise uniqueness holds for the SDE (8.8). Moreover, for any $\mathcal{F}_0$–measurable initial condition ξ, there exists a strong solution X.

Remark 8.12. Condition (8.11) means that b and σ are *globally Lipschitz* in x with constant K, uniformly in $t \geqslant 0$. Condition (8.12) means that b and σ have *at most linear growth* in x (indeed as $|x| \to \infty$, $|x| \sim 1 + |x|$) uniformly in $t \geqslant 0$. Notice that in practice (8.11) almost implies (8.12), since by triangular inequality

$$|b(t,x)| \leqslant |b(t,0)| + |b(t,x) - b(t,0)| \leqslant \tilde{K}(1+|x|)$$

for $\tilde{K} := \max\{|b(t,0)|, K\}$, similarly for σ.

Exercise. We will often need to exchange sums and powers. Show that, up to a constant, this is allowed: If $x_1, \ldots, x_n \geqslant 0$ and $p \geqslant 1$, then

$$(x_1 + \cdots + x_n)^p \leqslant n^{p-1}(x_1^p + \cdots + x_n^p).$$

We will split the proof in several steps. We start by recalling a fundamental tool from ODE theory.

Lemma 8.13. (Grönwall's lemma) *Let $T \in (0, +\infty]$ and let $f\colon [0,T] \to \mathbb{R}$ be a bounded and measurable function. Assume that for some $\alpha, \beta \geqslant 0$ we have*

$$f(t) \leqslant \alpha + \beta \int_0^t f(s)\,\mathrm{d}s \qquad \forall t \in [0,T].$$

Then

$$f(t) \leqslant \alpha e^{\beta t} \qquad \forall t \in [0,T].$$

Proof.
Proof was skipped in the lectures for the sake of time; the result is standard and I hope you have already seen it before in analysis courses. In case you haven't, here is included for completeness a proof (among the numerous existing ones).

Iterating the assumption we get

$$\begin{aligned}
f(t) \leqslant & \alpha + \beta \int_0^t \left(\alpha + \beta \int_0^{s_1} f(s_2) \mathrm{d}s_2 \right) \mathrm{d}s_1 \\
\leqslant & \alpha + \alpha(\beta t) + \beta^2 \int_0^t \int_0^{s_1} \left(\alpha + \beta \int_0^{s_2} f(s_3) \mathrm{d}s_3 \right) \mathrm{d}s_2 \mathrm{d}s_1 \\
= & \alpha + \alpha(\beta t) + \alpha \frac{(\beta t)^2}{2} + \beta^3 \int_0^t \int_0^{s_1} \int_0^{s_2} f(s_3) \mathrm{d}s_3 \mathrm{d}s_2 \mathrm{d}s_1 \\
\leqslant & \cdots \leqslant \alpha + \alpha(\beta t) + \alpha \frac{(\beta t)^2}{2} + \cdots + \alpha \frac{(\beta t)^{n-1}}{(n-1)!} + \beta^n \int_0^t \int_0^{s_1} \ldots \int_0^{s_{n-1}} f(s_n) \mathrm{d}s_n \cdots \mathrm{d}s_1.
\end{aligned}$$

The first n terms on the right hand side are the beginning of the power series $\alpha e^{\beta t}$, so if the integral term vanishes for $n \to \infty$, then the proof is complete. But if $C \geqslant 0$ is such that $|f(t)| \leqslant C$ for all $t \in [0, T]$, then

$$\left| \beta^n \int_0^t \int_0^{s_1} \ldots \int_0^{s_{n-1}} f(s_n) \mathrm{d}s_n \cdots \mathrm{d}s_1 \right| \leqslant \beta^n \int_0^t \int_0^{s_1} \ldots \int_0^{s_{n-1}} C \, \mathrm{d}s_n \cdots \mathrm{d}s_1 = C \frac{\beta^n t^n}{n!},$$

which indeed converges to zero as $n \to \infty$. □

Proposition 8.14. (Pathwise uniqueness for locally monotone coefficients) *Assume that the coefficients b and σ are locally bounded and satisfy the following* local monotonicity assumption*: for every $T \in (0, +\infty)$ and $n \in \mathbb{N}$ there exists a constant $K_{T,n} > 0$ such that*

$$2[b(t,x) - b(t,y)] \cdot (x - y) + |\sigma(t,x) - \sigma(t,y)|^2 \leqslant K_{T,n} |x - y|^2 \tag{8.13}$$

for all $t \in [0, T]$ and $|x|, |y| \leqslant n$. Then pathwise uniqueness holds for the SDE (8.8).

Exercise. Check that if b and σ satisfy the global Lipschitz condition (8.11), then (8.13) holds. More generally, a sufficient condition guaranteeing the validity of (8.13) is to require that b and σ are *locally Lipschitz continuous*, uniformly on finite times, namely to require that for every $T \in (0, +\infty)$ and $n \in \mathbb{N}$ there exists a constant $\tilde{K}_{T,n} > 0$ such that

$$|b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leqslant \tilde{K}_{T,n} |x - y| \tag{8.14}$$

for all $t \in [0, T]$ and $|x|, |y| \leqslant n$.

——— End of the lecture on February 5th ———

Proof. Let X and Y be two solutions, defined on the same probability space, driven by the same B and with same initial datum ξ. For any $n \in \mathbb{N}$, define the stopping time

$$\tau_n = \inf \{t \geqslant 0 \colon |X_t| \vee |Y_t| \geqslant n\}.$$

Then $\mathbb{E}[|X_t^{\tau_n} - Y_t^{\tau_n}|^2] \leqslant 4n^2 < \infty$ for all $t \geqslant 0$. Our goal is to control $|X_t^{\tau_n} - Y_t^{\tau_n}|^2$ using Itô's formula. To this end, setting $Z_t := X_t^{\tau_n} - Y_t^{\tau_n}$, notice that

$$Z_t^i = \int_0^t [b_i(s, X_s) - b_i(s, Y_s)] \, \mathbb{1}_{[0,\tau_n]}(s) \, \mathrm{d}s + \sum_{k=1}^m \int_0^t [\sigma_{ik}(X_s) - \sigma_{ik}(Y_s)] \, \mathbb{1}_{[0,\tau_n]}(s) \, \mathrm{d}B_s^k$$

so that

$$\langle Z^i, Z^i \rangle_t = \sum_{k=1}^m \int_0^t |\sigma_{ik}(X_s) - \sigma_{ik}(Y_s)|^2 \, \mathbb{1}_{[0,\tau_n]}(s) \, \mathrm{d}s.$$

Applying Itô's formula to $|Z_t|^2$ (notice that $f(x)=|x|^2$ is such that $\nabla f(x)=2x$, $D^2 f(x)=2I_d$) we then find

$$\begin{aligned}
\mathrm{d}|X_t^{\tau_n}-Y_t^{\tau_n}|^2 =& \mathrm{d}|Z_t|^2 = 2\sum_{i=1}^d Z_t^i\cdot \mathrm{d}Z_t^i + \sum_{i=1}^d \mathrm{d}\langle Z^i, Z^i\rangle_t\\
=& 2\sum_{i=1}^d [b_i(s,X_s)-b_i(s,Y_s)](X_s^i-Y_s^i)\mathbb{1}_{[0,\tau_n]}(s)\mathrm{d}s\\
&+2\sum_{i=1}^d\sum_{k=1}^m [b_i(s,X_s)-b_i(s,Y_s)][\sigma_{ik}(s,X_s)-\sigma_{ik}(s,Y_s)]\,\mathbb{1}_{[0,\tau_n]}(s)\,\mathrm{d}B_s^k\\
&+\sum_{i=1}^d\sum_{k=1}^m\int_0^t |\sigma_{ik}(s,X_s)-\sigma_{ik}(s,Y_s)|^2\,\mathbb{1}_{[0,\tau_n]}(s)\,\mathrm{d}s.\\
=& (2[b(s,X_s)-b(s,Y_s)]\cdot(X_s-Y_s)+|\sigma(s,X_s)-\sigma(s,Y_s)|^2)\mathbb{1}_{[0,\tau_n]}(s)\mathrm{d}s\\
&+2\sum_{i=1}^d\sum_{k=1}^m [b_i(s,X_s)-b_i(s,Y_s)][\sigma_{ik}(s,X_s)-\sigma_{ik}(s,Y_s)]\,\mathbb{1}_{[0,\tau_n]}(s)\,\mathrm{d}B_s^k.
\end{aligned}$$

In the above computation, we used several times the fact that we can write indifferently X_s or $X_s^{\tau_n}$ as long as we are multiplying by $\mathbb{1}_{[0,\tau_n]}(s)$. Since b and σ are locally bounded coefficients and $|X_s|\mathbb{1}_{[0,\tau_n]}\leqslant n$, similarly for Y, the stochastic integrals appearing at the end of the computation are genuine martingale (the integrands are bounded by deterministic constants). Therefore taking expectation (in integral form) we find

$$\begin{aligned}
\mathbb{E}[|X_t^{\tau_n}-Y_t^{\tau_n}|^2] =& \mathbb{E}\left[\int_0^t 2([b(s,X_s)-b(s,Y_s)]\cdot(X_s-Y_s)+|\sigma(s,X_s)-\sigma(s,Y_s)|^2)\,\mathbb{1}_{[0,\tau_n]}(s)\mathrm{d}s\right]\\
\leqslant& \mathbb{E}\left[\int_0^t K_{T,n}\,|X_s-Y_s|^2\,\mathbb{1}_{[0,\tau_n]}(s)\mathrm{d}s\right]\\
\leqslant& K_{T,n}\int_0^t \mathbb{E}[|X_s^{\tau_n}-Y_s^{\tau_n}|^2]\,\mathrm{d}s
\end{aligned}$$

where in the second passage we applied assumption (8.13) and in the last one Fubini's theorem. As the argument holds for any $t\in[0,T]$, applying Grönwall's lemma (for $f(t)=\mathbb{E}[|X_t^{\tau_n}-Y_t^{\tau_n}|^2]$, $\alpha=0$, $\beta=K_{T,n}$) we conclude that

$$\mathbb{E}[|X_t^{\tau_n}-Y_t^{\tau_n}|^2]=0 \qquad \forall t\geqslant 0, n\in\mathbb{N}.$$

Since X^{τ_n} and Y^{τ_n} are continuous, they must be indistinguishable. Now we let $n\to\infty$ to conclude that X and Y are indistinguishable as well. □

Next, we want to construct a strong solution by a *fixed point procedure*. To this end, recall (a version of) Banach's fixed point theorem:

Banach's fixed point theorem: Let $(E,\|\cdot\|_E)$ be a Banach space and let $\mathcal{I}\colon E\to E$ be a *contraction*, namely a function such that

$$\|\mathcal{I}(x)-\mathcal{I}(y)\|_E\leqslant\theta\,\|x-y\|_E \quad \forall x,y\in E$$

for some $\theta\in(0,1)$. Then $\mathcal{I}$ admits a unique fixed point $\bar{x}\in E$, namely there exists a unique solution $\bar{x}\in E$ to the equation $\mathcal{I}(\bar{x})=\bar{x}$.

In order to apply Banach's theorem, we want to rephrase the SDE problem (8.8) as an appropriate fixed point equation. Suppose we are already given a filtered space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ on which (B, ξ) are defined, and let $\mathbb{G}$ denote the (augmented) filtration generated by ξ, B. For a parameter $\lambda \geqslant 0$ to be chosen later, let us consider the space of processes

$$E^\lambda = \{H = (H_t)_{t \geqslant 0} \text{ continuous, } \mathbb{G}\text{-adapted, } \mathbb{R}^d\text{-valued such that } \|H\|_{E^\lambda} < \infty\}$$

where

$$\|H\|_{E^\lambda} := \sup_{t \geqslant 0} e^{-\lambda t} \mathbb{E}\bigg[\sup_{s \in [0,t]} |H_s|^2\bigg]^{\frac{1}{2}}.$$

Exercise. Show that, for any $\lambda \geqslant 0$, E^λ is a Banach space.

In the proof of the next result, Minkowski's integral inequality will be useful.

Minkowski's integral inequality (special case): Let $Z = Z(t, \omega)$ be a measurable nonnegative process; then for any $p \in [1, \infty]$ and any $t \in [0, +\infty]$ it holds that

$$\left\| \int_0^t Z_s \,\mathrm{d}s \right\|_{L^p(\Omega)} \leqslant \int_0^t \|Z_s\|_{L^p(\Omega)} \,\mathrm{d}s. \tag{8.15}$$

With these preparations, we can now show strong existence of solutions.

Proposition 8.15. *Let b, σ satisfy the assumptions of Theorem 8.11. Then for any $\mathcal{F}_0$-measurable $\xi \in L^2(\Omega)$, there exists a strong solution X to the SDE (8.8).*

Proof. Given a continuous, $\mathbb{G}$-adapted $\mathbb{R}^d$-valued process H, let us define another $\mathbb{G}$-adapted, continuous $\mathbb{R}^d$-valued process $\mathcal{I}(H)$ by

$$\mathcal{I}(H) = \xi + \int_0^\cdot b(s, H_s)\mathrm{d}s + \int_0^\cdot \sigma(s, H_s)\mathrm{d}B_s.$$

Our goal is to show that, for appropriately chosen λ, $\mathcal{I}$ as defined above maps E^λ into itself and is in fact a contraction. Once we have done this, by Banach's fixed point theorem, we will be able to conclude that there exists a process $X \in E^\lambda$ such that $X = \mathcal{I}(X)$, which is then by definition a strong solution to the SDE (recall that $\mathbb{G}$ is the (augmented) filtration generated by ξ and B).

We omit the verification that $\mathcal{I}$ maps E^λ into itself, as it is almost identical to the computations that we will present below to show its contractivity, up to roughly using assumption (8.12) in the passages where we will invoke (8.11) instead. Let H^1, $H^2 \in E^\lambda$, then

$$\mathcal{I}(H^1)_r - \mathcal{I}(H^2)_r = \int_0^r [b(s, H_s^1) - b(s, H_s^2)]\mathrm{d}s + \int_0^r [\sigma(s, H_s^1) - \sigma(s, H_s^2)]\mathrm{d}B_s$$

so that

$$\sup_{r \in [0,t]} |\mathcal{I}(H^1)_r - \mathcal{I}(H^2)_r| \leqslant \int_0^t |b(s, H_s^1) - b(s, H_s^2)|\mathrm{d}s + \sup_{r \in [0,t]} \left| \int_0^r [\sigma(s, H_s^1) - \sigma(s, H_s^2)]\mathrm{d}B_s \right|.$$

Taking the $L^2(\Omega)$-norm on both sides, using first standard Minkowki's inequality and then its integral-in-time version (8.15) (for $p = 2$) we find

$$\begin{aligned}\left\| \sup_{r \in [0,t]} |\mathcal{I}(H^1)_r - \mathcal{I}(H^2)_r| \right\|_{L^2(\Omega)} &\leqslant \int_0^t \|b(s, H_s^1) - b(s, H_s^2)\|_{L^2(\Omega)} \,\mathrm{d}s \\ &\quad + \left\| \sup_{r \in [0,t]} \left| \int_0^r [\sigma(s, H_s^1) - \sigma(s, H_s^2)]\mathrm{d}B_s \right| \right\|_{L^2(\Omega)} \\ &=: I_t^1 + I_t^2.\end{aligned}$$

On the first term, we may apply assumption (8.11) and the definition of $\|H^1 - H^2\|_{E^\lambda}$ to find

$$I_t^1 \leqslant K\int_0^t \|H_s^1 - H_s^2\|_{L^2(\Omega)}\,\mathrm{d}s \leqslant K\,\|H^1 - H^2\|_{E^\lambda}\int_0^t e^{\lambda s}\,\mathrm{d}s = \frac{K}{\lambda}\,e^{\lambda t}\|H^1 - H^2\|_{E^\lambda}.$$

On the second term, we can use Doob's martingale inequality, and then compute the quadratic variation of the stochastic integral (writing everything componentwise, similarly to Proposition 8.14) to find

$$\begin{aligned} I_t^2 \leqslant & C\,\mathbb{E}\left[\left|\int_0^t [\sigma(s,H_s^1) - \sigma(s,H_s^2)]\mathrm{d}B_s\right|^2\right]^{1/2}\\ = & C\mathbb{E}\left[\int_0^t |\sigma(s,H_s^1) - \sigma(s,H_s^2)|^2 \mathrm{d}s\right]^{1/2}\\ \leqslant & CK\left(\int_0^t \mathbb{E}[|H_s^1 - H_s^2|^2]\mathrm{d}s\right)^{1/2}\\ \leqslant & CK\,\|H^1 - H^2\|_{E^\lambda}\left(\int_0^t e^{2\lambda s}\,\mathrm{d}s\right)^{1/2} = \frac{CK}{\sqrt{2\lambda}}\,e^{\lambda t}\,\|H^1 - H^2\|_{E^\lambda} \end{aligned}$$

where again we used assumption (8.11) and the definition of $\|H^1 - H^2\|_{E^\lambda}$. Overall we get

$$\left\|\sup_{r\in[0,t]} |\mathcal{I}(H^1)_r - \mathcal{I}(H^2)_r|\right\|_{L^2(\Omega)} \leqslant \left(\frac{K}{\lambda} + \frac{CK}{\sqrt{2\lambda}}\right) e^{\lambda t}\,\|H^1 - H^2\|_{E^\lambda};$$

multiplying both sides by $e^{-\lambda t}$ and taking supremum over $t \geqslant 0$ we obtain

$$\|\mathcal{I}(H^1) - \mathcal{I}(H^2)\|_{E^\lambda} \leqslant \left(\frac{K}{\lambda} + \frac{CK}{\sqrt{2\lambda}}\right)\|H^1 - H^2\|_{E^\lambda}.$$

Therefore choosing $\lambda > 0$ large enough so that

$$\frac{K}{\lambda} + \frac{CK}{\sqrt{2\lambda}} < 1$$

we can conclude that $\mathcal{I}$ is a contraction on E^λ. □

We can finally complete the

Proof of Theorem 8.11. Pathwise uniqueness follows from Proposition 8.14 (as the global Lipschitz condition (8.11) implies the validity of (8.13)); Proposition 8.15 implies strong existence for $\xi \in L^2(\Omega)$, so it only remains to relax this last assumption to allow for any $\mathcal{F}_0$-measurable, possibly non-integrable ξ.

Given such ξ, thanks to Proposition 8.15, for all $n \in \mathbb{N}$ there is a pathwise unique strong solution X^n to

$$\mathrm{d}X_t^n = b(t, X_t^n)\mathrm{d}\,t + \sigma(t, X_t^n)\mathrm{d}B_t, \qquad X_0^n = \xi\mathbb{1}_{|\xi|\leqslant n}. \tag{8.16}$$

We claim that the sequence $\{X^n\}_n$ is a Cauchy sequence in the ucp convergence; if that is the case, then there must exist a continuous process X such that $X^n \to X$ in ucp, and thanks to the Lipschitz continuity of b and σ and stability of stochastic integrals under ucp convergence (Corollary 6.22) we can conclude that X solves the SDE with $X_0 = \xi$. Moreover X is $\mathbb{G}$-adapted, since it is the limit of $\mathbb{G}$-adapted processes, thus a strong solution.

Let X^n solve (8.16), then since ξ is $\mathcal{G}_0$-measurable we can multiply both sides by $\mathbb{1}_{|\xi|\leqslant n}$ and bring it inside the integrals to find

$$\mathrm{d}(X_t^n\mathbb{1}_{|\xi|\leqslant n}) = b(t, X_t^n\mathbb{1}_{|\xi|\leqslant n})\mathbb{1}_{|\xi|\leqslant n}\mathrm{d}\,t + \sigma(t, X_t^n\mathbb{1}_{|\xi|\leqslant n})\mathbb{1}_{|\xi|\leqslant n}\mathrm{d}B_t, \quad X_0^n\mathbb{1}_{|\xi|\leqslant n} = \mathbb{1}_{|\xi|\leqslant n};$$

Here we used two facts: a) if $|\xi|>n$, changing the values inside b and σ doesn't change anything, since they get multiplied by 0 anyway; b) since ξ is $\mathcal{G}_0$-measurable, so is $\mathbb{1}_{|\xi|\leqslant n}$, and so we can move it inside and outside stochastic integrals as we please (see the Exercise below). We can do the same operation for X^m with $m\geqslant n$, in which case since $\mathbb{1}_{|\xi|\leqslant n}\mathbb{1}_{|\xi|\leqslant m}=\mathbb{1}_{|\xi|\leqslant n}$ we find

$$\mathrm{d}(X_t^m\mathbb{1}_{|\xi|\leqslant n})=b(t,X_t^m\mathbb{1}_{|\xi|\leqslant n})\mathbb{1}_{|\xi|\leqslant n}\mathrm{d}t+\sigma(t,X_t^m\mathbb{1}_{|\xi|\leqslant n})\mathbb{1}_{|\xi|\leqslant n}\mathrm{d}B_t,\quad X_0^m\mathbb{1}_{|\xi|\leqslant n}=\mathbb{1}_{|\xi|\leqslant n}.$$

Roughly speaking, $X_t^n\mathbb{1}_{|\xi|\leqslant n}$ and $X_t^m\mathbb{1}_{|\xi|\leqslant n}$ solve the same SDE; indeed going through the same identical argument as in the proof of Proposition 8.14, one can show that they are indistinguishable. As a consequence it must hold

$$\mathbb{P}(X^n\neq X^m)\leqslant\mathbb{P}(|\xi|>n)\to 0\quad\text{as } n\to\infty$$

proving the desided claim and thus concluding the proof. ☐

Exercise. Let $s<t$ fixed, $(\Omega,\mathcal{F},\mathbb{F},\mathbb{P})$ given. Let $M\in\mathcal{M}_{\text{loc}}^c$, $H\in L_{\text{loc}}^2(M)$ and Z be a $\mathcal{F}_s$-measurable random variable. Then ($\mathbb{P}$-a.s.)

$$\int_s^t ZH_r\,\mathrm{d}M_r=Z\int_s^t H_r\,\mathrm{d}M_r$$

—— End of the lecture on February 6th ——

Example 8.16. (Ornstein-Uhlenbeck process) Let $d=m=1$, $a,\sigma>0$ and $\mu\in\mathbb{R}$ and consider the SDE

$$\mathrm{d}X_t=a(\mu-X_t)\mathrm{d}t+\sigma\mathrm{d}B_t,\qquad X_0=x. \tag{8.17}$$

Clearly the assumptions of Theorem 8.11 are satisfied, so there is a pathwise unique strong solution. If $X_t<\mu$, then the drift is positive and X has a tendency to increase. If $X_t>\mu$, then the drift is negative and X has a tendency to decrease. Such behavior is called *mean-reverting* (the mean being μ), and the Ornstein-Uhlenbeck process models a random process that fluctuates around its mean μ.

Let us derive the solution to (8.17) explicitly by the *variation of constants* method: the homogeneous part of the equation is given by

$$\mathrm{d}Y_t=-aY_t\,\mathrm{d}t$$

which is solved by $Y_t=ce^{-at}$ for $c\in\mathbb{R}$. We take $c=1$ look for a solution X which "looks like Y up to a multiplicative perturbation"; namely we make the *solution ansatz*

$$X_t=Y_t\,Z_t=e^{-at}\,Z_t.$$

By the integration by parts formula, Z then satisfies

$$\begin{aligned}\mathrm{d}Z_t&=\mathrm{d}(e^{at}X_t)=a\,e^{at}X_t\,\mathrm{d}t+e^{at}\,\mathrm{d}X_t\\&=e^{at}(aX_t\,\mathrm{d}t+a(\mu-X_t)\mathrm{d}t+\sigma\mathrm{d}B_t)\\&=\mu a e^{at}\mathrm{d}t+\sigma e^{at}\mathrm{d}B_t.\end{aligned}$$

Integrating in time (noting that $ae^{at}=(e^{at})'$) we find

$$e^{at}X_t=Z_t=Z_0+\mu(e^{at}-1)+\int_0^t\sigma e^{as}\mathrm{d}B_s$$

which finally yields ($X_0=Z_0=x$)

$$X_t=e^{-at}x+\mu(1-e^{-at})+\int_0^t e^{-a(t-s)}\sigma\mathrm{d}B_s. \tag{8.18}$$

Given the explicit formula (8.18) one can then verify that X is indeed a solution to (8.17), thus the unique one. The expected value of X_t is

$$\mathbb{E}[X_t] = x e^{-at} + \mu(1 - e^{-at}),$$

which converges to μ as $t \to \infty$ (recall that $a > 0$). Note also that X_t is a Gaussian random variable (since the integrand in the stochastic integral is deterministic), and its variance is given by

$$\mathrm{var}(X_t) = \mathbb{E}\left[\left(\int_0^t e^{-a(t-s)}\sigma \mathrm{d}B_s\right)^2\right] = \int_0^t e^{-2a(t-s)}\sigma^2 \mathrm{d}s = \frac{\sigma^2}{2a}(1 - e^{-2at}),$$

which converges to $\sigma^2/2a$. So as $t \to \infty$, the law of X_t converges to $\mathcal{N}\left(\mu, \frac{\sigma^2}{2a}\right)$ (in fact, if we had more time to discuss these topic, one could show that $\mathcal{N}\left(\mu, \frac{\sigma^2}{2a}\right)$ is the unique *invariant measure* of the SDE (8.17)).

Exercise. What happens for $a < 0$ as $t \to \infty$?

Due to lack of time, the next Examples 8.17-8.18 were omitted in the lectures and are not examinable. I leave them here for those interested.

Example 8.17. (ODE in Brownian time) If $\sigma = 0$, i.e. we are considering an ODE, then the same arguments used in Theorem 8.11 show that for globally Lipschitz $f \in C(\mathbb{R}^d, \mathbb{R}^d)$ and $x_0 \in \mathbb{R}^d$ we can find a unique solution $x \in C^1(\mathbb{R}, \mathbb{R}^d)$ to the equation

$$\frac{\mathrm{d}}{\mathrm{d}t} x(t) = f(x(t)) \quad \forall t \in \mathbb{R}, \qquad x(0) = x_0.$$

(The difference is that now we are solving the equation both forward and *backward* in time, i.e. we allow negative t). Let us write $\varphi(t) := x(t)$, so that $\varphi \in C(\mathbb{R}, \mathbb{R}^d)$. If $f \in C^1(\mathbb{R}^d, \mathbb{R}^d)$, then

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\varphi(t) = \frac{\mathrm{d}}{\mathrm{d}t} f(\varphi(t)) = \nabla f(\varphi(t)) \frac{\mathrm{d}}{\mathrm{d}t}\varphi(t) = \nabla f(\varphi(t))\, f(\varphi(t)) \tag{8.19}$$

where $(\nabla f(x) f(x))_i = \sum_{j=1}^d \partial_j f_i(x)\, f_j(x)$ for $i = 1, \ldots, d$); therefore $\varphi \in C^2(\mathbb{R}, \mathbb{R}^d)$. Iterating, it's easy to check that if $f \in C^2(\mathbb{R}^d, \mathbb{R}^d)$, then $\varphi \in C^3(\mathbb{R}, \mathbb{R}^d)$. Consider now the stochastic process $X_t := \varphi(B_t)$ for a one-dimensional Brownian motion B, i.e. we are considering a "Brownian random time change" (notice however that $t \mapsto B_t$ is by no means injective). By the Stratonovich chain rule (Corollary 7.7), X solves

$$\mathrm{d}X_t = \frac{\mathrm{d}\varphi}{\mathrm{d}t}(B_t) \circ \mathrm{d}B_t = f(\varphi(B_t)) \circ \mathrm{d}B_t = b(X_t) \circ \mathrm{d}B_t \tag{8.20}$$

which is a *Stratonovich SDE*. By instead applying Itô formula, using the formula for $\frac{\mathrm{d}^2}{\mathrm{d}t^2}\varphi$, coming from (8.19), we get the equivalent *Itô form* of SDE (8.20):

$$\mathrm{d}X_t = f(X_t)\mathrm{d}B_t + \frac{1}{2}(\nabla f\, f)(X_t)\mathrm{d}t. \tag{8.21}$$

The SDE (8.21) could have been obtained directly from (8.20), using the rules to convert Stratonovich integrals into Itô integrals (Definition 6.28) and recursively applying the structure of the SDE (8.20) itself.

The conditions $f \in C^2(\mathbb{R}^d, \mathbb{R})$ and global Lipschitz continuity of f are actually sufficient to guarantee the pathwise uniqueness of X, because in this case the drift $\frac{1}{2}\nabla f f$ in the Itô form (8.21) is locally Lipschitz continuous and so Proposition 8.14 applies. Therefore, $X_t = \varphi(B_t)$ is the *only* solution to our SDE (8.21), as well as to (8.20).

With some more tricks of this type ($\rightsquigarrow$*Doss-Sussmann transformation*), given any 1-dimensional Brownian motion B, we can write solutions to SDEs of the form

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sigma(X_t) \circ \mathrm{d}B_t$$

more or less explicitly in terms of solutions to suitable ODEs. At this point it might seem as if SDEs are a boring extension of ODEs. But in our arguments it was absolutely crucial that the noise is 1-dimensional ($m=1$; for $m>1$ we could not interpret B as a "random time change").

For multidimensional Brownian motion, this works only under very particular conditions on b and σ; typically, one must require the ODE flows induced by different vector fields σ_j (the ones coming from writing the SDE as (8.9) to commute, thus enforce Lie Bracket conditions of the form $[\sigma_j, \sigma_k] = 0$ for all $j, k \in \{1, \ldots, m\}$.

Example 8.18. (Brownian motion on the unit circle) Let us consider a concrete example of an ODE in Brownian time: We take $\varphi: \mathbb{R} \to \mathbb{R}^2$ as a rotation on the unit circle,

$$\varphi'(t) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \varphi(t), \qquad \varphi(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

The explicit solution is $\varphi(t) = \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix}$. So if B is a one-dimensional Brownian motion, then we call the following process *Brownian motion on the unit circle:*

$$X_t = \varphi(B_t) = \begin{pmatrix} \cos(B_t) \\ \sin(B_t) \end{pmatrix} \in \mathbb{R}^2, \qquad t \geqslant 0.$$

Note that $|X_t| = 1$ for all $t \geqslant 0$ (otherwise, "Brownian motion on the unit circle" would be a very bad name). By the considerations above, we have with $f(x) = \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix}$:

$$\begin{aligned} \mathrm{d}X_t &= \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} X_t \circ \mathrm{d}B_t \\ &= \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} X_t \,\mathrm{d}B_t + \frac{1}{2}(\nabla f f)(X_t)\mathrm{d}t \\ &= \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} X_t \,\mathrm{d}B_t - \frac{1}{2} X_t \,\mathrm{d}t, \end{aligned}$$

so X solves a linear SDE (both in Itô and Stratonovich form). The assumptions of Theorem 8.11 are satisfied for this equation, so X is the unique solution.

We can also formulate this differently: We let the Brownian motion B run on $\mathbb{R}$, but we consider its value modulo 2π. Then we identify $[0, 2\pi)$ with the unit circle S^1 via $\varphi(\theta) := e^{i\theta} = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix}$, where we identified $\mathbb{C}$ with $\mathbb{R}^2$.

Exercise. Find an explicit solution to the SDE $\mathrm{d}X_t = 2\sqrt{X_t} \circ \mathrm{d}B_t$, $X_0 = 1$, up to a suitable stopping time τ. Note that the square root function is smooth and Lipschitz continuous on $[\varepsilon, \infty)$ for any $\varepsilon > 0$. Derive the equivalent Itô formulation of the SDE (up to the stopping time), using the definition of Stratonovich integral and the fact itself that X is a solution. Does the formula for the explicit solution X you found *extend* to times $t \geqslant \tau$?

8.4 Weak solutions and uniqueness in law

Recall the definition of weak solution to the SDE

$$X_t = \xi + \int_0^t b(s, X_s)\mathrm{d}s + \int_0^t \sigma(s, X_s)\mathrm{d}B_s$$

from Definition 8.5. In practice, one is often interested in computing statistics of the solution X, rather than its exact values depending on the Brownian path B. This motivates the following definition.

Definition 8.19. *We say that* weak uniqueness *holds for the SDE (8.8) if any two given weak solutions* $(\Omega^i, \mathcal{F}^i, \mathbb{F}^i, \mathbb{P}^i, \xi^i, B^i, X^i)$ *such that* $\mathrm{Law}_{\mathbb{P}^1}(\xi^1) = \mathrm{Law}_{\mathbb{P}^2}(\xi^2)$ *have the same finite-dimensional distributions. We may indifferently refer to this as* uniqueness in law *for the SDE, or we say that the solution* X *is* unique in law*.*

Remark 8.20. In other words, X is unique in law if its law as a $\mathbb{R}^d$-valued stochastic process $(X_t)_{t\geqslant 0}$ is uniquely determined. Since X is a continuous process, this also uniquely determines its law e.g. as a random variable in $C([0,T];\mathbb{R}^d)$ for any finite T. In other words, if weak uniqueness holds for the SDE, and we are given two solutions as above, then

$$\mathbb{E}_{\mathbb{P}^1}[\Phi(X^1)] = \mathbb{E}_{\mathbb{P}^2}[\Phi(X^2)]$$

for any bounded, measurable function $\Phi: C([0,T];\mathbb{R}^d) \to \mathbb{R}$.

Weak solutions are indeed a strictly larger class than strong solutions, and we may achieve uniqueness in law in situations where pathwise uniqueness fails, as the next example shows.

Example 8.21. (Tanaka equation) Consider the SDE

$$\mathrm{d}X_t = \mathrm{sign}(X_t)\mathrm{d}B_t, \qquad X_0 = 0,$$

where by convention $\mathrm{sign}(x) = \begin{cases} 1, & \text{if } x \geqslant 0, \\ -1, & \text{if } x < 0 \end{cases}$. Then:

i. Uniqueness in law holds: If X is a weak solution, then it is a local martingale started at 0 and with quadratic variation $\langle X\rangle_t = \int_0^t \mathrm{sign}(X_s)^2 \mathrm{d}s = t$, i.e. it is a Brownian motion.

ii. There exists a weak solution: Let X be a Brownian motion and define

$$B_t = \int_0^t \mathrm{sign}(X_s)\mathrm{d}X_s.$$

Then B is a Brownian motion in the filtration $(\mathbb{F}^X)^{+,\mathbb{P}}$, and

$$X_t = \int_0^t |\mathrm{sign}(X_s)|^2 \mathrm{d}X_s = \int_0^t \mathrm{sign}(X_s)\mathrm{d}B_s.$$

iii. Pathwise uniqueness fails: If X solves the SDE, also $-X$ solves the same SDE with initial condition $X_0 = 0$. Indeed, one can check that $\mathbb{P}$-a.s. $\int_0^{+\infty} \mathbb{1}_{\{X_s = 0\}} \mathrm{d}s = 0$ (Exercise: compute its expectation and use that X is a Brownian motion), so that

$$\begin{aligned} \mathrm{d}(-X)_t &= -\mathrm{sign}(X_t)\mathrm{d}B_t = -\mathbb{1}_{\{X_t \neq 0\}}\,\mathrm{sign}(X_t)\mathrm{d}B_t \\ &= \mathbb{1}_{\{X_t \neq 0\}}\,\mathrm{sign}(-X_t)\mathrm{d}B_t \\ &= \mathrm{sign}(-X_t)\mathrm{d}B_t. \end{aligned}$$

iv. Strong existence fails. A self-contained proof by contradiction of this fact would require more advanced tools like Itô Tanaka formula and the local time L of X, therefore we omit it. Let us only point out that, in view of points i.-iii. above, strong existence cannot hold by virtue of the upcoming Theorem 8.24.

Intuitively, strong uniqueness fails whenever the filtration generated by X is larger than the one generated by B. This roughly speaking means that X carries "more randomness" than B, and knowledge of B alone is not enough to reconstruct the solution X.

As Tanaka's equation suggests, an important motivation for studying weak solutions is that they may exist and/or be unique in law under more permissive conditions than those needed for strong existence/pathwise uniqueness.

Girsanov's theorem can play a key role in order to construct weak solutions and/or prove their uniqueness, as the next result shows.

Proposition 8.22. *Let $b: \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ be a uniformly bounded drift, namely such that*

$$|b(t,x)| \leqslant K \quad \forall t \geqslant 0, x \in \mathbb{R}^d \tag{8.22}$$

for some deterministic constant $K > 0$. Then for any $x_0 \in \mathbb{R}^d$, there exists a weak solution to the SDE

$$\mathrm{d}X_t = b(t, X_t)\mathrm{d}t + \mathrm{d}B_t$$

starting at x_0, which is unique in law. Moreover for any $T \in (0, +\infty)$ and any bounded measurable function $F: C([0,T];\mathbb{R}^d) \to \mathbb{R}$ we have

$$\mathbb{E}[F(X)] = \mathbb{E}\left[F(x_0 + B) \exp\left(\int_0^T b(s, x_0 + B_s) \cdot \mathrm{d}B_s - \frac{1}{2}\int_0^T |b(s, x_0 + B_s)|^2 \, \mathrm{d}s\right)\right]. \tag{8.23}$$

Proof. We divide the proof in two steps: we first *construct* a weak solution, and then later verify uniqueness in law. For simplicity, especially for existence, we work on a fixed finite interval $[0,T]$; up to technical details, one can actually allow $\mathbb{R}_+$ instead.

Step 1: weak existence. Consider the process $X_t := x_0 + W_t$, where we assume to be working on a probability space carrying a Brownian motion W. Then we may write

$$X_t = x_0 + \int_0^t b(s, X_s)\mathrm{d}s + \left(W_t - \int_0^t b(s, X_s)\mathrm{d}s\right) =: x_0 + \int_0^t b(s, X_s)\mathrm{d}s + B_t.$$

We define a new probability measure $\mathbb{Q}$ by

$$\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} = \exp\left(\int_0^T b(s, X_s) \cdot \mathrm{d}W_s - \frac{1}{2}\int_0^T |b(s, X_s)|^2 \, \mathrm{d}s\right),$$

where the right hand side has expectation 1 because Novikov's condition (Theorem 7.32) is satisfied, since by assumption

$$\mathbb{E}\left[\exp\left(\frac{1}{2}\int_0^T |b(s, X_s)|^2 \, \mathrm{d}s\right)\right] \leqslant \exp\left(\frac{TK^2}{2}\right) < \infty.$$

Under $\mathbb{Q}$, B is a Brownian motion on $[0,T]$ and therefore $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{Q}, B, X)$ is a weak solution the SDE on $[0,T]$. Moreover we have

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}}[F(X)] &= \mathbb{E}_{\mathbb{P}}\left[F(X)\, \mathcal{E}\left(\int_0^\cdot b(s, X_s) \cdot \mathrm{d}W_s\right)_T\right] \\
&= \mathbb{E}_{\mathbb{P}}\left[F(x_0 + W) \exp\left(\int_0^T b(s, x_0 + W_s) \cdot \mathrm{d}W_s - \frac{1}{2}\int_0^T |b(s, x_0 + W_s)|^2 \, \mathrm{d}s\right)\right] \\
&= \mathbb{E}_{\mathbb{Q}}\left[F(x_0 + B) \exp\left(\int_0^T b(s, x_0 + B_s) \cdot \mathrm{d}B_s - \frac{1}{2}\int_0^T |b(s, x_0 + B_s)|^2 \, \mathrm{d}s\right)\right]
\end{aligned}$$

proving (8.23) in this case.

Step 2: uniqueness in law. Let now $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, X, B)$ be a weak solution; fix a finite time horizon $[0,T]$. We may again apply Novikov to define a new probability measure $\tilde{\mathbb{Q}}$ via

$$\frac{\mathrm{d}\tilde{\mathbb{Q}}}{\mathrm{d}\mathbb{P}} = \exp\left(-\int_0^T b(s, X_s) \cdot \mathrm{d}B_s - \frac{1}{2}\int_0^T |b(s, X_s)|^2 \, \mathrm{d}s\right).$$

Then under $\tilde{\mathbb{Q}}$, since X is a solution to the SDE, we have

$$X_t = x_0 + \left(B_t + \int_0^t b(s, X_s) \mathrm{d}s \right) = x_0 + W_t$$

where W is a $\mathbb{F}$–Brownian motion on $[0,T]$ under $\tilde{\mathbb{Q}}$ and $W_t = X_t - x_0$. By plugging the above relations into the definition of the density $\frac{\mathrm{d}\tilde{\mathbb{Q}}}{\mathrm{d}\mathbb{P}}$, we can compute its inverse:

$$\begin{aligned}
\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\tilde{\mathbb{Q}}} = \left(\frac{\mathrm{d}\tilde{\mathbb{Q}}}{\mathrm{d}\mathbb{P}} \right)^{-1} &= \exp\left(\int_0^T b(s, X_s) \cdot \mathrm{d}B_s + \frac{1}{2} \int_0^T |b(s, X_s)|^2 \, \mathrm{d}s \right) \\
&= \exp\left(\int_0^T b(s, X_s) \cdot \mathrm{d}\left(W_\cdot - \int_0^\cdot b(r, X_r) \mathrm{d}r \right) + \frac{1}{2} \int_0^T |b(s, X_s)|^2 \, \mathrm{d}s \right) \\
&= \exp\left(\int_0^T b(s, X_s) \cdot \mathrm{d}W_s - \frac{1}{2} \int_0^T |b(s, X_s)|^2 \, \mathrm{d}s \right) \\
&= \exp\left(\int_0^T b(s, x_0 + W_s) \cdot \mathrm{d}W_s - \frac{1}{2} \int_0^T |b(s, x_0 + W_s)|^2 \, \mathrm{d}s \right).
\end{aligned}$$

Therefore

$$\mathbb{E}_{\mathbb{P}}[F(X)] = \mathbb{E}_{\tilde{\mathbb{Q}}}\left[F(x_0 + W) \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\tilde{\mathbb{Q}}} \right]$$

recovering formula (8.23). As the argument holds for any $T < \infty$ and any bounded measurable $F: C([0,T]; \mathbb{R}^d) \to \mathbb{R}$, formula (8.23) implies that the finite-dimensional marginals of X are uniquenely determined. □

Exercise. Let $b: \mathbb{R} \to \mathbb{R}$ be bounded and measurable. Going through a similar argument, show that weak uniqueness holds for the SDE

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \mathrm{sign}(X_t)\mathrm{d}B_t, \qquad X_0 = x_0 \in \mathbb{R}^d.$$

Remark 8.23. (More about Girsanov, not examinable) Proposition 8.22 admits many extensions and variations. In general, Girsanov allows to "transform SDEs into other SDEs" by adding/subtracting a drift term, and consequently allows to "transfer" uniqueness in law results from one SDE to the other.

For instance, the condition (8.22) of boundedness of b can be relaxed to allow linear growth, and the argument from Step 2 about uniqueness in law part holds much more generally (but then Step 1 might fail, and one must find a different way to construct a weak solution, without using Girsanov).

Moreover, even though it largely simplified the analysis, we don't need noise to be additive in the argument, i.e. we don't need to take $\sigma = I_d$. For instance, it suffices to know that $\sigma: \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^{d \times d}$ is locally Lipschitz and *uniformly bounded and nondegenerate*, in the sense that $\sigma(t,x)$ is an invertible $d \times d$ matrix and

$$|\sigma(t,x)| + |\sigma(t,x)^{-1}| \leqslant K \quad \forall (t,x) \in \mathbb{R}_{\geqslant 0} \times \mathbb{R}^d.$$

More generally, by writing $b = \sigma\sigma^{-1} b$, so that

$$\int_0^\cdot b(s, X_s)\mathrm{d}s + \int_0^\cdot \sigma(s, X_s)\mathrm{d}B_s = \int_0^\cdot \sigma(s, X_s)\mathrm{d}\left(B + \int_0^\cdot \sigma(r, X_r)^{-1} b(r, X_r)\mathrm{d}r \right)_s,$$

we see that the key condition concerns integrabilty of $(\sigma^{-1} b)(r, X_r)$. So σ can be allowed to be degenerate, but only at points where b becomes 0 at the same time.

We refer to Chapter 5.3.B from [14] for a deeper discussion on the topic.

Extra comment on the literature (not examinable):

So far we showed that solutions to SDEs exist under similar conditions as solutions to ODEs (cf. Theorem 8.11). But Proposition 8.22 shows that SDEs can behave much better, and existence and uniqueness can be achieved in scenarios where the same is not true for the corresponding ODE. An example is given by $b = \mathbb{1}_{\{0\}}$, cf. Example 8.4: for $\sigma = 0$ and $x_0 = 0$ there exist no solutions, but for $\sigma = 1$ there exists a unique-in-law one. In fact in this case the solution is given by $X_t = \sigma B_t$, since then $\int_0^t \mathbb{1}_{\{0\}}(X_s) \mathrm{d}s = 0$ for all t. The noise drives us immediately away from the singularity 0 of our drift.

This is the simplest example of a *regularization by noise* phenomenon, where randomness helps improving the wellposedness of the system. In fact, a must stronger version of Proposition 8.22 holds true, as was shown by Veretennikov [27] in the 80's: for bounded, measurable drift $b: \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma = I_d$, strong existence and pathwise uniqueness holds for the SDE. The proof is much more advanced and strongly relies on PDE tools. There is a lot of ongoing research in regularization by noise phenomena; see [8] for a monograph on the topic.

The following result is extremely useful and clarifies the relations between the various notions of solutions and uniqueness thereof we have encountered so far. Point i. provides a simple criterion for uniqueness in law, while Point ii. additionally guarantees strong existence of solutions.

Theorem 8.24. (Yamada–Watanabe) *Let b, σ be locally bounded, measurable coefficients and consider the SDE (8.8). Then:*

i. Pathwise uniqueness implies uniqueness in law.

ii. If there exists a weak solution $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, \xi, B, X)$ to the SDE and pathwise uniqueness holds, then X is a strong solution, namely it is adapted to the (usual augmentation of the) filtration generated by ξ and B.

Point ii. may be summarized as "weak existence + pathwise uniqueness ⇒ strong existence".

Proof. See Chapter 5.3.D from Karatzas-Shreve [14], in particular Proposition 5.3.20 and Corollary 5.3.23. □

The power of Theorem 8.24 lies in the fact that there are many situations where one can develop ad hoc arguments to verify pathwise uniqueness and weak existence separately (which then jointly imply both uniqueness in law and strong existence, so that we "get everything"). It admits a less known, somewhat dual statement, due to Cherny [2].

Theorem 8.25. (Cherny) *Let b, σ be locally bounded, measurable coefficients and consider the SDE (8.8). Assume that there exists a strong solution and that uniqueness in law holds; then pathwise uniqueness holds as well.*

Schematically: "strong existence + uniqueness in law ⇒ pathwise uniqueness".

Proof. We refer to [2] for the proof, which is omitted. The statement is less obvious then what it might look like: to prove it, Cherny first shows that uniqueness in law for X actually implies uniqueness in law for the pair (X, B), where B is the noise driving the SDE satisfied by X. □

9 Further topics

This chapter contains a selection of more advanced results we might shortly touch upon in the very final lectures of the course, due to lack of time. In fact, a fitting title for the chapter could have been "Regrets". Regardless of what we will be able to cover, the content of this chapter is not examinable.

9.1 Connections between SDEs and PDEs

In this section, we restrict our attention to SDEs with *time-independent* (also called *autonomous*) coefficients, i.e. $b: \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma: \mathbb{R}^d \to \mathbb{R}^{d\times m}$ are only functions of x. For simplicity, we also assume them to be *continuous* and *uniformly bounded*, namely there exists $K>0$ such that

$$|b(x)| + |\sigma(x)| \leqslant K \quad \forall x \in \mathbb{R}^d;$$

for short, we will write $b, \sigma \in C_b$. Everything can be generalized to suitable time-dependent, less regular, unbounded coefficients b, σ (e.g. measurable and satisfying certain growth conditions), at the price of additional technicalities.

Let X be a weak solution to the SDE

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t, \qquad X_0 = \xi. \tag{9.1}$$

In the following, we write

$$C_b^{1,2}(\mathbb{R}_+ \times \mathbb{R}^d) := \{f \in C^{1,2}: f, \partial_t f, \partial_{x_i} f, \partial_{x_i x_j} f \text{ are continuous and bounded in x}\}.$$

Lemma 9.1. *Let $b, \sigma \in C_b$. Corresponding to the SDE (9.1), define the differential operator*

$$\mathcal{L}f(x) := b(x)\cdot\nabla f(x) + \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(x)\partial_{x_i x_j} f(x), \qquad f \in C^2(\mathbb{R}^2), \tag{9.2}$$

where

$$a_{ij}(x) := (\sigma\sigma^*)_{ij}(x) = \sum_{k=1}^{m} \sigma_{ik}(x)\sigma_{jk}(x).$$

Let X solve the SDE (9.1). Then for all $f \in C_b^{1,2}(\mathbb{R}_+ \times \mathbb{R}^d)$ the following process is a martingale:

$$M_t^f := f(t, X_t) - f(0, X_0) - \int_0^t (\partial_t + \mathcal{L}) f(s, X_s)\mathrm{d}s.$$

Sketch of proof. By applying Itô's formula to f, using recursively the SDE satisfied by X to compute $\langle X^i, X^j\rangle$, one finds

$$f(t, X_t) - f(0, X_0) - \int_0^t (\partial_s + \mathcal{L}) f(s, X_s) = \sum_{i=1}^{d}\sum_{k=1}^{m}\int_0^t \partial_i f(s, X_s)\sigma_{ik}(X_s)\mathrm{d}B_s^k.$$

Since we assumed σ and $\partial_i f$ bounded, the last stochastic integral is a genuine martingale (and not just a local martingale). □

By virtue of the above result, the differential operator $\mathcal{L}$ as defined in (9.2) is often referred to as the *infinitesimal generator* associated to the SDE (9.1).

Example 9.2. If $b=0$ and $m=d$ and σ is the identity matrix (i.e. if X is a d-dimensional Brownian motion), then

$$\mathcal{L} = \frac{1}{2}\Delta$$

is simply (a multiple of) the Laplace operator.

Definition 9.3. (Kolmogorov backward equation) *Let $\varphi: \mathbb{R}^d \to \mathbb{R}$ be Borel measurable and locally bounded. The partial differential equation (PDE)*

$$\partial_t u(t,x) = \mathcal{L}u(t,x), \qquad u(0,x) = \varphi(x), \qquad (t,x) \in \mathbb{R}_+ \times \mathbb{R}^d, \tag{9.3}$$

is called the Kolmogorov backward equation.

Theorem 9.4. *Let $b, \sigma \in C_b$, let X be a weak solution to the SDE (9.1) with random initial condition ξ, and let $u \in C_b^{1,2}$ be a solution to the Kolmogorov backward equation (9.3). Then*

$$\mathbb{E}[u(t,\xi)] = \mathbb{E}[\varphi(X_t)] \qquad \forall t \geqslant 0.$$

If for all $x \in \mathbb{R}^d$ there exists a weak solution X^x to the SDE (9.1) with $X_0^x = x$, then u is unique and

$$u(t,x) = \mathbb{E}[\varphi(X_t^x)] \qquad \forall (t,x) \in \mathbb{R}_+ \times \mathbb{R}^d. \tag{9.4}$$

Sketch of proof. The second claim immediately follows from the first one. For fixed $t > 0$, let $f(s,x) := u(t-s,x)$ for $(s,x) \in [0,t] \times \mathbb{R}^d$; then f satisfies

$$\partial_s f(s,x) + \mathcal{L} f(s,x) = 0 \quad \text{for } (s,x) \in [0,t] \times \mathbb{R}^d, \qquad f(0,x) = u(t,x).$$

By virtue of Lemma 9.1, M^f is then a martingale on $[0,t]$, so that

$$0 = \mathbb{E}[M_t^f] = \mathbb{E}[f(t,X_t)] - \mathbb{E}[f(0,X_0)] = \mathbb{E}[u_0(X_t)] - \mathbb{E}[u(t,\xi)]. \qquad \square$$

Remark 9.5. Remark 9.6.

i. The theorem proves *uniqueness* and a *stochastic representation* for the solution u to a *partial differential equation*, specifically the Kolmogorov backward equation, but we had to assume that u exists (which would be an input from PDE theory). We could also ask if it is possible to show that

$$u(t,x) := \mathbb{E}[\varphi(X_t^x)]$$

is indeed a solution, i.e. to prove *existence*. Under suitable regularity and growth conditions on b, σ, φ this is indeed possible, but the proof is quite technical (the idea is to use suitable generalizations of Kolmogorov's continuity criterion to prove that $u \in C^{1,2}$).

ii. As a simple consequence of the stochastic representation (9.4) we obtain the following *maximum principle* for the solution to the Kolmogorov backward equation:

$$\sup_{t \geqslant 0, x \in \mathbb{R}^d} u(t,x) \leqslant \sup_{x \in \mathbb{R}^d} \varphi(x).$$

The differential operator $\mathcal{L}u$ depends on partial derivatives of order 1 and 2. Theorem 9.4 can be slightly generalized to include additional terms of order 0 (i.e. depending on u itself); the proof of the next result is omitted.

Theorem 9.7. (Feynman–Kac representation) *Let $b, \sigma, c \in C_b$ and let $u \in C_b^{1,2}$ be a solution to the PDE*

$$\partial_t u = \mathcal{L}u + cu, \qquad u(0) = \varphi.$$

If for all $x \in \mathbb{R}^d$ there exists a solution X^x to the SDE (9.1) with $X_0^x = x$, then u is unique and

$$u(t,x) = \mathbb{E}\left[\varphi(X_t^x) \exp\left(\int_0^t c(X_s^x) \mathrm{d}s\right)\right].$$

As the name suggests, the *backward* PDE (9.3) admits another closely related PDE, of *forward* type. To this end, we need some notations.

We write $\mathcal{P}(\mathbb{R}^d)$ for the space of probability measures on $\mathbb{R}^d$; we say that a function μ: $\mathbb{R}_+ \to \mathcal{P}(\mathbb{R}^d)$ is *weakly continuous*, and we write $\mu \in C(\mathbb{R}_+, \mathcal{P}(\mathbb{R}^d))$, if for all $\varphi \in C_b(\mathbb{R}^d)$ the function

$$\mathbb{R}_+ \ni t \mapsto \mu_t(\varphi) := \int_{\mathbb{R}^d} \varphi(x) \mu_t(\mathrm{d}x) \in \mathbb{R}$$

is continuous.

Definition 9.8. *Let $b, \sigma \in C_b$ and let*

$$\mathcal{L}^* f := -\mathrm{div}\,(b f) + \frac{1}{2} \sum_{i,j=1}^{d} \partial_{ij}(a_{ij} f), \tag{9.5}$$

for $a = \sigma\sigma^*$ as before. A function $\mu \in C(\mathbb{R}_+, \mathcal{P}(\mathbb{R}^d))$ is called a weak solution to the Kolmogorov forward equation *(also known as* Fokker-Planck equation *or* master equation*)*

$$\partial_t \mu = \mathcal{L}^* \mu, \qquad \mu_0 = \bar{\mu} \tag{9.6}$$

if for all $\varphi \in C_c^2$ and all $t \geqslant 0$:

$$\mu_t(\varphi) = \bar{\mu}(\varphi) + \int_0^t \mu_s(\mathcal{L}\varphi) \mathrm{d}s.$$

The notation $\mathcal{L}^*$ is meant to stress the fact that, at least formally, this differential operator is the *dual* of the operator $\mathcal{L}$ previously defined in (9.2). In fact, whenever f and g are regular enough functions (e.g. C_c^∞, infinitely differentiable and compactly supported) and $b, \sigma \in C^2$, one can check by integration by parts that

$$\int_{\mathbb{R}^d} (\mathcal{L}f)(x)\, g(x)\, \mathrm{d}x = \int_{\mathbb{R}^d} f(x) (\mathcal{L}^* g)(x)\, \mathrm{d}x.$$

Correspondingly, the PDEs (9.3) and (9.6) are *dual* to each other, or in a *duality relation*.

Theorem 9.9. *Let $b, \sigma \in C_b$ and let X be a solution to the SDE (9.1). Then $\mu_t := \mathrm{law}(X_t)$ is a weak solution to the Kolmogorov forward equation*

$$\partial_t \mu = \mathcal{L}^* \mu, \qquad \mu_0 = \mathrm{law}(\xi).$$

Sketch of proof. Clearly μ_t is a probability measure and $\mu_0 = \mathrm{law}(\xi)$. Moreover, since X has continuous paths, it follows from dominated convergence that $\mu \in C(\mathbb{R}_+, \mathcal{P}(\mathbb{R}^d))$. Applying Lemma 9.1 to $f = \varphi$ and taking expectation (so that the martingale term M^φ vanishes), one finds

$$\begin{aligned} \mu_t(\varphi) &= \mathbb{E}[\varphi(X_t)] \\ &= \mathbb{E}\left[\varphi(X_0) + \int_0^t \mathcal{L}\varphi(X_s) \mathrm{d}s\right] \\ &= \mathbb{E}[\varphi(\xi)] + \int_0^t \mu_s(\mathcal{L}\varphi) \mathrm{d}s \end{aligned}$$

where the last step follows from Fubini. □

So far we have only discussed time-dependent PDEs, which should be thought of as *evolutionary problems*: given an initial prescribed profile (e.g. φ in (9.3)), we want to see how it looks like at positive times, where it is given by $u(t, \cdot)$. However, *static* problems also admit stochastic representations; here the prescribed *boundary condition* for the PDE plays an important role.

Definition 9.10. *Let $\mathcal{L}$ be the differential operator from (9.2) and let $D \subset \mathbb{R}^d$ be a bounded domain (i.e. open, connected and contained in a compact). Let $\psi \in C(\partial D)$. A function $u \in C(\overline{D}, \mathbb{R}) \cap C^2(D, \mathbb{R})$ solves the* Dirichlet problem on D with boundary condition ψ *if*

$$\begin{cases} \mathcal{L}u(x) = 0 & \text{for } x \in D, \\ u(x) = \psi(x) & \text{for } x \in \partial D. \end{cases} \tag{9.7}$$

The associated stochastic representation is given by the next statement.

Theorem 9.11. *Let $u \in C(\overline{D}, \mathbb{R}) \cap C^2(D, \mathbb{R})$ be a solution to the Dirichlet problem. Assume that for all $x \in D$ there exists a solution X^x to the SDE (9.1) such that*

$$\tau_D = \inf \{t \geqslant 0 \colon X_t^x \in \partial D\}$$

is almost surely finite. Then u is unique and

$$u(x) = \mathbb{E}[\psi(X_{\tau_D}^x)], \qquad x \in D.$$

Sketch of proof. Let $x \in D$, τ_D as defined above. By Lemma 9.1 applied to $f = u$, $Y_t := u(X_t^x)$ is a (local) martingale started at 0, therefore (up to tehcnicalities involving localizations) by the stopping theorem

$$u(x) = \mathbb{E}[Y_0] = \mathbb{E}[Y_{\tau_D}] = \mathbb{E}[u(X_{\tau_D}^x)] = \mathbb{E}[\psi(X_{\tau_D}^x)]$$

since $X_{\tau_D}^x \in \partial D$ by definition. □

9.2 Local solutions and criteria for absence of blow-up

In applications one often encounters equations with coefficients that do not satisfy the assumptions of Theorem 8.11. Recall for example the logistic growth model from Example 8.2 (with $M = 1$ for simplicity):

$$\mathrm{d}X_t = r(1 - X_t)X_t\,\mathrm{d}t + \mu X_t\,\mathrm{d}B_t. \tag{9.8}$$

The quadratic function $f(x) = r(1 - x)x = rx - rx^2$ is not globally Lipschitz continuous ($f'(x) = r - 2rx$ explodes as $|x| \to \infty$) and it does not have linear growth ($|f(x)| \sim r|x|^2$ as $|x| \to \infty$). On the other hand, it is locally Lipschitz continuous, so we know that uniqueness holds; in analogy to ODEs, we still expect a solution to exist "locally", only that it might blow-up in finite time; furthermore since $f(x)$ can only explode "with negative values" and solutions should stay nonnegative (X_t models a population), it is reasonable to expect blow-up not to happen in this case.

To prove this, we need to extend the SDE theory to allow for *local solutions* and to develop criteria to exclude blow-up from happening.

Definition 9.12. *Let (b, σ) be as in Definition 8.5. We say that $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, \xi, B, X, \tau)$ is a* local weak solution *to the SDE on $[0, \tau]$ if the following hold:*

i. $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, \xi, B)$ satisfy the same conditions as before;

ii. τ is a $\mathbb{F}$-stopping time and $X = (X_t)_{t \geqslant 0}$ is a d-dimensional continuous $\mathbb{F}$-adapted process;

iii. $\mathbb{P}$-almost surely, for all $t \geq 0$ and $i \in \{1, \ldots, d\}$ it holds that

$$X_{t \wedge \tau}^i = \xi + \int_0^{t \wedge \tau} b_i(s, X_s)\mathrm{d}s + \sum_{j=1}^m \int_0^{t \wedge \tau} \sigma_{ij}(s, X_s)\mathrm{d}B_s^j. \tag{9.9}$$

We say that X *is a* local strong solution *on* $[0,\tau]$ *if* X *is* $\mathbb{G}$*-adapted and* τ *is a* $\mathbb{G}$*-stopping time, where* $\mathbb{G}$ *is the (augmented) filtration generated by* (ξ, B).

One can readily readapt the concept of pathwise uniqueness in this case (specifically, given two local solutions (X, τ^1) and (Y, τ^2) on the same reference space, they must coincide on the common interval of existence $[0, \tau^1 \wedge \tau^2]$).

Proposition 9.13. *Assume that the coefficients* b *and* σ *are locally bounded and moreover* locally Lipschitz: *for all* $T > 0$ *and* $n \in \mathbb{N}$ *there exists* $K_{T,n}$ *such that*

$$|b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leqslant K_{T,n}|x-y|$$

for all $t \in [0,T]$ *and* $|x|, |y| \leqslant n$*. Then, for any* $\mathcal{F}_0$*–measurable initial condition* ξ*, there exists a local strong solution* X *to the SDE, defined till any stopping time* τ_U *of the form*

$$\tau_U = \inf\{t \geqslant 0 \colon X_t \in U^c\}$$

where U *is an open bounded set in* $\mathbb{R}^d$*. Furthermore, this solution is pathwise unique.*

Sketch of proof. Consider a sequence of *cutoff functions* $\{\varphi_n\}_n$, namely infinitely differentiable $\varphi_n \colon \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}$ such that

$$\varphi_n(t,x) = 1 \text{ for } t \in [0,n] \text{ and } |x| \leqslant n, \qquad \varphi_n(t,x) = 0 \text{ for } t \geqslant n+1 \text{ or } |x| \geqslant n+1$$

(such sequences can be shown to exist). Correspondingly, define the coefficients

$$b^n := b\varphi_n, \quad \sigma^n := \sigma\varphi_n.$$

By construction, for fixed n, (b^n, σ^n) are globally Lipschitz and bounded, therefore by Theorem 8.11 there exists a strong solution to the associated SDE; denote it by X^n. Correspondingly, define $\tau_n := \inf\{t \geqslant 0 \colon |X_t| \geqslant n\}$.

Let $m \geqslant n$; since $b^n(t,x) = b^m(t,x)$ as long as $t \leqslant n$, $|x| \leqslant n$, arguing as in the proof of Theorem 8.11, it follows that the solutions X^n and X^m coincide, as long as they both do not leave the ball of radius n. As a consequence, one can consistently define a stochastic process X (adapted to the filtration $\mathbb{G}$) on $[0, \tau_n]$ by

$$X_{t\wedge\tau_n} := X^n_{t\wedge\tau_n};$$

by construction, X is a local solution to the SDE on $[0, \tau_n]$.

Given U bounded open set, we can then take any n such that $U \subset B_n$, where $B_n = \{x \in \mathbb{R}^d \colon |x| < n\}$, and correspondigly define

$$\tau_U = \inf\{t \geqslant 0 \colon X_t \in U^c\} = \inf\{t \geqslant 0 \colon X^n_t \in U^c\}.$$ □

Remark 9.14. Let b, σ be differentiable in x with continuous $\partial_i b$ and $\partial_i \sigma$; then by the mean-value theorem, b and σ are locally Lipschitz continuous.

Definition 9.15. *Under the hypothesis of Proposition 9.13, consider the stopping time*

$$\tau^* := \lim_{n\to\infty} \tau_{B_n}, \quad \textit{where } B_n = \{x \in \mathbb{R}^d \colon |x| < n\}.$$

τ^* *is called the* explosion time *of the SDE with coefficients* (b, σ) *and initial condition* ξ *and* (X, τ^*) *is called the* maximal solution *to the SDE starting at* ξ*.*

Remark 9.16. By definition, it holds

$$\sup_{t\in[0,\tau^n]} |X_t| < \infty, \qquad \lim_{t\to\tau^*} |X_t| = +\infty.$$

In this sense, X is a maximal solution because it is no longer extendable to intervals $[0,\tilde{\tau}]$ with $\tilde{\tau}>\tau^*$.

Definition 9.17. *In the setting of Definition 9.15, we say that X is a* global solution *if explosion does not occur, namely if*

$$\mathbb{P}(\tau^*=+\infty)=1.$$

To show that blow-up does not occurr, there is a general class of techniques based on *a priori estimates* and *Lyapunov functions* (cf. also Exercise Sheet 14). The most basic Lyapunov function, which is often enough in applications, is $V(x)=|x|^2$, as the next result shows.

Proposition 9.18. (A priori estimate under weak coercivity) *Assume that the following* weak coercivity condition *holds: for all $T>0$ there is $K_T>0$ such that*

$$2b(t,x)\cdot x+|\sigma(t,x)|^2\leqslant K_T(1+|x|^2),\qquad (t,x)\in[0,T]\times\mathbb{R}^d.$$

Let ξ be $\mathcal{F}_0$–measurable, $\xi\in L^2$ and let X be a local solution to the SDE on the interval $[0,\tau]$. Then

$$\mathbb{E}[|X_{t\wedge\tau}|^2]\leqslant(\mathbb{E}[|\xi|^2]+K_t\,t)e^{K_t t}\quad\forall t\geqslant 0. \tag{9.10}$$

Proof. The argument is very similar to that of Proposition 8.14. Let

$$\tilde{\tau}_n=\inf\{t\geqslant 0\colon |X_t|\geqslant n\},\quad \tau_n:=\tilde{\tau}_n\wedge\tau$$

so that $|X_t^{\tau_n}|\leqslant|\xi|\vee n$ and so X is a local bounded solution to the SDE on $[0,\tau_n]$. We will apply Itô's formula to $|X_t^{\tau_n}|^2$. For that purpose note that

$$\partial_i|x|^2=2x_i,\qquad \partial_{ij}|x|^2=2\delta_{ij},$$

and that using recursively the SDE we have

$$\langle X^i,X^j\rangle_{t\wedge\tau}=\sum_{k=1}^m\int_0^t\sigma_{ik}(s,X_s)\sigma_{jk}(s,X_s)\mathrm{d}s;$$

thus

$$\begin{aligned}|X_t^{\tau_n}|^2=|\xi|^2+\sum_{i=1}^d\int_0^{t\wedge\tau_n}2(X_s^{\tau_n})^i b_i(s,X_s^{\tau_n})\mathrm{d}s+\sum_{i=1}^d\sum_{k=1}^m\int_0^{t\wedge\tau_n}\sigma_{ik}(s,X_s^{\tau_n})^2\mathrm{d}s\\ +\sum_{i=1}^d\sum_{k=1}^m\int_0^{t\wedge\tau_n}2(X_s^{\tau_n})^i\sigma_{ik}(s,X_s^{\tau_n})\mathrm{d}B_s^k\end{aligned}$$

Since σ is locally bounded and everything is stopped, the last term (in integral form) is a genuine martingale, therefore taking expectation on both sides we end up with

$$\begin{aligned}\mathbb{E}[|X_t^{\tau_n}|^2]&=\mathbb{E}[|\xi|^2]+\mathbb{E}\bigg[\int_0^{t\wedge\tau_n}(2X_s^{\tau_n}\cdot b(s,X_s^{\tau_n})+|\sigma(s,X_s^{\tau_n})|^2)\mathrm{d}s\bigg]\\ &\leqslant\mathbb{E}[|\xi|^2]+\mathbb{E}\bigg[\int_0^{t\wedge\tau_n}K_t(1+|X_s^{\tau_n}|^2)\mathrm{d}s\bigg]\\ &\leqslant\mathbb{E}[|\xi|^2]+K_t\,t+K_t\int_0^t\mathbb{E}[|X_s^{\tau_n}|^2]\mathrm{d}s.\end{aligned}$$

The conclusion then follows by Grönwall's lemma. □

Combining the previous results, one can deduce a stronger global well-posedness result than our previous Theorem 8.11, which only allowed global Lipschitz coefficients with linear growth:

Theorem 9.19. *Assume that the following local Lipschitz and weak coercivity conditions hold: for any $T\in(0,+\infty)$ and $n\in\mathbb{N}$, there are $K_{T,n},K_T\geqslant 0$ such that for all $t\in[0,T]$:*

$$|b(t,x)-b(t,y)|+|\sigma(t,x)-\sigma(t,y)|\leqslant K_{T,n}|x-y| \qquad \forall |x|,|y|\leqslant n,$$
$$2b(t,x)\cdot x+|\sigma(t,x)|^2\leqslant K_T(1+|x|^2) \qquad \forall x\in\mathbb{R}^d.$$

Then for any $\mathcal{F}_0$–measurable initial condition ξ, the associated maximal solution X is global:

$$\mathbb{P}(\tau^*=+\infty)=1.$$

In particular, global strong existence and pathwise uniqueness hold for the SDE.

Sketch of proof. First assume $\xi\in L^2$. The local Lipschitz conditions ensures existence of a maximal solution X on $[0,\tau^*)$, where $\tau^*=\lim_{n\to\infty}\tau^n$. For fixed $m\in\mathbb{N}$, by (9.10) and Fatou's lemma we have

$$\mathbb{E}[|X_{m\wedge\tau^*}|^2]\leqslant\liminf_{n\to\infty}\mathbb{E}[|X_{m\wedge\tau^n}|^2]\leqslant(\mathbb{E}[|\xi|^2]+K_m\,m)e^{K_m m}.$$

But since $|X_{\tau^*}|=+\infty$, it follow that $|X_{m\wedge\tau^*}|=|X_m|$ $\mathbb{P}-a.s.$, namely $\mathbb{P}(\tau^*>m)=1$. As the argument holds for all $m\in\mathbb{N}$, conclusion follows.

The case of general ξ can be handled similarly to the proof of Theorem 8.11. □

Example 9.20. (Logistic growth model) Recall the logistic growth model (9.8) from the beginning of the section, so that $b(x)=r(1-x)x$ and $\sigma(x)=\mu x$. Since the coefficients are locally Lipschitz, for any initial condition ξ we know that there exists a pathwise unique maximal solution X to the SDE, defined on $[0,\tau^*)$.

Notice that $Y\equiv 0$ is also solution to the SDE (started at 0). Then by the *comparison principle* (Sheet 14, which can be extended to the more general situation considered here), if $X_0=\xi>0$ $\mathbb{P}$-a.s., then we must have $X_t\geqslant 0$ for all $t\in[0,\tau^*)$ (as it should be: if the model is meaningful, the population should never get negative).

Exploiting this fact, the coefficients become monotone for $x\in\mathbb{R}_+$: indeed b satisfies

$$b(x)x=r(1-x)x^2=rx^2-rx^3\leqslant rx^2$$

as long as $x\geqslant 0$. As a consequence, solutions do not blow-up for this SDE and we recover strong global existence and pathwise uniqueness for (9.8).

9.3 The martingale representation theorem

Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space carrying a $\mathbb{R}^d$-valued Brownian motion B; in this section, we fix the reference filtration $\mathbb{F}=(\mathbb{F}^B)^{+,\mathbb{P}}$ is the usual augmentation of the canonical filtration of B.

We have seen throughout examples that, while the stochastic integrals

$$M_t=\int_0^t h_s\cdot\mathrm{d}B_s$$

with deterministic integrands h are quite rigid (they are centered Gaussian), general integrand $H\in L^2_{\mathrm{loc}}(B)$ can display a much reacher behaviour.

In fact, one can show that *any* Brownian random variable of mean zero can be realized as a stochastic integral:

Theorem 9.21. (Martingale representation theorem in $\mathbb{R}^d$, version 1) *Let B be a d-dimensional Brownian motion, let $\mathbb{F} = (\mathbb{F}^B)^{+,\mathbb{P}}$ be the (augmented) canonical filtration of B. If $Z \in L^2(\mathcal{F}_\infty, \mathbb{P})$, then there is a unique $H \in L^2(B)$ (i.e. H is progressive and $\mathbb{E}[\int_0^\infty |H_s|^2 \mathrm{d}s] < \infty$), such that*

$$Z = \mathbb{E}[Z] + \int_0^\infty H_s \cdot \mathrm{d}B_s = \mathbb{E}[Z] + \sum_{i=1}^d \int_0^\infty H_s^i \mathrm{d}B_s^i. \tag{9.11}$$

Ideas behind the proof. The result can be established by abstract Hilbert space theory (similarly to what we did in Lemma 6.6) in three main steps:

Step 1: uniqueness. Let H^1, H^2 be such that (9.11) holds, then we find

$$\int_0^\infty (H_s^1 - H_s^2) \cdot \mathrm{d}B_s = \int_0^\infty H_s^1 \cdot \mathrm{d}B_s - \int_0^\infty H_s^2 \cdot \mathrm{d}B_s = 0$$

from which we conclude by Itô isometry that $H^1 \equiv H^2$ in $L^2(B)$.

Step 2: "special classes of martingales". For deterministic integrands $h \in L^2(\mathbb{R}_+; \mathbb{R}^d)$, one can prove that the collection of associated exponential martingales

$$\mathcal{X} = \operatorname{span}\left\{\mathcal{E}(h)_\infty = \exp\left(\int_0^\infty h(s) \cdot \mathrm{d}B_s - \frac{1}{2}\int_0^\infty h(s)^2 \mathrm{d}s\right) : h \in L^2(\mathbb{R}_+; \mathbb{R}^d)\right\}$$

is dense in $L^2(\mathcal{F}_\infty)$.

Step 3: existence. Consider now the collection of random variables admitting a representation of the form (9.11):

$$\mathcal{Z} = \left\{Z \in L^2 : Z = \mathbb{E}[Z] + \int_0^\infty H_s \cdot \mathrm{d}B_s \text{ for some } H \in L^2(B)\right\}.$$

The goal is to show that $\mathcal{Z} = L^2$. To this end, it suffices to show that $\mathcal{Z}$ is closed and dense in L^2. $\mathcal{Z}$ being closed follows again from Itô isometry, while $\mathcal{Z}$ being dense follows from $\mathcal{Z}$ containing $\mathcal{X}$: since exponential martingales themselves solve SDEs, we have

$$\mathcal{E}(h)_\infty = 1 + \int_0^{+\infty} \mathcal{E}(h)_s h_s \cdot \mathrm{d}B_s$$

which gives the explicit formula for H for processes in $\mathcal{X}$. □

Theorem 9.21 has the important consequence that any Brownian local martingale can be represented as a stochastic integral w.r.t. B; in particular, Brownian local martingales are always necessarily continuous.

Theorem 9.22. (Martingale representation theorem in $\mathbb{R}^d$, version 2) *Let B be a d-dimensional Brownian motion and let $\mathbb{F} = (\mathbb{F}^B)^{+,\mathbb{P}}$ be the (augmented) canonical filtration of B. For every $\mathbb{R}$-valued càdlàg local martingale M with $M_0 \in L^1$ there exists a unique $\mathbb{R}^d$-valued process $H \in L^2_{\mathrm{loc}}(B)$ such that*

$$M_t = \mathbb{E}[M_0] + \int_0^t H_s \cdot \mathrm{d}B_s \qquad \forall t \geqslant 0,$$

in the sense that the two processes are indistinguishable. In particular, M is almost surely continuous.

Ideas behind the proof. There are three main steps.

Step 1: square integrable martingales. First assume M be uniformly integrable with $M_\infty \in L^2$. Then applying Theorem 9.21 to $H = M_\infty$ one finds

$$M_\infty = \mathbb{E}[M_\infty] + \int_0^\infty H_s \cdot \mathrm{d}B_s = \mathbb{E}[M_0] + \int_0^\infty H_s \cdot \mathrm{d}B_s$$

from which the formula at time t follows by conditioning w.r.t. $\mathcal{F}_t$ on both sides. It is also clear that such process H is unique, by Itô isometry as before.

Step 2: localization. Given a càdlàg local martingale M, one wants to first show that there exists a localizing sequence $(\tau_n)_n$ such that $M_\infty^{\tau_n} \in L^2$ for all n. This is technically more challenging than it looks, because we do not know whether M is continuous, so standard hitting times won't work (the process M could jump exactly at the hitting time and get "much higher" than the desired barrier n). Still, at the end of the day one can succeed in constructing such $(\tau_n)_n$.

Step 3: conclusion. Applying Step 1 to $M_\infty^{\tau_n}$ one gets

$$M_\infty^{\tau_n} = \mathbb{E}[M_\infty^{\tau_n}] + \int_0^\infty H_s^n \cdot \mathrm{d}B_s = \mathbb{E}[M_0] + \int_0^\infty H_s^n \cdot \mathrm{d}B_s.$$

By properties of stochastic integrals and uniqueness of the martingale representation it then follows that $H^n = H \mathbb{1}_{[0,\tau_n]}$ for some progressive $H \in L^2_{\mathrm{loc}}(B)$. □

Remark 9.23.

i. The martingale representation theorem shows that every local martingale in a Brownian filtration is almost surely continuous. So the condition $\mathbb{F} = (\mathbb{F}^B)^{+,\mathbb{P}}$ is necessary, otherwise discontinuous martingales exist: consider for example $M_t = N_t - \lambda t$ for a Poisson process N with intensity $\lambda > 0$, which does not have a continuous modification.

ii. Here is another counterexample to illustrate the necessity of the condition $\mathbb{F} = (\mathbb{F}^B)^{+,\mathbb{P}}$: Let B and W be independent Brownian motions and assume that for the random variable $W_1 \in L^2$ we have

$$W_1 = W_0 + \int_0^1 H_s \mathrm{d}B_s = \int_0^1 H_s \mathrm{d}B_s$$

for some $H \in L^2(B)$. Then we get the contradiction:

$$1 = \mathbb{E}[W_1^2] = \mathbb{E}\left[W_1 \int_0^1 H_s \mathrm{d}B_s\right] = \mathbb{E}\left[\left\langle W, \int_0^\cdot H_s \mathrm{d}B_s \right\rangle_1\right] = \mathbb{E}\left[\int_0^1 H_s \mathrm{d}\underbrace{\langle W, B\rangle_s}_{\equiv 0}\right] = 0.$$

The martingale representation theorem can be used to describe the only possible effect of an equivalent change of measure on the dynamics of Brownian motion.

Corollary 9.24. *Let B be a $\mathbb{R}^d$-valued Brownian motion and let $\mathbb{F} = (\mathbb{F}^B)^{+,\mathbb{P}}$ be the (augmented) canonical filtration of B. Let $\mathbb{Q} \sim \mathbb{P}$ be an equivalent probability measure. Then there exists a unique $\mathbb{R}^d$-valued process $H \in L^2_{\mathrm{loc}}(B)$ such that*

$$Z_t = \frac{\mathrm{d}\mathbb{Q}|_{\mathcal{F}_t}}{\mathrm{d}\mathbb{P}|_{\mathcal{F}_t}} = \exp\left(\int_0^t H_s \cdot \mathrm{d}B_s - \frac{1}{2}\int_0^t |H_s|^2 \mathrm{d}s\right) \qquad \forall t \geqslant 0,$$

and under $\mathbb{Q}$ the process

$$\tilde{B} = B - \int_0^\cdot H_s \mathrm{d}s$$

is a d-dimensional Brownian motion.

Proof. We know that the density process Z is a uniformly integrable (càdlàg) $\mathbb{P}$–martingale (cf. Lemma 7.20), so in particular Z is almost surely continuous by Theorem 9.22. Moreover, $Z_t > 0$ almost surely for all $t \geqslant 0$ by Lemma 7.21, and thus there exists a stochastic logarithm $L \in \mathcal{M}_{\text{loc}}^c$ with

$$Z_t = \exp\left(L_t - \frac{1}{2}\langle L\rangle_t\right), \qquad t \geqslant 0.$$

We have $Z_0 = \mathbb{E}[Z_0] = 1$ by Blumenthal's 0-1 law, and therefore $L_0 = \log Z_0 = 0$. By the martingale representation theorem we can thus write

$$L_t - \frac{1}{2}\langle L\rangle_t = \int_0^t H_s \cdot \mathrm{d}B_s - \frac{1}{2}\int_0^t |H_s|^2\,\mathrm{d}s$$

for a unique $H \in L^2_{\text{loc}}(B)$. The final claim follows from Girsanov's theorem. □

Remark 9.25. This result shows that, in a Brownian filtration, the only possible effect of switching to an equivalent probability measure is to add an absolutely continuous drift $\int_0^\cdot H_s\,\mathrm{d}s$ to B, with derivative in $L^2([0,t])$ for all $t \geqslant 0$, i.e. such that $\int_0^t |H_s|^2\,\mathrm{d}s < \infty$. This is a (closed subspace of) the Sobolev space $W^{1,2}_{\text{loc}}(\mathbb{R}_+; \mathbb{R}^d)$. Such functions $\int_0^\cdot H_s\,\mathrm{d}s$ are sometimes called *Cameron-Martin paths*.

Appendix A Probability theory background material

A.1 Gaussian random variables

We recall here several fundamental facts about Gaussian random variables. We start by considering the one-dimensional case.

Definition A.1. (Gaussian/normal distribution) *A random variable X is called* standard Gaussian *or* standard normal, *and we write $X \sim \mathcal{N}(0,1)$, if it has probability density*

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Let $m \in \mathbb{R}$ and $\sigma \geqslant 0$. A random variable Y has the Gaussian distribution, *or* normal distribution, $\mathcal{N}(m, \sigma^2)$ *if there exists a standard normal variable Z (namely $Z \sim \mathcal{N}(0,1)$) such that*

$$Y = m + \sigma Z. \tag{A.1}$$

Equivalently, $Y \sim \mathcal{N}(m, \sigma^2)$ if and only if its characteristic function is given by

$$\mathbb{E}[e^{iuY}] = e^{ium - \sigma^2 u^2/2}, \qquad u \in \mathbb{R}. \tag{A.2}$$

A random variable Y is a centered Gaussian *(or* centered normal*) if it has distribution $\mathcal{N}(0, \sigma^2)$, namely $m = 0$.*

Exercise. Let $X \sim \mathcal{N}(0,1)$. By direct computation, show that

$$\mathbb{E}[e^{\lambda X}] = e^{\frac{\lambda^2}{2}} \quad \forall \lambda \in \mathbb{R}$$

and that

$$\mathbb{E}[e^{\lambda |X|^2}] = \begin{cases} +\infty & \text{if } \lambda \geqslant 1/2 \\ \frac{1}{\sqrt{1 - 2\lambda}} & \text{if } \lambda < 1/2 \end{cases}.$$

Let $Y \sim \mathcal{N}(m, \sigma^2)$. Recall that if $\sigma > 0$, then Y has a density

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}.$$

For $\sigma = 0$ we have $\mathbb{P}(Y = m) = 1$. Recall also that

$$m = \mathbb{E}[Y], \qquad \sigma^2 = \mathrm{Var}(Y).$$

Remark A.2.

i. If $Y \sim \mathcal{N}(m, \sigma^2)$ and $\tilde{Y} \sim \mathcal{N}(\tilde{m}, \tilde{\sigma}^2)$ are independent Gaussian random variables, then (A.2) yields that $Y + \tilde{Y} \sim \mathcal{N}(m + \tilde{m}, \sigma^2 + \tilde{\sigma}^2)$. Namely, sum of independent Gaussian variables is still Gaussian.

ii. If $Y \sim \mathcal{N}(m, \sigma^2)$, then $\lambda Y \sim \mathcal{N}(m, \lambda^2\sigma^2)$. Moreover if $Y \sim \mathcal{N}(0, \sigma^2)$, so that $Y = \sigma Z$ for some $Z \sim \mathcal{N}(0, 1)$, then for any $p > 0$ it holds:

$$\mathbb{E}[|Y|^p] = \sigma^p \, \mathbb{E}[|Z|^p] = \mathbb{E}[Y^2]^{p/2} c_p,$$

for $c_p = \mathbb{E}[|Z|^p] \in (0, \infty)$. So up to a constant c_p, the p-th absolute moment of Y is simply the second moment raised to the power $\frac{p}{2}$.

Exercise. It is not true for general centered random variables that $\mathbb{E}[|Y|^p] \leqslant C\mathbb{E}[|Y|^2]^{\frac{p}{2}}$. Can you find a centered random variable Y with $\mathbb{E}[Y^2]^{\frac{p}{2}} < \infty = \mathbb{E}[|Y|^p]$ for some $p > 2$?

The following result shows that Gaussian random variables are in some sense very rigid: the limit in distribution of Gaussian random variables still has to be Gaussian.

Lemma A.3. *Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of Gaussian random variables such that $X_n \sim \mathcal{N}(m_n, \sigma_n^2)$. Then (X_n) converges in distribution to a random variable X if and only if there exist $m \in \mathbb{R}$ and $\sigma \geqslant 0$ such that $m_n \to m$ and $\sigma_n \to \sigma$. In that case $X \sim \mathcal{N}(m, \sigma^2)$. Moreover, if (X_n) converges even in probability to X, then (X_n) also converges in L^p to X, for any $p \geqslant 1$.*

Proof. (Sketch of proof) By Lévy's continuity theorem (Stochastik 1 or Theorem 15.23 in [15]), convergence in distribution is equivalent to pointwise convergence of the characteristic function. It is easy to see that if $m_n \to m$ and $\sigma_n \to \sigma$, then the characteristic functions converge. Conversely, assume that X_n converges in distriution to some X.

Step 1: Taking the absolute value of the characteristic function, we see that σ_n^2 converges to some $\sigma^2 \in [0, \infty]$. The case $\sigma^2 = \infty$ can be ruled out because then the limit of the characteristic function would be $\mathbb{1}_{u=0}$ which is discontinuous, while any characteristic function is continuous.

Step 2: Now that we know that (σ_n) converges, we obtain that (m_n) is bounded: Otherwise along a subsequence (X_n) would converge in probability to $\pm\infty$, which is incompatible with weak convergence.

Step 3: (m_n) has at most one limit point in $\mathbb{R}$: If there were two limits m and m', then along different subsequences (X_n) would converge weakly to different limits $\mathcal{N}(m, \sigma^2)$ resp. $\mathcal{N}(m', \sigma^2)$, which is impossible because we assumed weak convergence.

It remains to show that if $X_n \xrightarrow{\mathbb{P}} X$ and $p \geqslant 1$, then even $X_n \xrightarrow{L^p} X$. From the previous considerations and because convergence in probability implies convergence in distribution we know that $m_n \to m$ and $\sigma_n^2 \to \sigma_n$. In particular, for any $q \geqslant 1$ there exists $K_q > 0$ such that

$$\sup_n \mathbb{E}[|X_n|^q] \leqslant \sup_n K_q(\mathbb{E}[|X_n - m_n|^q] + m_n^q) = \sup_n K_q(C_q \sigma_n^q + m_n^q) < \infty.$$

Taking $q=2p$, we deduce from the de la Vallée-Poussin criterion (Stochastics II) that $(|X_n|^p)$ is uniformly integrable. By another result from Stochastics II, this yields $X_n \xrightarrow{L^p} X$. □

Next, we study the $\mathbb{R}^d$-valued case, which in some sense can be reduced to the real-valued case:

Definition A.4. *Let $d \in \mathbb{N}$ and let X be a random variable with values in $\mathbb{R}^d$. Then we say that X is* (centered) Gaussian *or* (centered) normal *if for any $u \in \mathbb{R}^d$ the linear combination*

$$u \cdot X = \sum_{j=1}^d u_j X_j$$

of the entries of X is (centered) Gaussian. We also call $(X_1, \ldots, X_d)$ jointly Gaussian.

Equivalently, there exist $m \in \mathbb{R}^d$ and a symmetric positive semi-definite matrix $C \in \mathbb{R}^{d\times d}$ such that X has the characteristic function

$$\mathbb{E}[e^{iu\cdot X}] = e^{iu\cdot m - (u^T C u)/2}, \qquad u \in \mathbb{R}^d.$$

Moreover,

$$\mathbb{E}[u\cdot X] = u\cdot m, \qquad \operatorname{var}(u\cdot X) = u^T C u.$$

We say that X has mean *(or* expectation*) m and* covariance *C and write $X \sim \mathcal{N}(m, C)$. X is centered if and only if $m=0$.*

Exercise.

i. Easy: If $X = (X_1, \ldots, X_d)$ is Gaussian, show that X_j is a one-dimensional Gaussian for all $j = 1, \ldots, d$.

ii. Easy: "Linear functions of Gaussians are Gaussian": Let X be an $\mathbb{R}^d$-valued Gaussian random variable and let $A \in \mathbb{R}^{n\times d}$. Show that AX (matrix times vector) is an $\mathbb{R}^n$-valued Gaussian random variable. In particular, if $X \sim \mathcal{N}(m, C)$, then $AX \sim \mathcal{N}(Am, ACA^T)$.

iii. Hard: If X_1 and X_2 are one-dimensional Gaussian random variables, is it true that (X_1, X_2) is a two-dimensional Gaussian? Or can you find a counterexample?

If $m = 0 \in \mathbb{R}^d$ and $C = \mathbb{I} \in \mathbb{R}^{d\times d}$ is the unit matrix, X is an $\mathcal{N}(0, \mathbb{I})$ variable if and only if $(X_1, \ldots, X_d)$ are independent standard Gaussians. Indeed, for independent standard Gaussians we get

$$\mathbb{E}[e^{iu\cdot X}] = \prod_{j=1}^d \mathbb{E}[e^{iu_j X_j}] = \prod_{j=1}^d e^{-\frac{1}{2}u_j^2} = e^{-\frac{1}{2}u^T \mathbb{I} u} = e^{-\frac{1}{2}|u|^2}.$$

In particular, for any $d \in \mathbb{N}$ there exists a d-dimensional $\mathcal{N}(0, \mathbb{I})$ variable.

Lemma A.5. *Let $m \in \mathbb{R}^d$ and let $C \in \mathbb{R}^{d\times d}$ be symmetric and positive semi-definite.*

i. There exists (a probability space with) a random variable $X \sim \mathcal{N}(m, C)$: Just let $Y \sim \mathcal{N}(0, \mathbb{I})$ and

$$X := m + \sqrt{C}\, Y,$$

where $\sqrt{C}$ is the square root of C, that is the unique symmetric and positive semi-definite matrix such that $\sqrt{C}\sqrt{C} = C$ (such a $\sqrt{C}$ always exists).

ii. *"Uncorrelated Gaussians are independent": Let $X \sim \mathcal{N}(m, C)$. Then the coordinates $(X_1, \ldots, X_d)$ are independent if and only if C is a diagonal matrix.*

iii. *Let $X \sim \mathcal{N}(m, C)$. Then X has a density p_X with respect to the d-dimensional Lebesgue measure if and only if C is invertible. In that case*

$$p_X(x) = \frac{1}{(2\pi)^{d/2} (\det(C))^{1/2}} \exp\left(-\frac{1}{2}(x-m)^T C^{-1}(x-m)\right). \tag{A.3}$$

Exercise.

i. Let (X, Y) be jointly Gaussian and such that $\operatorname{cov}(X, Y) = 0$. Show that X and Y are independent. In particular, if X and Y are centered and jointly Gaussian, this implies that orthogonality is equivalent to independence.

ii. Formulate point iii. for the special case $d = 1$.

Exercise. Extend point i. of the previous exercise to the following one: if $(X_t)_{t \in \mathbb{T}}$ and $(Y_s)_{s \in \mathbb{T}'}$ are jointly Gaussian centered processes, then $(X_t)_{t \in \mathbb{T}}$ and $(Y_s)_{s \in \mathbb{T}'}$ are independent if and only if $\mathbb{E}[X_t Y_s] = 0$ for all $t \in \mathbb{T}$ and $s \in \mathbb{T}'$.

Hint: reduce to finite-dimensional distributions by Dynkin's lemma, cf. Lemma A.10.

Example A.6. Let Y, Z be independent with $Z \sim \mathcal{N}(0, 1)$ and $\mathbb{P}(Y = 1) = \mathbb{P}(Y = -1) = \frac{1}{2}$. Let $X_1 = Z$ and $X_2 = YZ$. Then $X_1 \sim \mathcal{N}(0, 1)$, and

$$\mathbb{E}[e^{iuX_2}] = \mathbb{E}[\mathbb{E}[e^{iuZY} | Y]] = \mathbb{E}[e^{-(uY)^2/2}] = e^{-u^2/2},$$

where we used that $Y^2 = 1$. So also $X_2 \sim \mathcal{N}(0, 1)$. Moreover,

$$\operatorname{cov}(X_1, X_2) = \mathbb{E}[X_1 X_2] = \mathbb{E}[YZ^2] = \mathbb{E}[Y]\mathbb{E}[Z^2] = 0 \cdot 1 = 0.$$

But of course (X_1, X_2) are not independent: Knowing X_1, we know $|X_2|$ with certainty.

Exercise. Why does this example not contradict point ii. of the previous lemma? And can you solve the hard part iii. of the long blue question right before Lemma A.5 now?

A.2 Dynkin's lemma and monotone class theorems

Let us briefly discuss two versions of the monotone class theorem, which will be useful throughout the lecture. We will often be able to verify some property for particularly simple random variables, and then ask ourselves whether the property holds for a larger class of random variables. This can often be shown with the help of monotone class theorems. We follow Appendix 4 of [7].

Let Ω be a set. We write 2^Ω for the subsets of Ω. Recall the following definition:

Definition A.7. *A family $\mathcal{D} \subset 2^\Omega$ is called a* λ-system, *or also a* Dynkin system, *if*

i. *$\Omega \in \mathcal{D}$;*

ii. *if $A, B \in \mathcal{D}$ with $A \subset B$, then $B \setminus A \in \mathcal{D}$;*

iii. *if $(A_n)_{n \in \mathbb{N}} \subset \mathcal{D}$ are increasing, i.e. $A_n \subset A_{n+1}$ for all n, then $\bigcup_n A_n \in \mathcal{D}$.*

A family $\mathcal{E} \subset 2^\Omega$ is called a π-system *if it is closed under finite intersections:*

$$A, B \in \mathcal{E} \qquad \Rightarrow \qquad A \cap B \in \mathcal{E}.$$

Exercise. (elementary, but less trivial then it looks!) Verify that $\mathcal{D}$ is equivalently a λ-system if the following hold:

1. $\Omega \in \mathcal{D}$;
2. if $A \in \mathcal{D}$, then $A^c \in \mathcal{D}$;
3. if $(A_n)_{n\in\mathbb{N}} \subset \mathcal{D}$ are pairwise disjoint, then $\bigcup_n A_n \in \mathcal{D}$.

Theorem A.8. (Dynkin's π-λ theorem, also called Dynkin's lemma) *Let $\mathcal{D} \subset 2^\Omega$ be a λ-system and let $\mathcal{E} \subset \mathcal{D}$ be a π-system. Then $\sigma(\mathcal{E}) \subset \mathcal{D}$.*

Proof. Probability I, or Theorem 4.2 from [7]. □

Definition A.9. *Let $\mathcal{A}$ be a σ-algebra on Ω. We say that $\mathcal{E}$ is a basis for $\mathcal{A}$ if $\mathcal{E}$ is a π-system and $\sigma(\mathcal{E}) = \mathcal{A}$.*

[Comment: the terminology "basis" in this context is not fully standard and not adopted by many authors, but convenient for this appendix]

Exercise. Use Dynkin's lemma to prove the following fact: if μ, ν are two probability measures on $(\Omega, \mathcal{A})$ and $\mathcal{E}$ is a basis for $\mathcal{A}$, then $\mu = \nu$ on $(\Omega, \mathcal{A})$ if and only if $\mu(A) = \nu(A)$ for all $A \in \mathcal{A}$.

Let us quickly recall some relevant cases of bases:

- On $\mathbb{R}$, the collection of closed intervals $\mathcal{R} := \{[a,b]: a < b\}$ form a basis for $\mathcal{B}(\mathbb{R})$; same for open intervals $\{(a,b): a < b\}$, or unbouded intervals $\{(-\infty, a]: a \in \mathbb{R}\}$, etc.
- On $\mathbb{R}^d$, the collection of closed rectangles $\mathcal{R}^d := \{A = \Pi_{i=1}^d A_i | A_i \in \mathcal{R}\}$ are a basis for $\mathcal{B}(\mathbb{R}^d)$.
- Given a topological space (E, τ), the collection τ of open sets form a basis for $\mathcal{B}(E)$; same for the collection of closed sets.
- If $(E_1, \mathcal{A}_1)$ and $(E_2, \mathcal{A}_2)$ are measurable spaces, $\mathcal{E}_i$ are bases for $\mathcal{A}_i$ and we endow $E_1 \times E_2$ with $\mathcal{A} = \sigma(\mathcal{A}_1 \times \mathcal{A}_2)$ the product σ-algebra, then the "rectangles"
$$\mathcal{R}^{E_1 \times E_2} := \{A = A_1 \times A_2 | A_1 \in \mathcal{E}_1, A_2 \in \mathcal{E}_2\}$$
form a basis for $\mathcal{A}$.
- Given $n \in \mathbb{N}$, $t_1, \ldots, t_n \in \mathbb{R}_+$ and $B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R})$, consider the set
$$E = \{\omega \in \mathbb{R}^{\mathbb{R}_+} : \omega(t_i) \in B_i \text{ for } i = 1, \ldots, n\};$$
let $\mathcal{E}$ denote the collection of all such sets, upon varying n, t_i and B_i. Then $\mathcal{E}$ is a basis for $\mathcal{B}(\mathbb{R})^{\otimes \mathbb{R}_+}$.

Exercise. Let X_1, X_2 be random variables taking values in two measurable spaces $(E_1, \mathcal{A}_1)$ and $(E_2, \mathcal{A}_2)$ and let $\mathcal{E}_i$ be bases for $\mathcal{A}_i$. Using Dynkin's lemma and the last point from the above bullet list, show that X_1 and X_2 are independent if and only if
$$\mathbb{P}(X_1 \in A_1, X_2 \in A_2) = \mathbb{P}(X_1 \in A_1)\mathbb{P}(X_2 \in A_2)$$
for all $A_1 \in \mathcal{E}_1$ and $A_2 \in \mathcal{E}_2$.

Lemma A.10. *Let $(X_t)_{t\in\mathbb{T}}$ and $(Y_s)_{s\in\mathbb{T}'}$ are (real-valued) stochastic processes, over (possibly different) index sets $\mathbb{T}$ and $\mathbb{T}'$. Then $(X_t)_{t\in\mathbb{T}}$ and $(Y_s)_{s\in\mathbb{T}'}$ are independent if and only if $(X_{t_i})_{i=1}^n$ and $(Y_{s_j})_{j=1}^m$ are independent, for all $n, m \in \mathbb{N}$ and all $(t_i)_{i=1}^n \subset \mathbb{T}^n$, $(s_j)_{j=1}^m \subset (\mathbb{T}')^m$.*

Exercise. Prove the above lemma.

Theorem A.8 only concerns sets, but naturally leads to results concerning *monotone* classes of functions. Such results are extremely useful whenever one wants to establish certains properties being true for a large class of functions f, upon only verifying them for simpler cases (typically $f = \mathbb{1}_A$ for some "nicely chosen" A). We present two versions of such results.

Theorem A.11. (Monotone class theorem I) *Let H be a vector space of bounded real-valued functions on Ω and $\mathcal{E}$ be a π-system. Assume that*

i. H contains all constant functions;

ii. $\mathbb{1}_A \in H$ for all $A \in \mathcal{E}$;

iii. if $(h_n)_{n\in\mathbb{N}}$ is an increasing sequence of positive functions in H such that $h := \lim_n h_n \geq 0$ is bounded, then $h \in H$.

Then H contains all bounded $\sigma(\mathcal{E})$-measurable functions.

Proof. See Theorem 4.3 from [7]. □

Exercise. Let X, Y be independent random variables taking values in $\mathbb{R}^d, \mathbb{R}^m$ respectively; let $f: \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$ be a measurable bounded function. Use Theorem A.11 to prove that

$$\mathbb{E}[f(X,Y)|Y](\omega) = g(Y(\omega)) \quad \text{for} \quad g(y) := \mathbb{E}[f(X,y)]$$

where $\mathbb{E}[\,\cdot\,|Y]$ denotes conditional expectation w.r.t. $\sigma(Y)$.

In the next statement, given a family of functions C from Ω to $\mathbb{R}$, we denote by $\sigma(C)$ the σ-algebra generated by C, i.e. the smallest σ-algebra on Ω such that $f: (\Omega, \sigma(C)) \to \mathbb{R}$ is measurable for all $f \in C$. Equivalently this is the σ-algebra generated by the sets $f^{-1}([a,b])$, for $[a,b] \subset \mathbb{R}$ and $f \in C$.

Theorem A.12. (Monotone class theorem II) *Let H be a vector space of bounded real-valued functions on Ω such that*

i. H contains all constant functions;

ii. if $(h_n)_{n\in\mathbb{N}} \subset H$ converges uniformly to h, then $h \in H$;

iii. if $(h_n)_{n\in\mathbb{N}}$ is an increasing sequence of positive functions in H such that $h := \lim_n h_n \geq 0$ is bounded, then $h \in H$.

If $C \subset H$ is closed under pointwise multiplication (that is, $fg \in C$ whenever $f, g \in C$), then H contains all bounded $\sigma(C)$-measurable functions.

Proof. See Corollary 4.4 from [7]. □

Exercise. (hard) Given a probability measure μ on $\mathbb{R}_+$, we define its Laplace transform by

$$L_\mu(a) := \int_{\mathbb{R}_+} e^{-ax} \mu(\mathrm{d}a) \quad \forall a \geqslant 0.$$

Apply Theorem A.12 to prove that L_μ characterizes uniquely μ, in the following sense: if μ and ν are two probability measures on $\mathbb{R}_+$ such that $L_\mu = L_\nu$, then $\mu = \nu$.

Bibliography

[1] Mathias Beiglböck and Pietro Siorpaes. Pathwise versions of the Burkholder-Davis-Gundy inequality. *Bernoulli*, 21(1):360–373, 2015.

[2] A. S. Cherny. On strong and weak uniqueness for stochastic differential equations. *Teor. Veroyatnost. i Primenen.*, 46(3):483–497, 2001.

[3] Giuseppe Da Prato and Jerzy Zabczyk. *Stochastic equations in infinite dimensions*, volume 152 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, Second edition, 2014.

[4] Claude Dellacherie and Paul-André Meyer. *Probabilities and potential*, volume 29 of *North-Holland Mathematics Studies*. North-Holland Publishing Co., Amsterdam, 1978.

[5] G. Di Nunno and Yu. A. Rozanov. On measurable modification of stochastic functions. *Teor. Veroyatnost. i Primenen.*, 46(1):175–180, 2001.

[6] Rick Durrett. *Probability: theory and examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, Fourth edition, 2010.

[7] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes: Characterization and convergence*. John Wiley & Sons, 1986.

[8] Franco Flandoli. *Random perturbation of PDEs and fluid dynamic models*, volume 2015 of *Lecture Notes in Mathematics*. Springer, Heidelberg, 2011. Lectures from the 40th Probability Summer School held in Saint-Flour, 2010, École d'Été de Probabilités de Saint-Flour. [Saint-Flour Probability Summer School].

[9] Peter K. Friz and Martin Hairer. *A course on rough paths*. Universitext. Springer, Cham, 2014. With an introduction to regularity structures.

[10] Peter K. Friz and Nicolas B. Victoir. *Multidimensional stochastic processes as rough paths*, volume 120 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. Theory and applications.

[11] J. van der Hoeven et al. GNU TeXmacs. `https://www.texmacs.org`, 1998.

[12] Jean Jacod. Mouvement brownien et calcul stochastique. Unpublished lecture notes, `https://www.lpsm.paris//cours/DEA-07.pdf`, 2008.

[13] Jean Jacod and Albert N. Shiryaev. *Limit theorems for stochastic processes*. Springer, 2nd edition, 2003.

[14] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*. Springer, 1988.

[15] Achim Klenke. *Probability Theory - A Comprehensive Course*. Springer, 2008.

[16] Jean-François Le Gall. *Brownian motion, martingales, and stochastic calculus*, volume 274 of *Graduate Texts in Mathematics*. Springer, 2016.

[17] Thomas M. Liggett. *Continuous time Markov processes*, volume 113 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2010. An introduction.

[18] Peter Mörters and Yuval Peres. *Brownian motion*, volume 30 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2010. With an appendix by Oded Schramm and Wendelin Werner.

[19] J. R. Norris. *Markov chains*, volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998. Reprint of 1997 original.

[20] Bernt Øksendal. *Stochastic differential equations*. Universitext. Springer-Verlag, Berlin, Sixth edition, 2003. An introduction with applications.

[21] Martin Ondreját and Jan Seidler. On existence of progressively measurable modifications. *Electron. Commun. Probab.*, 18:0, 2013.

[22] Philip E. Protter. *Stochastic integration and differential equations*. Springer, 2nd edition, 2004.

[23] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*. Springer, 3rd edition, 1999.

[24] L. C. G. Rogers and David Williams. *Diffusions, Markov processes, and martingales. Vol. 2*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000. Itô calculus, Reprint of the second (1994) edition.

[25] Pietro Siorpaes. Applications of pathwise Burkholder-Davis-Gundy inequalities. *Bernoulli*, 24(4B):3222–3245, 2018.

[26] Daniel W. Stroock. *Probability theory*. Cambridge University Press, Cambridge, Second edition, 2011. An analytic view.

[27] A. Ju. Veretennikov. Strong solutions and explicit formulas for solutions of stochastic integral equations. *Mat. Sb. (N.S.)*, 111(153)(3):434–452, 1980.

[28] Heinrich von Weizsäcker and Gerhard Winkler. *Stochastic integrals*. Advanced Lectures in Mathematics. Friedr. Vieweg & Sohn, Braunschweig, 1990. An introduction.