C8.1 Stochastic differential equations

Harald Oberhauser

These are PRELIMINARY(!) lecture notes for C8.1 "Stochastic differential equations" .
PLEASE SEND ANY typos/comments/suggestions/criticism to oberhauser@maths.ox.ac.uk
This version: Wednesday 22nd November, 2017, 20:09

Abstract. It is an understatement to say that differential equations are useful to describe quantities that evolve over time. In many applications (such as biology, finance or engineering, to name a few) these quantities can often be profoundly affected by stochastic fluctuations, noise, and randomness. The theory of stochastic differential equations provides a qualitative and quantitative understanding of the effects of such perturbations. This course is an introduction to this theory.

Contents

Acknowledgement. These notes are based on previous notes of Ben Hambly and Zhongmin Qian.

CHAPTER 1

A motivation

One of simplest cases of a stochastic perturbation is an ODE driven by a vector field μ that is affected by noise: let us model this perturbation by a sequence of random variables $N = (N_t)_{t\geq 0}$ which influence the evolution over time via another vector field σ,

$$\frac{dY_t}{dt} = \mu(Y_t) \underbrace{+\sigma(Y_t)N_t}_{\text{Noise}}. \tag{1.0.1}$$

There is a multitude of situations where one ends up with an equation of the form (1.0.1) (for example the management of risk, mathematical finance, uncertainty in biological processes, etc; we refer to [**Gardiner, 2009, Gardiner et al., 1985, Øksendal, 2003**] for a wealth of examples). Throughout this course we mainly focus on the well-posedness of (1.0.1).

There are many possible choices for the distribution of the noise process N. A common situation is that the noise N is a *white noise*, that is

(1) (independence) $\forall s \neq t$, $N(t)$ and $N(s)$ are independent,
(2) (stationarity) $\forall t_1 \leq \cdots \leq t_n$ the law of $(N(t_1+t), \cdots, N(t_n+t))$ does not depend on t,
(3) (centered) $\mathbb{E}[N_t] = 0$, $\forall t \geq 0$.

It is clear, that above properties imply that the trajectory $t \mapsto N_t$ will not be continuous[1] and therefore we cannot simply apply standard ODE results for every sample path $N(\omega)$ since the regularity of the vector fields is well-beyond classic ODE theory. We have to exploit the probabilistic structure of the noise. Putting mathematical rigour aside, let us rewrite above differential equation as an integral equation, that is we integrate (1.0.1) against time and work with $dB_t := N_t dt$. Since integration smoothes out, the process B will have continuous trajectories, $B_0 = 0$ and by 1) and 2) above it has stationary and independent increments. One can show that this already implies that B is a BM (this is not suprising considering the CLT; the "Levy–Ito decomposition theorem" makes this precise)[2]. This informal(!) argument motivates us to rewrite (1.0.1) as a stochastic integral equation

$$Y_t = \int_0^t \mu(Y_t)\,\mathrm{d}r + \int_0^t \sigma(Y_t)\,\mathrm{d}B_t. \tag{1.0.2}$$

resp. in differential notation

$$\mathrm{d}Y_t = \mu(Y_t)\,\mathrm{d}t + \sigma(Y_t)\,\mathrm{d}B_t.$$

So far we shifted the problem to give meaning to the second integral $\int_0^t \sigma(Y_r)\,\mathrm{d}B_r$. If $t \mapsto B_t$ would be a path of finite length (finite variation), we could just use classic Riemann–Stieltjes integration. But you already know that although B can be rigorously defined as a stochastic process with continuous trajectories $t \mapsto B_t(\omega)$, these trajectories are somewhat "degenerate": they are of infinite length (also called unbounded variation) and therefore excluding the use of Riemann–Stieltjes integration, highly oscillatory, statistically self-similar and possess a rich fractal structure, etc.; see Figure **??** for a two-dimensional BM sampled at different time intervals. Nevertheless it is possible to exploit the probabilistic structure of the noise and

[1]Even worse, if we assume that $\mathbb{E}[N_t^2] = 1$ then $(\omega, t) \mapsto N_t(\omega)$ is not even measurable (see [**Øksendal, 2003, Hida, 1980**]).

[2]Any continuous real-valued process X with stationary independent increments can be written as $X_t = X_0 + \mu t + B_t$ where B is a BM and μ is a constant.

develop a vast generalization of classic ODE theory that can deal with such "wild" trajectories. Qulitatively and quantitavely understanding the influence of such perturbations of is what makes stochastic differential equations such a fascinating and rich subject. You already know that there exists a beautiful and highly influential theory that gives meaning to such integrals against "degenerate" trajectories. We are going to recall this theory and start to think about the well-posedness of (1.0.2).

CHAPTER 2

Martingale theory in continuous time

In this chapter we briefly recall the basic theory of martingales in continuous time as covered for example in last year's "*B8.2 Continuous martingales and stochastic Calculus*". We refer to the lecture notes for B8.1, B8.2 [**Obloj, 2015**] or alternatively [**Revuz and Yor, 1999**] and the books in the reading list in Appendix A for a thorough treatment. We prove very few of the results in Chapter 2 and therefore many proofs in Chapter 2 are marked as **not examinable.** However, you should ***be familiar with the statements themselves and be able to apply them***: all of the other chapters rely on these obects and their properties as presented in Chapter 2, like stochastic integrals, the quadratic variation process, semimartingales, etc.

2.1. Processes, filtrations and stopping times

Fix two measurable spaces $(\Omega, \mathcal{F})$ and $(S, \mathcal{S})$. We refer to $(\Omega, \mathcal{F})$ as *sample space* and to $(S, \mathcal{S})$ as *state space*. We call a collection of random variables $X = (X_t)_{t\geq 0}$,

$$X_t : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$$

a stochastic process and we sometimes use the notation $X(t, \omega)$ for $X_t(\omega)$. In virtually all situations that we are interested in, it is justified to assume a stronger measurability requirement, namely that

$$(t, \omega) \mapsto X(t, \omega)$$

is $\mathcal{B}([0, \infty)) \times \mathcal{F}$ measurable, and we call such a stochastic process X *measurable*. Further, we refer to the index t as *time* and for every fixed $\omega \in \Omega$ we say that

$$t \mapsto X_t(\omega)$$

is the *trajectory or sample path* of X associated with ω. Obviously, two stochastic processes, $X = (X_t)$ and $Y = (Y_t)$ are the same if

$$X(t, \omega) = Y(t, \omega) \, \forall (t, \omega)$$

but it turns out that weaker concepts of "being the same" are more useful:

(1) X and Y are *indistinguishable* if $\mathbb{P}(X_t = Y_t, \forall t) = 1$.
(2) X and Y are *modifications* if $\mathbb{P}(X_t = Y_t) = 1$ for every $t \geq 0$,
(3) X and Y have the *same finite dimensionals marginals* if for every $k \in \mathbb{N}$, $t_1 < \cdots < t_k$, $A \in \mathcal{S}^{\otimes k}$

$$\mathbb{P}\left(\left(X_{t_1}, \ldots, X_{t_k}\right) \in A\right) = \mathbb{P}\left(\left(Y_{t_1}, \ldots, Y_{t_k}\right) \in A\right)$$

and in this case we also say that X and Y are *versions* of each other.

Note that 1 implies 2 and 2 implies 3. The weakest notion of the above, namely having the same finite dimensional marginals, has (unlike being a modifications or indistinguishable) an immediate extension to the case when X and Y are defined on different probability spaces, that is if $X_t : (\Omega, \mathcal{F}, \mathbb{P}) \to (S, \mathcal{S})$ and $Y_t : (\Omega', \mathcal{F}', \mathbb{P}') \to (S, \mathcal{S})$. Recall also Kolmogorov's extension theorem which guarantees that there exists a probability space and a stochastic process that has a given set of marginals (under natural consistency assumptions).

Since we usually think about the index t as time, we want to formalize the accumulation of information over time and of processes that "do not look into the future": this is done using filtrations. [1] A *filtration* of a measure space (in our case the *sample space*) $(\Omega, \mathcal{F})$ is a nondecreasing family $\{\mathcal{F}_t : t \geq 0\}$ of sub-σ-fields of $\mathcal{F}$, i.e. $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$ for all $s \leq t$. Given a stochastic process X and a filtration $(\mathcal{F}_t)$ we say that X *is adapted* to $(\mathcal{F}_t)$ if for every $t \geq 0$, X_t is $\mathcal{F}_t$-measurable. We call the smallest filtration for which this is true the *filtration generated by* X and denote it by

$$\mathcal{F}_t := \sigma\left(X_s : 0 \leq s \leq t\right).$$

Given a filtration, we can speak of past and future by setting

$$\mathcal{F}_{t-} := \sigma\left(\bigcup_{s<t} \mathcal{F}_t\right), \mathcal{F}_{t+} := \sigma\left(\bigcap_{s>t} \mathcal{F}_s\right) \text{ and } \mathcal{F}_{0-} := \mathcal{F}_0.$$

Obviously $\mathcal{F}_{t-} \subset \mathcal{F}_t \subset \mathcal{F}_{t+}$ and in general these inclusions are strict. Note that $\sigma\left(\bigcup_{s<t} \mathcal{F}_{t-}\right) = \sigma\left(\bigcup_{s<t} \mathcal{F}_t\right) = \sigma\left(\bigcup_{s<t} \mathcal{F}_{t+}\right)$ and we denote this σ-algebra with $\mathcal{F}_\infty$. We call a filtration *right-continuous* if $\mathcal{F}_t = \mathcal{F}_{t+}$ *[resp. resp. left-continous if* $\mathcal{F}_t = \mathcal{F}_{t+}$*]*. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a filtration $(\mathcal{F}_t)$, we denote this as $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ and refer to a filtered probability space. We say that$(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ [or the filtration $(\mathcal{F}_t)$] satisfies the *usual conditions* if $(\mathcal{F}_t)$ is right-continuous and $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets in $\mathcal{F}$.

CONVENTION 1. Unless stated otherwise, we always assume that the usual conditions hold.[2]

Given a filtration $(\mathcal{F}_t)$ we can use this to gradually define smaller classes of processes: X is *progressively measurable with respect to* $(\mathcal{F}_t)$ if for each $t \geq 0$

$$(s, \omega) \mapsto X_s(\omega) \text{ is } \left(\mathcal{B}([0, t]) \otimes \mathcal{F}_t\right) - \text{measurable}$$

for all $(s, \omega) \in [0, t] \times \Omega$. Every measurable process that is adapted to $(\mathcal{F}_t)$ has a modification that is progressively measurable with respect to $(\mathcal{F}_t)$. Using the progressively measurable version allows to avoid some measure theoretic trouble. A stronger requirement is that of predictability: X is predictable if

$$(s, \omega) \mapsto X_s(\omega) \text{ is } \mathcal{P} - \text{measurable}$$

where $\mathcal{P}$ denotes the σ-algebra on $[0, \infty) \times \Omega$ that is generated by the space of continuous and $(\mathcal{F}_t)$-adapted processes. To sum up,

$$\text{predictable} \subset \text{progressively measurable} \subset \text{adapted} \subset \text{measurable}.$$

Using filtrations we can also speak of random and stopping times: we a call any random variable

$$\tau : (\Omega, \mathcal{F}) \to ([0, \infty], \mathcal{B}([0, \infty]))$$

a *random time*. We call a random time a *stopping time* if for every (t, ω)

$$\tau^{-1}(\omega) \in \mathcal{F}_t$$

We say that a property π holds *locally* for a process X if there exists a sequence of stopping times, (τ_n) such that $\tau_n \to \infty$ a.s. and such that the stopped process $\left(X_{\tau_n \wedge t} 1_{\tau_n > 0}\right)_t$ has the property π for every $n \geq 1$. We call a family of random variables $\{Z_\alpha\}_{\alpha \in \mathcal{I}}$ indexed by some set $\mathcal{I}$ *uniformly integrable* if

$$\sup_{\alpha \in \mathcal{I}} \mathbb{E}\left[|Z_\alpha| 1_{|Z_\alpha| > n}\right] \to 0 \text{ as } n \to \infty.$$

(in the case of stochastic process, $\mathcal{I} = [0, \infty)$).

Finally, recall a common convention:

[1] We use the word "information" very loosely here but one way to make precise how a σ-algebra contains "information" is the Doob–Dynkin lemma: given two real-valued random variables U, V on $(\Omega, \mathcal{F})$ it holds that V is $\sigma(U)$-measurable if and only $V = f(U)$ for a $\mathcal{B}(\mathbb{R})$-measurable function f.

[2] This is justified since we can replace $\mathcal{F}_t$ by $\sigma\left(\mathcal{F}_{t+}, \{N \in \mathcal{F} : \mathbb{P}(N) = 0\}\right)$ to get a filtration that satisfies the usual conditions.

REMARK 2.1. Fix a measure space (S, Σ, μ) and denote the space of all real-valued Lebesgue measurable functions such that

$$|f|_p := \left(\int_S |f(x)|^p \, \mu(dx) \right)^{1/p} < \infty$$

with $\mathcal{L}^p(S, \Sigma, \mu)$. This is a linear space but $|.|_p$ is only a seminorm since it is invariant under modification of f on μ-nullsets. By using the quotient space

$$L^p(S, \Sigma, \mu) := \mathcal{L}^p(S, \Sigma, \mu) / \ker\left(|.|_p\right)$$

we arrive at normed space.

CONVENTION 2. It is custom to not be pedantic about such distinctions between an equivalence class and a member of this equivalence class and unless stated otherwise we follow this convention.

For example, we use the same notation for what are strictly speaking two different objects; we refer to a process when it is actually an equivalence class of indistinguishable processes in the same way as we refer to a function L^2 although it is an equivalence class of functions in $\mathcal{L}^p$

2.2. (Local) Martingales

Throughout the rest of Chapter 2 we fix a filtered probability space $(\Omega, \mathbb{P}, (\mathcal{F}_t) \mathcal{F})$.

DEFINITION 2.2. We call an $(\mathcal{F}_t)$-adapted process $M = (M_t)_{t \geq 0}$ such that $M_t \in L^1(d\mathbb{P})$ $\forall t \geq 0$ a

(1) a *submartingale* if $\mathbb{E}[M_t | \mathcal{F}_s] \geq M_s$,
(2) a *supermartingale* if $\mathbb{E}[M_t | \mathcal{F}_s] \leq M_s$,
(3) a *martingale* if $\mathbb{E}[M_t | \mathcal{F}_s] = M_s$,

holds (a.s.) for all $s < t$. If the sample paths $t \mapsto M_t$ are right-continuous, then we say that M is a *right-continuous* [sub/super] martingale; similary we speak about left-continuous or continuous [sub/super] martingales. We denote the space of martingales with $\mathcal{M}$, the subspace of continuous martingales with $\mathcal{M}_c$.

REMARK 2.3.

- $\mathcal{M}$ and $\mathcal{M}_c$ are vector spaces under pointwise addition and scalar multiplication of sample paths.
- If we want to emphasize with respect to which probability measure and filtration a process is a martingale, we say that it is $((\mathcal{F}_t), \mathbb{P})$-martingale or $(\mathcal{F}_t)$-martingale. If M is $((\mathcal{F}_t), \mathbb{P})$-martingale then it is also a $((\mathcal{G}_t), \mathbb{P})$-martingale under every filtration such that M is $(\mathcal{G}_t)$-adapted and $\mathcal{G}_t \subset \mathcal{F}_t$; especially this applies to $(\mathcal{G}_t) = \sigma(M)$. The converse is not true: if $\mathcal{F}_t \subset \mathcal{G}_t$ then there is no reason why M should be a $((\mathcal{G}_t), \mathbb{P})$-martingale. (The same applies to [sub/super] martingales).
- The prefixes sub- and super- come from a deep connection with sub- and superharmonic functions. Rogers & Williams give the useful mnemonic that the "b" in "sub" points upwards, and the "p" in "sup" downwards reflecting that the conditional expectation of a submartingale increases/a supermartingale decreases.

Recall that we π holds *locally* for a process X if there exists a sequence of stopping times, (τ_n) such that $\tau_n \longrightarrow \infty$ a.s. and such the stopped process $\left(X_{\tau_n \wedge t} 1_{\tau_n > 0}\right)_t$ has the property π for every $n \geq 1$. In the context of martingales this leads to so-called local martingales. This generalization of martingales is extremely useful: for example it frees us from worrying about integrability and allows to deal with processes that are only defined up to a stopping time.

DEFINITION 2.4. We call an adapted, right-continuous stochastic process $X = (X_t)_{t \geq 0}$ a *local martingale* if there exists a sequence of stopping times $(\tau_n)_n$ such that

$$\tau_n \leq \tau_{n+1} \text{ and } \lim_n \tau_n = \infty \text{ a.s.}$$

and such that

$$X^{\tau_n} 1_{\tau_n > 0} = \left(X_{t\wedge\tau_n} 1_{\tau_n>0}\right)_{t\geq 0}$$

is a sequence of martingales. We denote with $\mathcal{M}_{\text{loc}}$ the space of local martingales, and with $\mathcal{M}_{\text{c,loc}}$ the subspace of continuous local martingales.

Remark 2.5.

- The purpose of the factor $1_{\tau_n>0}$ is to deal with the case when X_0 is not integrable, thus when X_0 is constant we can ignore it.
- By replacing the stopping times τ_n with $\tau_n' = \tau_n \wedge n$ we get another sequence of stopping times (τ_n') such that $X^{\tau_n'}$ is u.i. (Therefore some textbooks include u.i. in above in the definition of a local martingale).
- If X is continuous we can use $\tau_n' = \tau_n \wedge \inf\{t : |X_t| = n\}$ to work with a sequence $\left(X^{\tau_n'}\right)$ of bounded martingales.

Any martingale is a local martingale (using for example $\tau_n = n$) but many important processes are only local martingales but not martingales — we recall such examples below. Many proofs that work for martingales have an immediate generalisation to local martingales by repeating a localized version of the argument. Let us first recall one of the most important examples of a martingale, namely Brownian motion.

Definition 2.6. We call a real-valued, stochastic process $B = (B_t)_{t\geq 0}$ defined on a probability space $(\Omega, \mathbb{P}, \mathcal{F})$ and adapted to a filtration $(\mathcal{F}_t)_{t\geq 0}$ a *one-dimensional Brownian motion started at* 0 *with respect to the filtration* $(\mathcal{F}_t)$ if

(1) $B_0 = 0$ and $t \mapsto B_t$ is continuous (a.s.),
(2) for $s \leq t$, $B_t - B_s$ is independend of $\mathcal{F}_s$,
(3) for $s \leq t$, $B_t - B_s \sim \mathcal{N}(0, t-s)$.

Last year, it was shown to you that the above is non-vacous: there exists a probability space that carries a Brownian motion (you might have seen several different ways to prove this). Already in the Brownian case, local martingales appear naturally under nonlinear transformations.

Example 2.7. Let B be a two-dimensional Brownian motion. Then

$$(\log|B_t|)_t$$

is a local martingale but not a martingale since $\mathbb{E}[\log|B_t|] \to \infty$ as $t \to \infty$.

Remark 2.8 (WARNING). One might be tempted to think of local martingales as martingales that just miss an integrability property. This is wrong! For example

$$\log|B_t| \in L^p \forall p < \infty, \forall t \geq 0$$

but $t \mapsto \mathbb{E}[\log|B_t|]$ is not constant. Local martingales are truly much more general than martingales.

Given a local martingale M, often it is useful to know if M is actually a martingale.

Proposition 2.9. *Let $M \in \mathcal{M}_{c,loc}$ and $M_t^\star \in L^1(\mathbb{P})$ $\forall t \geq 0$. Then $M \in \mathcal{M}_c$.*

Proof. Since $M_t \leq M_t^\star$ we have $M_t \in L^1(\mathbb{P})$ so it just remains to verify $\mathbb{E}[M_t|\mathcal{F}_s] = M_s$. Since $M^{\tau_n} \in \mathcal{M}_{\text{c,loc}}$ we have

$$\mathbb{E}\left[M_t^{\tau_n}|\mathcal{F}_s\right] = M_s^{\tau_n}$$

Since $M_t^\star \in L^1(\mathbb{P})$ we can apply dominated convergence and let n to ∞. □

Corollary 2.10. *A bounded, continuous local martingale is a continuous martingale.*

2.3. Inequalities, regularity, convergence and optional stopping

It often helps to think of sub- and supermartingales as randomized versions of decreasing and increasing sequences. This motivates us to look for "randomized versions" of classic results from analysis about monotone sequences, for example convergence. A powerful tool to do this are Doob's inequalities. Throughout this section recall that any statement about a supermartingale M can be turned into a statement about the submartingale $-M$. We use the notation

$$X_t^\star = \sup_{s\in[0,t]} |X_s|$$

THEOREM 2.11 (DOOB'S MAXIMAL INEQUALITIES). *Let M be right-continuous, a martingale or a positive submartingale. Then for every $t>0$, $\lambda>0$ and every*

- *$p\geq 1$ we have*
$$\lambda^p \mathbb{P}\left(M_t^\star \geq \lambda\right) \leq |M_t|_{L^p}^p,$$
- *$p>1$ we have*
$$\left|M_t^\star\right|_{L^p} \leq \frac{p}{p-1} |M_t|_{L^p}.$$

PROOF. **(PROOF NOT EXAMINABLE)**The result is first established in discrete time, then we choose an increasing squence of finite subsets of $[0,t]$ to approximate the continuous time case. The details are given in [**Obloj, 2015**]. □

Since $|M_t|\leq M_t^\star$ we get for every $p>1$, that $M_t^\star$ is L^p-bounded iff M has a uniform L^p-bound over $[0,t]$. For $p=2$ this becomes a key ingredient for stochastic integration theory.

COROLLARY 2.12. *Let M be a right-continuous martingale. Then*

$$\left|M_t^\star\right|_{L^2} < \infty \text{ if and only if } \sup_{s\in[0,t]} |M_t|_{L^2} < \infty.$$

If the above holds, then $M^t = (M_s)_{s\in[0,t]}$ is u.i.

As in the proof of Doob's maximal inequalities a common strategy is to first derive a result in in discrete time and then approximate the process in continuous time. This requires that sample paths can be well-approximated by such discrete approximations (like right continuous trajectories). Luckily, under weak conditions we can work with a modification that is right-continuous.

THEOREM 2.13. *Let M be a submartingale[3]. If*

$$t \mapsto \mathbb{E}[X_t]$$

is right-continuous then M has a modification that is a $(\mathcal{F}_t)$ submartingale with cadlag[4] paths. This modification is unique (up to indistinguishability).

PROOF. **(PROOF NOT EXAMINABLE)** □

Especially, every martingale has a cadlag modification.

CONVENTION 3. From now onwards, unless otherwise stated, we only consider cadlag [sub-/super-] martingales (without explicitly stating that they are cadlag).

We can now state the first convergence result.

THEOREM 2.14. *Let X be a submartingale. If $\sup_t \mathbb{E}\left[X_t^+\right] < \infty$ then*

$$X_\infty := \lim_{t\to\infty} X_t \text{ exists a.s. and } |X_\infty|_{L^1} < \infty.$$

[3]Recall our convention that we always assume the usual conditions. This theorem is even true if $(\mathcal{F}_t)$ is only right-continuous
[4]**c**ontinue **à d**roite**, l**imite **à g**auche (right-continuous with left limit).

Proof. **(PROOF NOT EXAMINABLE)**. Follows by using Doob's upcrossing inequalities (which we do not recally here) in discrete time and use the usual limit to get the result for the continuous time case. See details in [**Obloj, 2015**]. □

If a [sub-/super-] martingale converges to a limit M_∞ we want to add $t = \{\infty\}$ as time and study M via the conditional expecation $\mathbb{E}[M_\infty|\mathcal{F}_t]$.

Theorem 2.15. *Let $M \in \mathcal{M}$. The following are equivalent,*

(1) $\lim_{t\to\infty} M_t$ *converges in* L^1,
(2) *there exists a* $M_\infty \in L^1$ *such that* $M_t = \mathbb{E}[M_\infty|\mathcal{F}_t]\ \forall t$,
(3) M *is uniformly integrable.*

If the above holds, then

$$M_t \to_{t\to\infty} M_\infty \ a.s.$$

Moreover, above conditions are all satisfied if $\sup_t |M_t|_{L^p(\Omega,\mathcal{F},\mathbb{P})} < \infty$ *holds for a* $p > 1$. *The analoguous statement holds to super [sub-] martingales by replacing the equality in 2 with ≤[resp. ≥].*

Proof. **(PROOF NOT EXAMINABLE).** □

An important consequence is that a L^p bounded martingale is a uniformly integrable martingale.

We emphasize that above Theorem shows that a path-valued (thus infinite dimensional) random variable, namely the u.i. martingale $M = (M_t)_{t\geq 0}$, is determined by a real-valued random variable and a filtration, namely $((\mathcal{F}_t), M_\infty)$. This identification is a key ingredient for stochastic integration theory. Another very useful property of u.i. martingales is that the martingale property is robust under stopping times.

Theorem 2.16 (Optional stoppping theorem). *Let $X \in \mathcal{M}$ and u.i. Then*

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

for all stopping times σ, τ with $\sigma \leq \tau$.

Proof. **(PROOF NOT EXAMINABLE).** □

Remark 2.17. The u.i. assumption is not a technicality. For example if $\tau_x := \inf\{t : B_t = x\}$ then

$$b \equiv \mathbb{E}\left[B_{\tau_b}|\mathcal{F}_{\tau_a}\right] \neq B_{\tau_a} \equiv a \text{ for any } a < b.$$

(Thus BM is not a u.i. martingale).

2.4. From finite to quadratic variation

Given a path $\gamma : \mathbb{R}_+ \to \mathbb{R}$ we can measure its length on the interval $[0,t]$ as follows: for every finite partition $\pi = \{0 = t_0 < \cdots < t_n = t\}$ the length of the path that is the linear interpolation of the points $(\gamma_0, \gamma_{t_1}, \ldots, \gamma_{t_n})$ is clearly

$$\sum_{t_i\in\pi\cap[0,t]} \left|\gamma_{t_{i+1}} - \gamma_{t_i}\right|.$$

Thus it is natural to definte the length (henceforth referred to as total variation) of γ as

$$\sup_\pi \sum_{t_i\in\pi\cap[0,t]} \left|\gamma_{t_{i+1}} - \gamma_{t_i}\right|.$$

with $\sup_\pi$ is taken over all finite partitions π.

DEFINITION 2.18. Let A be an adapted cadlag process such that for all $t \geq 0$

$$\sup_{\pi} \sum_{t_i \in \pi \cap [0,t]} \left| A_{t_{i+1}} - A_{t_i} \right| < \infty, \tag{2.4.1}$$

with the $\sup_\pi$ is taken over all finite partitions $\pi = \{0 = t_0 < \cdots < t_n < \infty\}$. We call (2.4.1) the total variation of A on the interval $[0,t]$, and denote the set of adapted cadlag processes with finite total variation with BV.

Note that (2.4.1) does not have to converge, even if γ is continuous! Last year you saw that if γ is a typical realization of Brownian motion,$\gamma = B(\omega)$ for a.e. ω, then

$$\sup_{\pi} \sum_{t_i \in \pi \cap [0,t]} \left| B_{t_{i+1}} - B_{t_i} \right| = \infty$$

for any sequence of partitions (π_n) with $\mathrm{mesh}(\pi_n) \to_n 0$. Hence a Brownian particle travels an infinite distance over a finite interval $[0,t]$ so it moves with "infinite speed". However, we can say a bit more by recalling that BM is α-Holder continuous, i.e.

$$\sup_{s \neq t} \frac{\left| B_{t_{i+1}} - B_{t_i} \right|}{|t-s|^{\alpha}} < \infty \ \forall \alpha \in \left(0, \frac{1}{2}\right)$$

and as consequence

$$\sup_{\pi} \sum_{t_i \in \pi \cap [0,t]} \left| B_{t_{i+1}} - B_{t_i} \right|^{1/\alpha} < \infty \ \forall \alpha \in \left(0, \frac{1}{2}\right)$$

Thus we found a more more useful "inner clock" for the evolution of a Brownian particle, namely taking the $\frac{1}{\alpha}$-th power of its space increments. It turns out that if we switch from a $\sup_\pi$ over all partitions π to a $\lim_n$ along a sequnce of partitions with vanishing mesh, we can even take $\alpha = \frac{1}{2}$. Moreover, above generalizes to local martingales.

THEOREM 2.19. *Let $M \in \mathcal{M}_{c,loc}$. Then there exists a non-decreasing, continuous adapted process*

$$\langle M \rangle$$

such that $\langle M \rangle_0 = 0$ and

$$M^2 - \langle M \rangle \in \mathcal{M}_{c,loc}.$$

This process is unique (up to indistinguishability). Furthermore, for any interval $[0,t]$ and any sequence of partitions $(\pi_n)_n$

$$\pi_n = \left\{0 = t_0^n < t_1^n < \cdots < t_{m_n}^n = t\right\}$$

with vanishing mesh, that is

$$\sup_{i:t_{i-1},t_i \in \pi_n} \left| t_i^n - t_{i-1}^n \right| \to 0 \text{ as } n \to \infty,$$

we have that

$$\langle M \rangle_t = \lim_n \sum_{i=1}^{m_n} \left(M_{t_i} - M_{t_{i-1}} \right)$$

where the limit is taken in probability. We call the process $\langle M \rangle$ the quadratic variation or bracket of M.

PROOF. **(PROOF NOT EXAMINABLE)** □

REMARK 2.20.

- Every increasing path is of finite total variation, thus $\langle M \rangle \in$ BV.
  - For Brownian motion $\langle B \rangle_t = t$ but in general $\langle M \rangle$ is not deterministic and it is often difficult to find an explicit expression for $\langle M \rangle$.

We motivated the bracket process by speaking about lenght/speed of a particle following the trajectory of a process. There are other ways to motivate the bracket - first let us generalize the bracket via polarisation to a bilinear operator.

DEFINITION 2.21. Given $X,Y \in \mathcal{M}_{\text{c,loc}}$ we define their *cross-variation process* as

$$\langle X,Y\rangle_t := \frac{1}{4}\left(\langle X+Y\rangle_t - \langle X-Y\rangle_t\right).$$

(Note that $\langle X,X\rangle = \langle X\rangle$). As above we can characterize the process $\langle X,Y\rangle$.

THEOREM 2.22. *Let $M,N \in \mathcal{M}_{c,loc}$. Then there exists a continuous, adapted process of finite variation denoted*

$$\langle M,N\rangle,$$

such that $\langle M,N\rangle_0 = 0$ and

$$MN - \langle M,N\rangle \in \mathcal{M}_{c,loc}.$$

This process is unique (up to indistinguishability). Furthermore, for any interval $[0,t]$ and any sequence of partitions $(\pi_n)_n$

$$\pi_n = \left\{0 = t_0^n < t_1^n < \cdots < t_{m_n}^n = t\right\}$$

with vanishing mesh, that is

$$\sup_i \left|t_i^n - t_{i-1}^n\right| \to 0 \text{ as } n \to \infty,$$

we have that

$$\langle M\rangle_t = \lim_n \sum_{i=1}^{m_n} \left(M_{t_i} - M_{t_{i-1}}\right)\left(N_{t_i} - N_{t_{i-1}}\right)$$

where the limit is taken in probability. We call the process $\langle M,N\rangle$ the quadratic covariation or bracket of M and N.

PROOF. **(PROOF NOT EXAMINABLE)** For example [**Revuz and Yor, 1999**, Chapter 4.1] □

Recall that the covariance of two random variables U,V that are centred, $\mathbb{E}[U] = \mathbb{E}[V] = 0$ is simply

$$\operatorname{Cov}(U,V) = \mathbb{E}[UV]. \tag{2.4.2}$$

Moreover, one has the simple estimate

$$\operatorname{Cov}(U,V) \le \sqrt{\operatorname{Var}(U)}\sqrt{\operatorname{Var}(V)}. \tag{2.4.3}$$

A (local) martingale has clearly centred increments so from this point of view $\langle M,N\rangle$ generalizes above covariation (2.4.2) since informally(!)

$$\begin{aligned} \text{"}\mathbb{E}[\langle M,N\rangle_t] &= \lim_n \sum_{t_i^n \in \pi \cap [0,t]} \mathbb{E}\left[\left(M_{t_{i+1}^n} - M_{t_i^n}\right)\left(N_{t_{i+1}^n} - N_{t_i^n}\right)\right] \\ &= \int_0^t \operatorname{Cov}(dM_s, dN_s)\text{."} \end{aligned}$$

Similarly, the Kunita–Watanabe inequality generalizes (2.4.3) to pathspace.

THEOREM 2.23 (KUNITA–WATANABE INEQUALITY). *Let $M,N \in \mathcal{M}_{c,loc}$ and H,K be measurable processes. Then*[5] *for all $t \ge 0$*

$$\int_0^t H_s K_s \left|d\langle M,N\rangle\right|_s \le \sqrt{\int_0^t H_s^2 \left|d\langle M\rangle\right|_s}\sqrt{\int_0^t K_s^2 \left|d\langle N\rangle\right|_s}.$$

---

[5] If $A \in \mathrm{BV}$ then $\frac{dA_t}{dt}$ exists $d\mathbb{P} \times d\lambda$-a.e. ($d\lambda$ denotes Lebesgue measure). Hence, for $\mathbb{P}$-a.e. ω we can define new measure on $(\mathbb{R}, \mathcal{B}([0,\infty)))$ $|dA_t(\omega)| := \left|\frac{dA_t(\omega)}{dt}\right| dt$. By the above $\langle M,N\rangle \in \mathrm{BV}$.

Moreover, for $\frac{1}{p}+\frac{1}{q}=1$

$$\mathbb{E}\left[\int_0^\infty H_s K_s \left|d\langle M,N\rangle\right|_s\right] \leq \left|\sqrt{\int_0^t H_s^2 \left|d\langle M\rangle\right|_s}\right|_{L^p} \left|\sqrt{\int_0^t K_s^2 \left|d\langle N\rangle\right|_s}\right|_{L^q}.$$

Proof. **(PROOF NOT EXAMINABLE)** For example [**Revuz and Yor, 1999**, Chapter 4.1] or [**Obloj, 2015**] □

We have seen several motivations for the bracket process (via the length of a path, the process which turns M^2 into a local martingale, as generalization of covariance) and we will see yet another way to think about the bracket process when we recall Ito's formula.

2.5. Ito's stochastic integral

We now give meaning to stochastic integrals

$$\int_0^t K_s dM_s. \tag{2.5.1}$$

A first naive try is to define the above integral, trajectory by trajectory via Riemann–Stieltjes integration

$$\int_0^t K_s(\omega)\, dM_s(\omega) \equiv \lim \sum_{t_i \in \pi(n)} K_{t_i}(\omega)\left(M_{t_{i+1}\wedge t}(\omega) - M_{t_{i+1}\wedge t}(\omega)\right)$$

But already if M is a Brownian motion we see that this is doomed to fail since Brownian motion is of unbounded variation and Riemann–Stieltjes integration relies heavily on the assumption of bounded variation (check your old notes if you do not believe this!). Ito realized that despite the breakdown of classic integration theory, one can develop a general theory of integration for (2.5.1) by

(1) restricting integrators M to (semi)martingales;
(2) restricting integrands K to those that "do not look into the future" of the integrator.

We will see below that the combination of (1) and (2) allows to exploit probabilistic cancellations in the Riemann sums using the martingale inequalities and gives shows well-posedness of (2.5.1).

2.5.1. A nice class of integrators: the space of continuous, L^2-bounded martingales H^2.

Theorem 2.24. *Denote with* $\mathbb{H}^2$ *the space of* L^2*-bounded martingales*[6],

$$\mathbb{H} \quad := \quad \left\{M \in \mathcal{M} : \sup_t |M_t|_{L^2(\mathbb{P})} < \infty\right\}$$

and with H^2 *the subspace of continuous and* L^2*-bounded martingales*

$$H^2 := \mathbb{H}^2 \cap \mathcal{M}_c.$$

Then

(1) $\mathbb{H}^2$ *is isomorphic to* $L^2(\Omega, \mathcal{F}_\infty, \mathbb{P})$ *and the isomorphism is given as*

$$\begin{aligned} (M_t)_{t\geq 0} &\mapsto \lim_{t\to\infty} M_t, \\ M_\infty &\mapsto (\mathbb{E}[M_\infty|\mathcal{F}_t])_{t\geq 0}. \end{aligned}$$

It follows that $\mathbb{H}^2$ *is a Hilbert space with inner product*

$$\langle M,N\rangle_{\mathbb{H}^2} = \mathbb{E}[M_\infty N_\infty] \text{ and norm } |M|_{\mathbb{H}^2} = \sqrt{\mathbb{E}\left[M_\infty^2\right]}.$$

(2) H^2 *is closed in* $\left(\mathbb{H}^2, |.|_{\mathbb{H}^2}\right)$*, hence again Hilbert.*

[6]Recall that by Convention 2, we strictly consider equivalence classes of indistinguishable $\mathcal{L}^2(d\mathbb{P})$-bounded martingales. Recall also Convention 3, that throughout we refer to cadlag martingales.

We also refer to $\mathbb{H}^2$ [resp. H^2] as the space of square integrable or L^2-bounded [and continuous] martingales.

Proof. We claim that $\mathbb{H}^2$ is isomorphic to $L^2(\Omega, \mathcal{F}_\infty, \mathbb{P})$. To see this let $M_\infty \in L^2(\Omega, \mathcal{F}_\infty, \mathbb{P})$, then the process

$$(\mathbb{E}[M_\infty|\mathcal{F}_t])_t$$

is in $\mathbb{H}^2$ since $|M_t|_{L^2(d\mathbb{P})} = \mathbb{E}\left[\mathbb{E}[M_\infty|\mathcal{F}_t]^2\right] \leq \mathbb{E}\left[\mathbb{E}\left[M_\infty^2|\mathcal{F}_t\right]\right] = \mathbb{E}\left[M_\infty^2\right]$. On the other hand, given $M \in \mathbb{H}^2$ we get by Doob that $M_\infty^\star \in L^2(\Omega, \mathcal{F}_\infty, \mathbb{P})$, therefore M is u.i. and $M_t = \mathbb{E}[Z|\mathcal{F}_t]$ for some $Z \in L^2(\Omega, \mathcal{F}_\infty, \mathbb{P})$ by Theorem 2.15.

It remains to see the closedness of H^2. Therefore consider a Cauchy sequence $(M^n) \subset H^2$ converging to some $M \in \mathbb{H}^2$. By Doob

$$\mathbb{E}\left[\sup\left|M_t^n - M_t\right|^2\right] \leq 4\left|M^n - M\right|_{\mathbb{H}^2} \to_n 0.$$

Since L^2 convergence implies a.s. convergence along a subsequence, there must exists a subsequence (n_k) such that

$$\sup_t\left|M_t^n - M_t\right| = 0 \ a.s.$$

□

In our applications the integrand will always be a continuous process. In this case we have a nice characterization of H^2.

Proposition 2.25. *Let $M \in \mathcal{M}_{c,loc}$. Then $M \in H^2$* iff

(1) $M_0 \in L^2(d\mathbb{P})$,
(2) $\mathbb{E}[\langle M\rangle_\infty] < \infty$.

In this case, $M^2 - \langle M\rangle$ is u.i. and for all stopping times $\sigma \leq \tau$,

$$\mathbb{E}[(M_\tau - M_\sigma)|\mathcal{F}_\sigma] = \mathbb{E}[\langle M\rangle_\tau - \langle M\rangle_\sigma|\mathcal{F}_\sigma].$$

Especially, if $M \in H_0^2$ it follows that

$$|M|_{\mathbb{H}^2} = |M_\infty|_{L^2(d\mathbb{P})} = \left|\langle M\rangle_\infty^{1/2}\right|_{L^2(d\mathbb{P})}.$$

Proof. **(PROOF NOT EXAMINABLE)** For example, see [**Revuz and Yor, 1999**, Chapter 4.1] or [**Obloj, 2015**] □

2.5.2. A nice class of integrands for every integrator $M \in H^2$: the space $L^2(M)$. Firstly, we restrict ourselves to integrands that are progressively measurable. Secondly, we expect two stochastic integrals $\int_s^t H_r dM_s$ and $\int_s^t G_r dM_r$ to coincide if M does not change on the interval $[s,t]$ (both should equal 0). Recalling that the bracket measures the "accumulated infinitesimal changes" of M, we take this as motivation for the definition of the so-called "Doleans measure" on $([0,\infty)\times\Omega, \mathcal{B}([0,\infty))\times\mathcal{F}_\infty)$,

$$\begin{aligned}\mu_M((s,t]\times A) &:= \mathbb{E}[\langle M\rangle_t - \langle M\rangle_s; A] \\ &= \mathbb{E}\left[\int_s^t 1_A(\omega)\, d\langle M\rangle_r\right].\end{aligned}$$

Thus for every integrator $M \in H^2$ we have identified a nice candidate for a class of integrands, namely the subspace of progressively measurable processes in the Hilbert space

$$L^2([0,\infty)\times\Omega, \mathcal{B}([0,\infty))\otimes\mathcal{F}_\infty, \mu_M)$$

The measure μ_M does not see a difference between processes H, G that only differ from each other on intervals where M does not change. From the definition of μ_M we see that

$$\begin{aligned} |H|_{L^2}^2 &= \int_0^\infty |H(t,\omega)|^2 \, d\mu_M(t,\omega) \\ &= \mathbb{E}\left[\int_0^\infty H_s^2 d\langle M\rangle_s\right]. \end{aligned}$$

Definition 2.26. For $M \in \mathbb{H}^2$ denote with $L^2(M)$ the Hilbert space of progressively measurable processes[7] with Hilbert norm

$$|H|_{L^2(M)}^2 := \mathbb{E}\left[\int_0^\infty H_s^2 d\langle M\rangle_s\right] < \infty.$$

Remark 2.27. $L^2(M)$ includes all bounded and cadlag, adapted processes though in our applications we will mostly deal with continuous bounded processes.

2.5.3. Construction of the stochastic integral. Let's make an educated guess what our stochastic integral should look like: if the integrand is especially simple, namely piecewise constant

$$K_t = K_{-1} 1_{\{0\}}(t) + \sum_{i\geq 0} K_i 1_{(t_i,t_{i+1}]}(t) \text{ for some } 0 = t_0 < t_1 < \cdots \text{ with } \lim_i t_i = \infty$$

with $K_i \in \mathcal{F}_{t_i}$, all K_i bounded uniformly in i, then we obviously want that

$$\int_0^t K dM \equiv \sum_{i\geq 1} K_i \left(M_{t_{i+1}\wedge t} - M_{t_i\wedge t}\right).$$

By linearity of cross covariation we get for every $N \in H^2$ that

$$\begin{aligned} \left\langle \int_0^{\cdot} K dM, N \right\rangle_t &= \sum_i K_i \left\langle M_{t_{i+1}\wedge t} - M_{t_i\wedge t}, N\right\rangle \\ &= \int_0^t K_s d\langle M, N\rangle_s \end{aligned}$$

On the right hand side we have a classic Riemann–Stieltjes integral against the quadratic cross variation $\langle M, N\rangle$. It turns out that above property already characterizes the stochastic integral and can be used to show its existence.

Theorem 2.28. *Let $M \in H^2$ and $K \in L^2(M)$. Then*

(1) *there exists a unique element $K \bullet M \in H_0^2$ such that*

$$\langle K \bullet M, N\rangle = K \cdot \langle M, N\rangle \ \ \forall N \in H^2.$$

(2) *the map*

$$K \mapsto K \bullet M$$

is an isometry from $\left(L^2(M), |.|_{L^2(M)}\right)$ into $\left(H_0^2, |.|_{H_0^2}\right)$, that is

$$|K|_{L^2(M)} \equiv \mathbb{E}\left[\int_0^\infty H_s^2 d\langle M\rangle_s\right] = \mathbb{E}\left[(K \bullet M)_\infty^2\right] \equiv |K \bullet M|_{H_0^2}$$

We call the process $K \bullet M$ the stochastic integral or Ito integral of the integrand K against the integrator M. We also denote it with $\int K dM$.

[7] Again: recall Convention 2, that strictly speaking we don't consider process but equivalence classes processes.

Proof. **(PROOF NOT EXAMINABLE)** Let $I, I' \in H_0^2$ and orthogonal to all $N \in H^2$. Then their difference is also orthogonal to all $N \in H^2$. Choosing $N = I - I'$gives

$$\langle I - I', I - I' \rangle = 0$$

which shows that I and I' are indistinguishable. To see the existence, first fix $M \in H_0^2$ and consider the map

$$H_0^2 \ni N \mapsto \mathbb{E}\left[(K \cdot \langle M, N \rangle)_\infty\right] \tag{2.5.2}$$

It is a linear functional since Riemann–Stieltjes integration and expectation are both linear operators. Further, by Kunita–Watanabe

$$\left|\int K_s^2 d\langle M, N \rangle\right| \leq |N|_{\mathbb{H}^2} |K|_{L^2(M)} \ \ \forall N \in H_0^2$$

the linear functional (2.5.2) is also continuous. By the Riesz representation theorem any linear continuous functional can be written as an inner product, hence there must exist $K \bullet M \in H_0^2$ such that □

Applied to the case when M is a BM stopped at a fixed time t, $M = \left(B_{t_1 \wedge t}\right)_t$, we have $M \in H^2$ since $|M|_{H^2} = \sqrt{\mathbb{E}\left[\langle B \rangle_t^2\right]} = t$ and we recover what is known as "Ito's isometry" (historically stochastic integration was first developed for BM)

$$\mathbb{E}\left[\int_0^t H_s^2 ds\right] = \mathbb{E}\left[\left(\int_0^t H_s dB_s\right)^2\right].$$

Unfortunately, Theorem 2.28 excludes example Brownian motion on the half-line since $\sqrt{\mathbb{E}\left[\langle B \rangle_\infty^2\right]} = \infty$ and hence $B \notin H^2$. However, if we stop Brownian motion at a fixed time then it becomes an element of H^2. We expect that a similar argument should work if we replace the fixed time t by a stopping time — put differently, the stochastic integral should depend only on local properties of the integrator. Below we show that this is indeed the case. This allows to extend the class of integrands from H^2 to $\mathcal{M}_{\mathrm{c,loc}}$ using stopping times and localization.

2.5.4. Extending the stochastic integral to $\mathcal{M}_{\mathrm{c,loc}}$ integrators. For the extension we first show that the stochastic integral just depends on local properties.

Proposition 2.29. *If τ is a stopping time then*

$$K \bullet M^\tau = (K \bullet M)^\tau$$

Definition 2.30. Let $M \in \mathcal{M}_{\mathrm{c,loc}}$. Denote with $L^2_{\mathrm{loc}}(M)$ the space of progressively measurable processes K such that there exists a sequence of increasing stopping times $(\tau_n), \tau_n \to \infty$ such that

$$\mathbb{E}\left[\int_0^{\tau_n} K_s^2 d\langle M \rangle_s\right] < \infty.$$

Theorem 2.31. *Let $M \in \mathcal{M}_{c,loc}$ and $K \in L^2_{loc}(M)$. Then there exist a unique $K \bullet M \in \mathcal{M}_{c,loc}$ such that*

$$\langle K \bullet M, N \rangle = K \cdot \langle M, N \rangle \ \ \forall N \in \mathcal{M}_{c,loc}.$$

If $M \in H^2$ and $K \in L^2(M)$ then $K \bullet M$ coincides with the stochastic integral defined in Theorem 2.28.

Proof. **(PROOF NOT EXAMINABLE)** □

We already know how to integrate against bounded variation processes, hence we can immediately extend the class of integrands even further.

Definition 2.32. We call a an adapted, continuous process X a *continuous semimartingale* if there exists a $M \in \mathcal{M}_{c,loc}$ and a $A \in BV_c$ such that

$$X_t = X_0 + M_t + A_t \ \forall t \geq 0$$

We denote the class of adapted, continuous semimartingales with $\mathcal{S}_c$. For a locally bounded process[8] K we define

$$K \bullet X := K \bullet M + K \bullet A.$$

Theorem 2.33. *Let $X \in \mathcal{S}_c$ and K be a locally bounded process. Then$K \bullet X \in \mathcal{S}_c$ and the map*

$$K \mapsto K \bullet X$$

has the following properties:

(1) $H \bullet (K \bullet M) = (HK \bullet M)$ *for any locally bounded* H, K,
(2) $(K \bullet M)^\tau = (K \bullet M^\tau)$ *for any stopping time* τ,
(3) *if* $M \in \mathcal{M}_{c,loc}$ *then* $K \bullet M \in \mathcal{M}_{c,loc}$,
(4) *if* $M \in \mathrm{BV}$ *then* $K \bullet M \in \mathrm{BV}$,
(5) *if* $K = K_{-1} 1_{\{0\}}(t) + \sum_{i \geq 1} K_i 1_{)t_i, t_{i+1}]}$, $K_i \in \mathcal{F}_{t_i}$ *then*

$$(K \bullet M)_t = \sum_{i \geq 1} K_i \left(M_{t_{i+1} \wedge t} - M_{t_i \wedge t} \right)$$

Moreover, if K is a left-continuous, locally bounded process then

$$\sum_{t_i \in \pi(n)} K_{t_i} \left(M_{t_{i+1} \wedge t} - M_{t_i \wedge t} \right) \rightarrow_{n \to \infty} (K \bullet M)_t \ \text{ in } \mathbb{P}-\text{probability}$$

whenever $\mathrm{mesh}(\pi(n)) \to 0$ *as* $n \to \infty$.

Proof. **(PROOF NOT EXAMINABLE)** □

Remark 2.34. You might ask:

- *Why semimartingales?* Could we have done things differently and not ended up with semimartingales as our class of integrators? The Bichteler–Dellacherie–Meyer theorem gives a negative answer: if M is a stochastic proceess and I_M an operator on locally bounded processes K such that $I_M(K)$ coincides on the class of simple processes with our definition and I_M is continuous then this is equivalent to M being a semimartingale. The proof is somewhat technical but we refer to [**Protter, 2004**].
- *Why left Riemann sums?* We could look for limits of type

$$\sum_{t_i \in \pi(n)} K_{t_i^\star} \left(M_{t_{i+1} \wedge t} - M_{t_i \wedge t} \right)$$

where $t_i^\star$ is any point in $[t_i, t_{i+1}]$. In the bounded variation case, the choice of $t_i^\star \in [t_i, t_{i+1}]$ does not matter but this is not the case here. For example, if we take $t_i^\star = t_{i+1}$ instead of $t_i^\star = t_i$ we end up with the so-called "Fisk–Stratonovich integral" and a simple calculation shows that

$$\lim_n \sum_{t_i \in \pi(n)} K_{t_{i+1}} \left(M_{t_{i+1} \wedge t} - M_{t_i \wedge t} \right) = (K \bullet M)_t + \frac{1}{2} \langle K, M \rangle_t .$$

One of the main reasons for the popularity of the Ito integral is that in many applications it is naturall that the integrand should not look into the future of the integrator, for example in finance, K is a trading strategy and M an asset price.

[8]That is, there exists constants (c_n) and a sequence of stopping times (τ_n) such that $|K^{\tau_n}| \leq c_n$ (for example, every continuous adapted process is locally bounded ($\tau_n = \inf\{t : |K_t| > n\}$). A locally bounded processes is in $L^2_{\mathrm{loc}}(M)$ if $M \in \mathcal{M}_{\mathrm{c,loc}}$ since$\left| \int_0^{\tau_n} K_s^2 d\langle M \rangle_s \right| \leq c_n^2 \langle M \rangle_{\tau_n}$. Therefore the integral is well defined.

- *Equivalence classes vs pathwise definitions?* Our stochastic integral only depends on the the equivalence class of the underlying integrator and integrand (which explains the notation $\left(\int K dM\right)(\omega)$ as opposed to $\int K(\omega)\, dM(\omega)$). There are more radical departures from classic stochastic calculus that allow to integrate against non-semimartingales like fractional Brownian motion and recent developments like rough path integration can deal with such signals in great generality. However, in many applications the semimartingale assumption is justified, for example in finance via the no-arbitrage condition.

CHAPTER 3

Ito's stochastic calculus and applications

Given a process X we can apply a function f to it to create the new process $f(X) = (f(X_t))_{t\geq 0}$. If f is linear and $X \in \mathcal{M}$ then $f(X) \in \mathcal{M}$ — similarly if f is convex/concave and X is a sub/supermartingale, $f(X)$ is again a sub/supermartingale (vian Jentzen's inequality). Ito's formula shows that there exists a class of processes that is closed under all non-linear transformations f that are sufficiently smooth, namely the class of semimartingales. Ito's formula itself is a consequence of a second order Taylor expansion

$$f(X_t) - f(X_s) = f'(X_s)(X_t - X_s) + \frac{f''(X_s)}{2}(X_t - X_s)^2 + o\left((X_t - X_s)^2\right) \text{ as } |t-s| \to 0$$

Now using a telescope sum

$$f(M_t) - f(M_0) = \sum f'\left(M_{t_i}\right)\left(M_{t_{i+1}} - M_{t_i}\right) + \sum \frac{f''\left(M_{t_i}\right)}{2}\left(M_{t_{i+1}} - M_{t_i}\right)^2 + o\left(\left(M_{t_{i+1}} - M_{t_i}\right)^2\right)$$

and sending $|t_{i+1} - t_i| \to 0$ we have (very heuristically!) "proved" Ito's famous change-of-variable formula.

Theorem 3.1. *Let $f \in C^2(\mathbb{R}, \mathbb{R})$ and $X \in \mathcal{S}_c$. Then $f(X)$ is again a semimartingale and we have $\mathbb{P}$-a.s. that*

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)\, dX_s + \frac{1}{2}\int_0^t f''(X_s)\, d\langle X\rangle_s \ \ \forall t \geq 0.$$

For brevity we also use the differential notation to express the above

$$df(X_t) = f'(X_t)\, dX_t + \frac{1}{2} f''(X_t)\, d\langle X\rangle_t$$

Proof. **(PROOF NOT EXAMINABLE)** One approach is to make our reasoning above rigourous, that is carefully show that the discrete sums convergence and the error term vanishes uniformly. Another approach is to show via a direct calcultation that it holds for all polynomials and conclude by using that polynomials are dense in $C(\mathbb{R}, \mathbb{R})$. □

Remark 3.2.

- Ito's formula not only tells us that $\mathcal{S}_c$ is invariant under C^2 transformation (unlike $\mathcal{M}_c$, thus giving another justification for the introduction of semimartingales). It even gives the explicit semimartingale decomposition of the new semimartingale. If

  $$X = X_0 + M + A \text{ for some } M \in \mathcal{M}_2^c, A \in BV$$

  then

  $$f(X_t) = f(X_0) + \underbrace{\int_0^t f'(X_s)\, dM_t}_{\in \mathcal{M}_2^{c,loc}} + \underbrace{\int_0^t f'(X_t)\, dA_t + \int_0^t f''(X_s)\, d\langle M\rangle_t}_{BV}.$$

- There are some easy extensions: in the case of a $d-$dimensional continuous semimartingales, that is $X=\left(X^1,\dots,X^d\right)$ with $X^i\in\mathcal{S}_c$ and $f\in C^2\left(\mathbb{R}^d,\mathbb{R}\right)$, we have
$$df(X)=\sum_{i=1}^d \partial_i f(X)\,dX^i+\frac{1}{2}\sum_{i,j=1}^d \partial_{ij}^2 f(S_t)\,d\left\langle X^i,X^j\right\rangle_t$$
where $\partial_i f$ denotes the first derivate in the i-th coordinate of f, etc. Note that if S takes only values in an open set we only need f to be defined on that set. Later we will see that one can reduce the smoothness assumption of f further using so-called local times.
- It is often said that Ito's calculus is a "second order" theory: unlike for the case for bounded variation paths x where the change of variable formula reads
$$df(x_t)=f'(x_t)\,dx_t$$
an additional second derivative of f appears. This happens since only a Taylor expansion up to order two will converge. The additional term leads to new phenomena that are not encountered in classic analysis.

3.1. Integration by parts

Another consequence of Ito's formula is that the integration by parts formula has an additional term. By applying the Ito-formula to $(x,y)\mapsto xy$ we immediately get

PROPOSITION 3.3. *Let $X,Y\in\mathcal{S}_c$. Then for every $t\geq 0$*
$$(X\bullet Y)_t=X_tY_t-X_0Y_0-(Y\bullet X)_t-\langle X,Y\rangle_t$$

REMARK 3.4. Applied with processes that start at 0 the above reads
$$\langle X,Y\rangle_t=X_tY_t-(Y\bullet X)_t-\langle X,Y\rangle_t$$
and with $X=Y$ it reduces to
$$\langle X\rangle_t=X_t^2-2(X\bullet X)_t\,.$$
Recall that for a bounded variation path X we have $X_t^2-2(X\bullet X)_t=0$ since the stochastic integral coincides in this case with the classic Riemman–Stieltjes integral. Thus $\langle X\rangle$ can be interpreted as measuring "how far we are away from a first order calculus".

3.2. The stochastic exponential (Doleans–Dade exponential)

In the deterministic case, the unique solution of
$$dy_t=y_t\,dt,y_0=1$$
is given by $y_t=\exp t$. We now replace dt by dB_t and look for a process Y such that
$$dY_t=Y_t\,dB_t,Y_0=1. \tag{3.2.1}$$
Assume such a process Y exists. Applying Ito to $\ln Y_t$ yields
$$\begin{aligned}d\lg Y_t &= \frac{1}{Y_t}dY_t-\frac{1}{2}\frac{1}{Y_t^2}d\langle Y\rangle_t\\ &= \frac{1}{Y_t}dY_t-\frac{1}{2}\frac{1}{Y_t^2}Y_t^2\,dt\\ &= dB_t-\frac{1}{2}dt\end{aligned}$$

hence we arrive at $Y_t = \exp\left(B_t - \frac{1}{2}t\right)$ as candidate for our solution; above calculation was somewhat formal (we assumed Y exists, is not C^2 everywhere, etc.) but by defining

$$Y_t := \exp\left(B_t - \frac{1}{2}t\right)$$

we can quickly verify that it indeed solves (3.2.1): $\left(B_t - \frac{1}{2}t\right)_t$ is a martingale and Ito applied to $x \mapsto \exp x$ shows 3.2.1. This justifies the name stochastic exponential and not surprisingly this turns out to be extremely useful process (it plays a role akin to the exponential funtion in classic calculus).

Theorem 3.5. *Let $X \in \mathcal{S}_c$, $X_0 = 0$. The process $\mathcal{E}(X)$,*

$$\mathcal{E}(X)_t := \exp\left(X_t - \frac{1}{2}\langle X\rangle_t\right),$$

is a continuous semimartingale and the unique solution of the SDE

$$dZ_t = Z_t dX_t \text{ with } Z_0 = 1. \tag{3.2.2}$$

We call $\mathcal{E}(X)$ the stochastic exponential (also Dolean–Dade exponential) of X.

Proof. Applying Ito to the semimartingale $X_t - \frac{1}{2}\langle X\rangle_t$ and the exponential $x \mapsto \exp x$ shows

$$\begin{aligned} d\mathcal{E}(X)_t &= \mathcal{E}(X)_t\, d\left(X_t - \frac{1}{2}\langle X\rangle_t\right) + \frac{1}{2}\mathcal{E}(X)_t\, d\left\langle X_t - \frac{1}{2}\langle X\rangle_t\right\rangle \\ &= \mathcal{E}(X)_t\, dX_t \end{aligned}$$

and clearly $\mathcal{E}(X)_0 = 1$. To see uniqueness let Z be any solution, set $Y_t := \frac{1}{\mathcal{E}(X)_t} = \exp\left(-X_t + \frac{1}{2}\langle X\rangle_t\right)$. By Ito

$$\begin{aligned} dY &= -Y dX + \frac{1}{2}Y d\langle X\rangle + \frac{1}{2}Y d\langle X\rangle \\ &= -Y dX + Y d\langle X\rangle \end{aligned} \tag{3.2.3}$$

Now apply Ito to ZY to get

$$\begin{aligned} d(YZ) &= ZdY + YdZ + d\langle Z, Y\rangle \\ &= -ZY dX + ZY d\langle X\rangle + YZ dX + d\langle Z, Y\rangle \\ &= ZY d\langle X\rangle + \langle Z \bullet X, -Y \bullet X\rangle = 0 \end{aligned}$$

where we used 3.2.3 for the first equality and Ito isometry plus linearity of the bracket for the final equality. Above says that for all $t \geq 0$

$$Z_t Y_t - Z_0 Y_0 = 0$$

hence $Z_t Y_t = Z_0 Y_0 = 1$ and so $Z_t \mathcal{E}(X)_t^{-1} = 1$. □

Remark 3.6. An useful consequence is that for $\lambda \in \mathbb{C}$, $\mathcal{E}(\lambda X)$ is a complex valued semimartingale (i.e. real and imaginary part are both in $\mathcal{S}_c$) that solves the SDE

$$dZ_t = \lambda Z_t dX_t.$$

Example 3.7 (Geometric Brownian motion). An important example is geometric brownian motion $\mathcal{E}(G)$ where $G_t = \mu t + \sigma B_t$, i.e.

$$\mathcal{E}(G) = \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t\right)$$

By above, $Y_0\mathcal{E}(G)$ is the unique solution of the SDE with constant coefficients (μ,σ),

$$\begin{aligned} dY_t &= Y_t dG_t \\ &= \sigma Y_t dt + \mu Y_t dB_t. \end{aligned}$$

PROPOSITION 3.8. *Let $M \in \mathcal{M}_{c,loc}$ with $M_0 = 0$. Then*

(1) $\mathcal{E}(M) \in \mathcal{M}_{c,loc}$, $\mathcal{E}(M)_t \geq 0$ *for all* $t \geq 0$,
(2) $\mathcal{E}(M)$ *is a continuous supermartingale,*
(3) $\mathbb{E}[\mathcal{E}(M)_\tau] \in [0,1]$ *for all stopping times* τ,
(4) $\mathcal{E}(M) \in \mathcal{M}_c$ *and u.i. iff* $\mathbb{E}[\mathcal{E}(M)_\infty] = 1$.

PROOF. (See also exercise Sheet 3). The first statement follows from (3.2.2). We claim that the second, third and fourth statement hold more generally for any $N \in \mathcal{M}_{\text{c,loc}}$ with $\mathbb{E}[N_0] = 1$ and $N_t \geq 0$ for all $t \geq 0$ and not just $\mathcal{E}(M)$. To see this, denote with (τ_n) a sequence of stopping times such that N^{τ_n} is a u.i. martingale (this is possible by Remark 2.5). By Fatou's lemma and optional stopping

$$\mathbb{E}[N_\tau|\mathcal{F}_\sigma] = \mathbb{E}\left[\liminf_n N_{\tau_n\wedge\tau}|\mathcal{F}_\sigma\right] \leq \liminf_n \mathbb{E}\left[N_{\tau_n\wedge\tau}|\mathcal{F}_\sigma\right] = \lim_n N_{\tau_n\wedge\sigma} = N_\sigma \tag{3.2.4}$$

for all stopping times $\sigma \leq \tau$, thus N is a supermartingale. As a consequence, applied with $\sigma = 0$ we have $\mathbb{E}[N_\tau] \leq \mathbb{E}[N_0] = 1$ which shows the third statement. For the fourth statement, note that N being a u.i. martingale, martingale convergence (Theorem 2.15) implies that N_∞ exists and $\mathbb{E}[N_\infty] = \mathbb{E}[N_0] = 1$; conversely if $\mathbb{E}[N_\infty] = 1$ use the supermartingale property of N from statement 2 with $\tau = \infty$ in 3.2.4 to get

$$\mathbb{E}[N_\infty|\mathcal{F}_\sigma] \leq N_\sigma.$$

Taking expectation gives $\mathbb{E}[N_\sigma] \geq \mathbb{E}[N_\infty] = 1$ but because N is a supermartingale (statement 2), $t \mapsto \mathbb{E}[N]$ is decreasing therefore also $1 = \mathbb{E}[N_0] \geq \mathbb{E}[N_\sigma]$, thus $\mathbb{E}[N_\sigma] = 1$ for all stopping times σ. Hence N is a martingale (a supermartingale is martingale iff $t \mapsto \mathbb{E}[N_t]$ is constant). It is u.i. since N_∞ exists in L^1 (again martingale convergence, Theorem 2.15). □

When we discuss changes of measure, it turns out that the case when $\mathcal{E}(M)$ is a martingale is especially important. We give a simple sufficient criteria:

PROPOSITION 3.9. *Let $M \in \mathcal{M}_{c,loc}$ and there exists a constant c such that for all $t \geq 0$*

$$\langle M \rangle_t \leq ct.$$

Then $\mathcal{E}(M) \in \mathcal{M}_c$.

PROOF. By Proposition 3.8 $\mathcal{E}(M) \in \mathcal{M}_{\text{c,loc}}$. Therefore we can find a sequence (τ_n) of stopping times s.t. $\left(\mathcal{E}(M)_{t\wedge\tau_n}\right)_t$ is a martingale for every $n \geq 1$. We would like to send n to ∞ in

$$\mathbb{E}\left[\mathcal{E}(M)_{t\wedge\tau_n}|\mathcal{F}_s\right] = \mathbb{E}\left[\mathcal{E}(M)_{s\wedge\tau_n}\right]. \tag{3.2.5}$$

The (conditional) dominated convergence theorem can be applied if

$$\mathbb{E}\left[\sup_{s\leq t}|\mathcal{E}(M)_s|\right] = \mathbb{E}\left[\sup_{s\leq t}\mathcal{E}(M)_s\right] < \infty \tag{3.2.6}$$

To see that this is satisfied, note that from the definition of $\mathcal{E}(M) \equiv \exp\left(M_t - \frac{1}{2}\langle M\rangle_t\right)$ we deduce that

$$\mathcal{E}(M)_t^2 = \mathcal{E}(2M_t)\exp(\langle M\rangle_t)$$

and use this in Doob's maximal inequality with $p = 2$

$$\begin{aligned} \mathbb{E}\left[\sup_{s\leq t}\mathcal{E}(M)^2_{s\wedge\tau_n}\right] &\leq 4\mathbb{E}\left[\mathcal{E}(M)^2_{t\wedge\tau_n}\right] \\ &= 4\mathbb{E}\left[\mathcal{E}(2M)_{t\wedge\tau_n}\exp\langle M\rangle_{t\wedge\tau_n}\right] \\ &\leq 4\exp ct\mathbb{E}[\mathcal{E}(2M)_0] = 4\exp ct. \end{aligned}$$

Now send $n \to \infty$ and by monotone convergence $\mathbb{E}\left[\sup_{s\leq t}\mathcal{E}(M)_s^2\right] \leq 4\exp ct$, so we have shown

$$\mathbb{E}\left[\sup_{s\leq t}\mathcal{E}(M)_s\right]^2 \leq \mathbb{E}\left[\sup_{s\leq t}\mathcal{E}(M)_s^2\right] < \infty.$$

Thus (3.2.6) holds and we can apply dominated convergence in (3.2.5). □

Two much more general criteria are due to Kazamaki and Novikov.

Proposition 3.10. *Let $M \in \mathcal{M}_{c,loc}$. If at least one of the conditions*

(1) *(Kazamaki condition) $\exp\left(\frac{1}{2}M\right)$ is a u.i. submartingale*

(2) *(Novikov condition) $\mathbb{E}\left[\exp\frac{1}{2}\langle M\rangle_\infty\right] < \infty$*

holds, then $\mathcal{E}(M)$ is a continuous and u.i. martingale.

Proof. 1) Exercise Sheet 4, exercise 2.

2) One can show that 2) implies the Kazamaki condition. This will be easy once we have proved the Burkholder–Davie–Gundy inequalities and we postpone the proof until then. Exercise Sheet 5, exercise 2) □

Remark 3.11. Despite their innocent looks, both conditions can often be hard to verify.

3.3. Brownian functional and martingale representation

In his section we assume that $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ carries a Brownian motion B. We denote the filtration generated by Brownian motion with $(\mathcal{F}_t^B)$.

Theorem 3.12. *Let $Y \in L^2(\Omega, \mathcal{F}_\infty^B, \mathbb{P})$. Then there exists a predictable process $H \in L^2(B)$ such that*

$$Y = \mathbb{E}[Y] + \int_0^\infty H_s dB_s. \tag{3.3.1}$$

This representation of Y is unique (up to indistinguishability).

To motivate the proof note that by the previous section, we know that if we take $Y = \mathcal{E}(B)_t$ then

$$Y = \mathcal{E}(B)_t = \mathbb{E}Y + \int_0^\infty \mathcal{E}(B)_s 1_{[0,t]}(s)\, dB_s$$

is of above form. More generally, if $Y = \mathcal{E}(M)_t$ for $M_t = \int_0^t f_s dB_s$, then

$$Y = \mathbb{E}Y + \int_0^\infty \underbrace{\mathcal{E}(M)_s f_s 1_{[0,t]}(s)}_{H_s} dB_s.$$

So if we can show that the linear span of the random variables $\mathcal{E}(M)_\infty$ with M a stochastic integral $M = f \bullet B$ with "nice enough" f is a dense subspace of $L^2(\Omega, \mathcal{F}_\infty^B, \mathbb{P})$, the existence of the represenation will follow.

Lemma 3.13. *Let*

$$\mathcal{J} = \left\{\mathcal{E}(M)_\infty : M = \int_0^\infty f dB_s \text{ and } f = \sum_{j=1}^k \lambda_j 1_{)t_j,t_{j+1}]} \text{for } k \in \mathbb{N}, \lambda_j \in \mathbb{R}, 0 \leq t_1 < \cdots < t_k < \infty\right\}.$$

The linear span of $\mathcal{J}$ is dense in $L^2(\Omega, \mathcal{F}_\infty^B, \mathbb{P})$.

For the proof of this lemma recall that

- if two complex analytic functions $f, g : \mathbb{C}^k \to \mathbb{R}$ agree on a connected subset of $\mathbb{C}^k$ then $f = g$,

- the Fourier transform of a measure μ on $\left(\mathbb{R}^k, \mathcal{B}\left(\mathbb{R}^k\right)\right)$ is
$$u \mapsto \int_{\mathbb{R}^k} e^{-i\langle u,x\rangle} \mu(dx)$$
and if this map equals $u \mapsto 0$ then μ is the zero measure.

Proof. **(Proof of Lemma** 3.13**. Not examinable)**It is sufficient to show that if

$$\langle Y, Z\rangle_{L^2(\Omega, \mathcal{F}^B_\infty, \mathbb{P})} = 0 \text{ for all } Z \in \mathcal{J}. \tag{3.3.2}$$

then $Y = 0$ a.s. Note that by definition of $\mathcal{J}$ every $Z \in \mathcal{J}$ is of the form $\exp\left(\int_0^\infty f dB - \frac{1}{2}\left\langle \int_0^\cdot f dB \right\rangle_\infty\right)$ but since f is step function with compact support,

$$\begin{aligned}
\int_0^\infty f dB &= \sum_{j=1}^k \lambda_j \left(B_{t_{j+1}} - B_{t_j}\right) \text{ and} \\
\left\langle \int_0^\cdot f dB \right\rangle_\infty &= \int_0^\infty f^2 d\langle B\rangle_s = \sum_{i=1}^k \lambda_i^2 \left(t_{j+1} - t_j\right)
\end{aligned}$$

Therefore

$$Z = \exp\left(\sum_{j=1}^k \lambda_j \left(B_{t_{j+1}} - B_{t_j}\right) - \frac{1}{2}\sum \lambda_j^2 \left(t_{j+1} - t_j\right)\right)$$

and (3.3.2) holds if for every $k \in \mathbb{N}$, $t_1 < \cdots t_k$ the map

$$\lambda = (\lambda_1, \ldots, \lambda_k) \mapsto \mathbb{E}\left[\exp\left(\sum_{i=1}^k \lambda_i \left(B_{t_{i+1}} - B_{t_i}\right)\right) Y\right] = 0.$$

This map is clearly real analytic in $\lambda \in \mathbb{R}^k$. Obviously its (unique!) extension to an analytic function to $\lambda \in \mathbb{C}^k$ is $\lambda \mapsto 0$. Evaluated at

$$(i\lambda_1, \ldots, i\lambda_k) \text{ with } \lambda_1, \ldots, \lambda_k \in \mathbb{R}$$

this gives

$$\mathbb{E}\left[\exp\left(\sum_{i=1}^k i\lambda_i \left(B_{t_{i+1}} - B_{t_i}\right)\right) Y\right] = 0.$$

Now define a measure $\mathbb{Q}$ on $\left(\Omega, \mathcal{F}^B_\infty\right)$ asa $d\mathbb{Q} := Y \cdot d\mathbb{P}$ (that is $\mathbb{Q}(A) = \mathbb{E}_{\mathbb{P}}[Y 1_A]$ for $A \in \mathcal{F}^B_\infty$). Above expression is the Fourier transform of the image of $\mathbb{Q}$ under the map

$$I(\omega) = \left(B_{t_1}(\omega) - B_0(\omega), \ldots, B_{t_k}(\omega) - B_{t_{k-1}}(\omega)\right).$$

That is

$$\lambda = (\lambda_1, \ldots, \lambda_k) \mapsto \mathbb{E}\left[\exp\left(\sum_{i=1}^k i\lambda_i \left(B_{t_{i+1}} - B_{t_i}\right)\right) Y\right] = \int_{\mathbb{R}^k} e^{-i\langle x,\lambda\rangle} d\mathbb{Q} \circ I^{-1}(x)$$

it the Fourier transform of the measure $\mathbb{Q} \circ I^{-1}$ on $\left(\mathbb{R}^k, \mathcal{B}\left(\mathbb{R}^k\right)\right)$ and equals 0. Hence $\mathbb{Q} \circ I^{-1}(S) = 0$ for all $S \in \mathcal{B}\left(\mathbb{R}^k\right)$ and therefore

$$\mathbb{Q}(A) = 0$$

for all $A \in \sigma(I) = \sigma\left(B_{t_1} - B_0, \ldots, B_{t_k} - B_{t_{k-1}}\right) = \sigma\left(B_{t_1}, \ldots, B_{t_k}\right)$. Since $d\mathbb{Q} = Y d\mathbb{P}$ and $\mathbb{P}$ is a probability measure, $Y = 0$ $\mathbb{P}$–a.s. □

We now have everything to give the proof of Theorem 3.12.

Proof. (of Theorem 3.12) Set

$$\mathcal{H}=\left\{Y\in L^{2}\left(\Omega,\mathcal{F}_{\infty}^{B},\mathbb{P}\right):\mathbb{E}[Y]+\int_{0}^{\infty}H_{s}dB_{s},H\in L^{2}(B)\text{and predictable}\right\}$$

and note that this is a linear space. Moreover, $\mathcal{J}\subset\mathcal{H}$ since directly by Ito's formula

$$\mathcal{E}(M)_{t}=1+\int_{0}^{t}\mathcal{E}(M)_{s}f_{s}dB_{s}\text{ for }M_{t}=\int_{0}^{t}f_{s}dB_{s}.$$

Since f has compact support, this holds also with $t=\infty$. Therefore $\mathcal{E}(M)_{\infty}\in\mathcal{H}$. The linear span of $\mathcal{J}$ is dense in $\mathcal{H}$ by above Lemma so the existence of the representation (3.3.1) follows if $\mathcal{H}$ is closed in $L^{2}\left(\Omega,\mathcal{F}_{\infty}^{B},\mathbb{P}\right)$.

We now show the closedness of $\mathcal{H}$: therefore note that if Y is of the form (3.3.1), then

$$\begin{aligned}Y^{2}&=\left(\mathbb{E}[Y]+\int_{0}^{\infty}H_{s}dB_{s}\right)^{2}\\&=\mathbb{E}[Y]^{2}+2\mathbb{E}[Y]\int_{0}^{\infty}H_{s}dB_{s}+\left(\int_{0}^{\infty}H_{s}dB_{s}\right)^{2}\end{aligned}$$

Taking expecations and the Ito isometry yields

$$\mathbb{E}\left[Y^{2}\right]=\mathbb{E}[Y]^{2}+\int_{0}^{\infty}\mathbb{E}\left[H_{s}^{2}\right]ds. \tag{3.3.3}$$

Now let $(Y^{n})\subset\mathcal{H}$ be Cauchy in $L^{2}\left(\Omega,\mathcal{F}_{\infty}^{B},\mathbb{P}\right)$ and apply (3.3.3) with $Y=Y_{m}-Y_{n}$. This shows

$$\mathbb{E}\left[(Y_{m}-Y_{n})^{2}\right]=\mathbb{E}[Y_{m}-Y_{n}]^{2}+\underbrace{\int_{0}^{\infty}\left(H_{s}^{m}-H_{s}^{n}\right)^{2}ds}_{=|H^{m}-H^{n}|_{L^{2}(B)}}\to 0\text{ as }m,n\to\infty.$$

Thus $(H^{n})_{n}$ is Cauchy in $L^{2}(B)$ and therefore converges to a $H\in L^{2}(B)$. To sum up,

$$\lim_{n}Y_{n}=\lim_{n}\mathbb{E}[Y_{n}]+\int_{0}^{\infty}H_{s}dB_{s}$$

(where the limit on the left is in $L^{2}\left(\Omega,\mathcal{F}_{\infty}^{B},\mathbb{P}\right)$). This shows the closedness $\mathcal{H}$ in $L^{2}(\Omega,\mathcal{F}_{\infty},\mathbb{P})$ and since $\mathcal{H}$ contains a dense subspace (by above Lemma) the existence of the representation (3.3.1) follows.

To see uniqueness, assume $Y=\mathbb{E}[Y]+\int_{0}^{\infty}H_{s}dB=\mathbb{E}[Y]+\int_{0}^{\infty}H_{s}'dB_{s}$. Then

$$0=\int_{0}^{\infty}\left(H_{s}-H_{s}'\right)dB_{s}$$

and by (3.3.3) applied with $Y=\int_{0}^{\infty}\left(H_{s}-H_{s}'\right)dB_{s}$ we have

$$0=\int_{0}^{\infty}\left(H_{s}-H_{s}'\right)^{2}ds.$$

Therefore $|H-H'|_{L^{2}(B)}=0$ or put differently $H_{s}(\omega)=H_{s}'(\omega)$ for a.e. (s,ω) which shows uniqueness of the representation (3.3.1). □

Theorem 3.14. *Let M be a continuous $\left(\mathcal{F}_{t}^{B}\right)$-martingale that is L^{2}-bounded (i.e. $M\in H^{2}$). Then there exists a predictable $H\in L^{2}(B)$ such that*

$$M_{t}=\mathbb{E}[M_{\infty}]+\int_{0}^{t}H_{s}dB_{s}.$$

Proof. By L^2-boundedness we have $M_t = \mathbb{E}\left[M_\infty|\mathcal{F}_t^B\right]$ for some $M_\infty \in L^2\left(\mathcal{F}_\infty^B, \mathbb{P}\right)$ (see Theorem 2.24). We apply the previous result

$$M_\infty = \mathbb{E}\left[M_\infty\right] + \int_0^\infty H_s dB_s$$

and take the conditional expectation $\mathbb{E}\left[\cdot|\mathcal{F}_t^B\right]$. □

Remark 3.15.

- Above Theorem extends to local martingales using localization but we do not develop this.
- One could ask if we can replace Brownian motion by another process. This is not true in general.
- In the exercise sheets you will see some concrete exampels where the martingale representation gets explicit. More generally and in analogy to classic calculus one expects H to be in a certain sense the first derivative of F. This is indeed the case: the process H can be expressed as a Frechet derivative of M_∞. This is developed in the so-called Malliavin calculus.

3.4. Lévy's characterization of Brownian motion

If B is a Brownian motion then

$$B \in \mathcal{M}_{\text{c,loc}} \text{ and } \langle B\rangle_t = t.$$

A famous result of Paul Levy shows this already characterizes Brownian motion in the class of continuous and adapted processes. The proof follows is a rather simple application of the stochastic exponential.

Theorem 3.16. *Let M be a continuous, adapted process with $M_0 = 0$. Then M is a Brownian motion if and only if*

$$M \in \mathcal{M}_{c,loc} \text{ and } \langle M\rangle_t = t.$$

Proof. If M is a BM then we already know $M \in \mathcal{M}_{\text{c}} \subset \mathcal{M}_{\text{c,loc}}$ with $\langle M\rangle_t = t$ a.s. To show the other direction, consider

$$\mathcal{E}(iuM)_t = \exp\left(iuM_t + \frac{u^2}{2}\langle M\rangle_t\right) \text{ with } u \in \mathbb{R}.$$

By Ito

$$\begin{aligned} d\mathcal{E}(iuM)_t &= \mathcal{E}(iuM)_t\, d\left(iuM_t + \frac{u^2}{2}\langle M\rangle_t\right) + \frac{1}{2}\mathcal{E}(iuM)_t\, d\left\langle iuM_t + \frac{u^2}{2}\langle M\rangle_t\right\rangle \\ &= \mathcal{E}(iuM)_t\, d\left(iuM_t + \frac{u^2}{2}\langle M\rangle_t\right) - \frac{u^2}{2}\mathcal{E}(iuM)_t\, d\langle M_t\rangle \\ &= iu\mathcal{E}(iuM)_t\, dM_t. \end{aligned}$$

Hence $\mathcal{E}(iuM)$ is a stochastic integral against a continuous local martingale, thus again a continuous local martingale. Moreover $|\mathcal{E}(iuM)_t| \leq 1$ and therefore a martingale by Proposition 3.8. Therefore

$$\mathbb{E}\left[\mathcal{E}(iuM)_t\,|\mathcal{F}_s\right] = \mathcal{E}(iuM)_s.$$

and rewriting the above

$$\mathbb{E}\left[\exp\left(iu\left(M_t - M_s\right)\right)|\mathcal{F}_s\right] = \exp\frac{u^2(t-s)}{2}$$

Taking expecation, this already implies that $M_t - M_s$ is normally distributed with mean zero and variance $(t-s)$. To see that $(M_t - M_s)$ is independent of $\mathcal{F}_s$, take $A \in \mathcal{F}_s$, $\mathbb{P}(A) > 0$ and define a new measure on $(\Omega, \mathcal{F})$ as

$$\mathbb{P}_A(B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \text{ for } B \in \mathcal{F}.$$

Then $\mathbb{E}_{\mathbb{P}(A)}[\exp(iu(M_t - M_s))] = \frac{\mathbb{E}_{\mathbb{P}}[1_A \exp(iu(M_t-M_s))]}{\mathbb{P}(A)} = \exp\left(-\frac{u^2}{2}(t-s)\right)$, hence $M_t - M_s \sim \mathcal{N}(0, t-s)$ wrt to $\mathbb{P}_A$. Thus for all bounded measurable g, $g \geq 0$,

$$\mathbb{E}_{\mathbb{P}_A}[g(M_t - M_s)] = \mathbb{E}[g(M_t - M_s)]$$

or unwrapping the definition of $\mathbb{P}(A)$

$$\mathbb{E}[g(M_t - M_s)] = \mathbb{P}(A)\mathbb{E}[g(M_t - M_s)] \ \forall A \in \mathcal{F}_s$$

The the indepence of increments follows.

□

3.5. Martingales as time changed BM: Dambis, Dubins–Schwarz

Let $M = (M_t)_{t\geq 0}$ be a continous and bounded martingale. If we can find a time change $t \mapsto \tau(t)$ such that the time changed process $M_\tau := \left(M_{\tau(t)}\right)_{t\geq 0}$ is continuous and has as bracket

$$\langle M_\tau \rangle_t = t$$

then by OST (since M is bounded, hence u.i.) $\mathbb{E}\left[M_{\tau(t)}|\mathcal{F}_{\tau(u)}\right] = M_{\tau(u)}$, we know that M is a $\left(\mathcal{F}_{\tau(u)}\right)$-martingale. By Levy's characterization it must be a Brownian motion wrt to $\left(\mathcal{F}_{\tau(t)}\right)_{t\geq 0}$. To find such a a time change τ is trivial when

$$t \mapsto \langle M \rangle_t \tag{3.5.1}$$

is strictly increasing: simply take the inverse of 3.5.1.

A famous theorem by Dambis and Dubins–Schwarz generalizes the above argument. This proves the somewhat suprising fact that one-dimensional continous local martingales that "move enough" are simply Brownian motion run at a different speed and this time-change is given by the bracket $\langle M \rangle$.

Theorem 3.17. *Let $M \in \mathcal{M}_{c,loc}$ with $M_0 = 0$ and*

$$\lim_t \langle M \rangle_t = \infty \ a.s. \tag{3.5.2}$$

Define for every $t \geq 0$

$$\tau_t := \inf\{u : \langle M \rangle_u > t\} \text{ and } \mathcal{G}_t := \mathcal{F}_{\tau_t}.$$

Then for each $s \geq 0$, τ_s is a $(\mathcal{F}_t)$-stopping time, $(\mathcal{G}_t)$ is a filtration on $(\Omega, \mathcal{F}, \mathbb{P})$ that satisifies the usual conditions and

1. *the process $B = (B_t)$ defined as $B_t := M_{\tau_t}$ is a Brownian motion on $(\Omega, \mathcal{F}, (\mathcal{G}_t), \mathbb{P})$,*
2. *$(M_t)_{t\geq 0} = \left(B_{\langle M \rangle_t}\right)_{\geq 0}$ a.s.*

We prepare the proof with a lemma.

Lemma 3.18. *Let $M \in \mathcal{M}_{c,loc}$. The intervals of constancy of $\langle M \rangle$ and M coincide.*

Proof. Exercise □

We can now proof the main result

Proof. If we can show that (B_t) is a continuous process then it remains by Levy's characaterization, Theorem 3.16, only to show that $B \in \mathcal{M}_{\text{loc}}$ and that $\langle B \rangle_t = t$. To see continuity note that for each $s \geq 0$, τ_s is the hitting time of the bracket process $\langle M \rangle$ of the set (s, ∞), hence a $(\mathcal{F}_t)$-stopping time. In fact,

$$t \mapsto \tau_t$$

is the right-continuous inverse of

$$s \mapsto \langle M \rangle_s .$$

The assumption (3.5.2) also guarantees that $\tau_s < \infty$ a.s. and clearly $s \mapsto \tau_s$ is an increasing process. The right-continuity of $t \mapsto \tau_t$ together with the continuity of $t \mapsto M_t$ implies that

$$t \mapsto B_t = M_{\tau_t}$$

is right-continuous (a.s.). It also has left-limits since

$$B_{s-} = \lim_{u \nearrow s} B_u = \lim_{u \nearrow s} M_{\tau_u} = M_{\tau_{s-}}. \tag{3.5.3}$$

It now follows that B is even continuous: by definition of τ, $\tau_{s-} < \tau_s$ iff $\langle M \rangle$ is constant on $[\tau_{s-}, \tau_s]$. By Lemma 3.18, $\tau_{s-} < \tau_s$ iff M is constant on $[\tau_{s-}, \tau_s]$, that is (3.5.3) reduces to

$$B_{s-} = M_{\tau_{s-}} = M_{\tau_s}.$$

To conclude via Levy's characterization, it remains to show $B \in \mathcal{M}_{\text{loc}}$ and that $\langle B \rangle_t = t$. By the characterization of the bracket this is equivalent to show

$$(B_t) \text{ and } \left(B_t^2 - t\right) \text{ are local } (\mathcal{G}_t)-\text{martingales.}$$

For $u \leq s$ consider

$$\mathbb{E}[B_s | \mathcal{G}_u] = \mathbb{E}\left[M_{\tau_s} | \mathcal{F}_{\tau_u}\right].$$

Note that $\mathbb{E}\left[M_{\tau_t}^2\right] = \mathbb{E}\left[\langle M \rangle_{\tau_t}\right] = t$ so the stopped process $M^{\tau_n} = \left(M_{\tau_n \wedge t}\right)_{t \geq 0}$ is L^2-bounded for $n \in \mathbb{N}$. Therefore we can apply OST. Moreover, the monotonicity of $s \mapsto \tau_s$ guarentess $\tau_u \leq \tau_s \leq \tau_n$ for $u \leq s \leq n$ so putting everything together gives

$$\mathbb{E}[B_s | \mathcal{G}_u] = \mathbb{E}\left[M_{\tau_s} | \mathcal{F}_{\tau_u}\right] = \mathbb{E}\left[M_{\tau_s}^{\tau_n} | \mathcal{F}_{\tau_u}\right] = M_{\tau_u}^{\tau_n} = M_{\tau_u} = B_u.$$

Similarly,

$$\mathbb{E}\left[B_s^2 - s | \mathcal{G}_u\right] = \mathbb{E}\left[M_{\tau_s \wedge \tau_n}^2 - \langle M \rangle_{\tau_s \wedge \tau_n} | \mathcal{F}_{\tau_u}\right] = M_{\tau_u \wedge \tau_n}^2 - \langle M \rangle_{\tau_u \wedge \tau_n} = M_{\tau_u}^2 - \langle M \rangle_{\tau_u} = B_u^2 - u.$$

This shows that B is a Brownian motion.

Finally, the second statement follows since by definition $B_{\langle M \rangle_t} = M_{\tau_{\langle M \rangle_t}}$ and again by Lemma 3.18, $s \mapsto \tau_s$ is constant on $\left[t, \tau_{\langle M \rangle_t}\right]$, hence $M_{\tau_{\langle M \rangle_t}} = M_t$. □

3.6. The maximum, the bracket and their moments: Burkholder–Davis–Gundy inequalities

The running maximum $M^\star$ of a (local) martingale appears naturally and is usually a complicated quantity. Doob's maximal inequlities gave us one powerful tool to bound its moments. The Burkholder–Davis–Gundy inequalities gives us another. It relates the L^p-norms for all $p > 0$ of the running maximum with that of the bracket process and the latter are often easier to handle. We already ran into a situation where we needed such an estimate (Kazamaki's criterion) and we will see more such applications when we work with SDEs and local times.

THEOREM 3.19 (BURKHOLDER–DAVIS–GUNDY). *For every $p \in (0, \infty)$ there are two constants $c_1 = c_1(p), c_2 = c_2(p) > 0$ such that*

$$c_1 \mathbb{E}\left[\langle M \rangle_\tau^{p/2}\right] \leq \mathbb{E}\left[\left(M_\tau^\star\right)^p\right] \leq c_2 \mathbb{E}\left[\langle M \rangle_\tau^{p/2}\right]$$

holds for any $M \in \mathcal{M}_{c,loc}, M_0 = 0$ and stopping time τ.

REMARK 3.20.

- Above is already non-trivial and useful with the constant stopping time $\tau = \infty$.
- An important class processes with complicated running maxima are stochastic integrals $M_t = \int_0^t K_s dN_s$. Since $\langle M \rangle_t = \int_0^t K_s^2 d\langle N \rangle_s$ and the latter integral is often easy to estimate, e.g. if N is a Brownian motion, the BDG inequalities turns out the be very useful in this context.

- Chapter 2 recalled L^2-theory of martingale integration (that is the spaces H^2). We then made a huge step to the general local martingale integration using localization. However, in between these two extremes one can develop a L^p-theory with sharper estimates: the space of martingales with $|M^\star_\infty|_{L^p(\Omega,\mathcal{F}_\infty,\mathbb{P})} < \infty$ forms a vector space and Theorem 3.19 tells us that the norms $M \mapsto |M^\star_\infty|_{L^p(\Omega,\mathcal{F}_\infty,\mathbb{P})}$ and $M \mapsto |\langle M\rangle_\infty|_{L^{p/2}(\Omega,\mathcal{F}_\infty,\mathbb{P})}$ are equivalent. We do not develop this general H^p theory in these notes.
- Compare the above with Proposition 2.25 which shows that for $p = 2$ a sharper statement holds, namely one can take $c_1(2) = c_2(2) = 1$. (This is not true for general p). Moreover, the BDG inequalities cover also the case of nearly no integrability, $p \in (0, 1)$.

By Levy's theorem 3.16, we can write M as a time-changed BM so it is sufficient to prove that

THEOREM 3.21. *If B is a standard Brownian motion, then for every $p \in (0,\infty)$ there are two constants $c_1 = c_1(p), c_2 = c_2(p) > 0$ such that*

$$c_{1,p}\mathbb{E}\left[\tau^{p/2}\right] \le \mathbb{E}\left[\left(B^\star_\tau\right)^p\right] \le c_{1,p}\mathbb{E}\left[\tau^{p/2}\right]$$

for every stopping time τ.

There are several approaches to show BDG. The classic approach is via so-called "good λ inequalities".

LEMMA 3.22. *We call a function $F : [0,\infty) \to \mathbb{R}$ moderately increasing with growth rate k if F is nondecreasing, $F(0) = 0$ and*

$$F(2\lambda) \le kF(\lambda).$$

Let X, Y be nonnegative random variables. If

$$\mathbb{P}(X > 2\lambda, Y \le \delta\lambda) \le \delta^2\mathbb{P}(X > \lambda) \ \ \forall\delta,\lambda > 0$$

then

$$\mathbb{E}[F(X)] \le c\mathbb{E}[F(Y)]$$

for every moderately increasing function F, where the constant c depends only on the growth rate K.

PROOF. By replacing F with $F \wedge n$ for $n \in \mathbb{N}$ and letting $n \to \infty$ we see that wlog we can assum that F is bounded. By our assumption we bound

$$\begin{aligned}\mathbb{P}(X > 2\lambda) &= \mathbb{P}(X > 2\lambda, Y \le \delta\lambda) + \mathbb{P}(X > 2\lambda, Y > \delta\lambda)\\ &\le \delta^2\mathbb{P}(X > \lambda) + \mathbb{P}(Y > \delta\lambda)\end{aligned}$$

Integrating against $dF(\lambda)$ shows

$$\begin{aligned}\mathbb{E}F\left(\frac{X}{2}\right) &= \int_0^\infty \mathbb{P}\left(\frac{X}{2} > \lambda\right)dF(\lambda)\\ &\le \delta^2\mathbb{E}[F(X)] + \mathbb{E}\left[F\left(\frac{Y}{\delta}\right)\right]\end{aligned}$$

Choosing δ s.t. $K\delta^2 < 1$ and N s.t. $2^N > \frac{1}{\delta}$ and using the growth assumption gives

$$\mathbb{E}\left[F\left(\frac{Y}{\delta}\right)\right] \le K^N\mathbb{E}[F(Y)]$$

and therefore

$$\mathbb{E}[F(X)] \le K\mathbb{E}\left[F\left(\frac{X}{2}\right)\right] \le K\delta^2\mathbb{E}[F(X)] + K^{N+1}\mathbb{E}[F(Y)].$$

□

BDG follows from the above with $B^\star_\tau$ and $\sqrt{\tau}$ as nonnegative random variable. So it only remains to show that such good λ inequalities hold.

LEMMA 3.23. *Let $\beta > 1$ and $\delta > 0$. Then $\forall \lambda > 0$*

$$\begin{aligned}
\mathbb{P}\left(B_\tau^\star > \beta\lambda, \sqrt{\tau} \le \delta\lambda\right) &\le \frac{\delta^2}{(\beta-1)^2}\mathbb{P}\left(B_\tau^\star > \lambda\right), \\
\mathbb{P}\left(\sqrt{\tau} > \beta\lambda, B_\tau^\star \le \delta\lambda\right) &\le \frac{\delta^2}{(\beta^2-1)}\mathbb{P}\left(\sqrt{\tau} > \lambda\right).
\end{aligned}$$

PROOF. If the results holds with $\tau \wedge n$ for all $n \in \mathbb{N}$ instead of τ, then it must hold also for τ. Thus we can assume wlog that τ is bounded. To see first inequality, define

$$\begin{aligned}
\sigma_1 &= \inf\{t : |B_{t\wedge\tau}| > \lambda\} \\
\sigma_2 &= \inf\{t : |B_{t\wedge\tau}| > \beta\lambda\} \\
\sigma_3 &= \inf\left\{t : \sqrt{|t\wedge\tau|} > \delta\lambda\right\}
\end{aligned}$$

We observe that

$$\left\{B_\tau^\star > \beta\lambda\right\} \subset \{\sigma_1 < \sigma_2 \le \tau\} \text{ and } \left\{\sqrt{\tau} \le \delta\lambda\right\} \subset \{\sigma_3 = \infty\}. \tag{3.6.1}$$

and therefore

$$\mathbb{P}\left(B_\tau^\star > \beta\lambda, \sqrt{\tau} \le \delta\lambda\right) \le \mathbb{P}\left(\left|B_{\tau\wedge\sigma_3\wedge\sigma_2} - B_{\tau\wedge\sigma_3\wedge\sigma_1}\right| \ge (\beta-1)\lambda\right).$$

Applying Chebycef gives

$$\mathbb{P}\left(B_\tau^\star > \beta\lambda, \sqrt{\tau} \le \delta\lambda\right) \le \frac{\mathbb{E}\left[\left|B_{\tau\wedge\sigma_3\wedge\sigma_2} - B_{\tau\wedge\sigma_3\wedge\sigma_1}\right|^2\right]}{(\beta-1)^2\lambda^2} \tag{3.6.2}$$

For bounded stopping times $\rho_1 \le \rho_2$ we have

$$\begin{aligned}
\mathbb{E}\left[\left(B_{\rho_2} - B_{\rho_1}\right)^2\right] &= \mathbb{E}\left[B_{\rho_2}^2\right] - 2\mathbb{E}\left[B_{\rho_2}B_{\rho_1}\right] + \mathbb{E}\left[B_{\rho_1}^2\right] \\
&= \mathbb{E}\left[B_{\rho_2}^2\right] - 2\mathbb{E}\left[B_{\rho_1}\mathbb{E}\left[B_{\rho_2}|\mathcal{F}_{\rho_1}\right]\right] + \mathbb{E}\left[B_{\rho_1}^2\right] \\
&= \mathbb{E}\left[B_{\rho_2}^2\right] - \mathbb{E}\left[B_{\rho_1}^2\right] = \mathbb{E}[\rho_2 - \rho_1]
\end{aligned}$$

where we used that use that $\left(B_t^2 - t\right)$ is a martingale. Applied with $\rho_2 = \tau\wedge\sigma_3\wedge\sigma_2$ and $\rho_1 = \tau\wedge\sigma_3\wedge\sigma_1$ (recall that wlog τ is bounded) we get

$$\mathbb{E}\left[\left|B_{\tau\wedge\sigma_3\wedge\sigma_2} - B_{\tau\wedge\sigma_3\wedge\sigma_1}\right|^2\right] = \mathbb{E}[\tau\wedge\sigma_3\wedge\sigma_2 - \tau\wedge\sigma_3\wedge\sigma_1].$$

Now we have the trivial estimate

$$\mathbb{E}[\tau\wedge\sigma_3\wedge\sigma_2 - \tau\wedge\sigma_3\wedge\sigma_1] \le (\tau\wedge\sigma_3)\mathbb{P}(\sigma_1 < \infty)$$

and using that by (3.6.1) $\tau\wedge\sigma_3 \le \tau \le \delta^2\lambda^2$ we plug this into (3.6.2) to get

$$\begin{aligned}
\mathbb{P}\left(B_\tau^\star > \beta\lambda, \sqrt{\tau} \le \delta\lambda\right) &\le \frac{\delta^2\lambda^2\mathbb{P}(\sigma_1 < \infty)}{(\beta-1)^2\lambda^2} \\
&\le \frac{\delta^2\mathbb{P}\left(B_\tau^\star > \beta\lambda\right)}{(\beta-1)^2}.
\end{aligned}$$

□

To see the other inequality, we run the same argument with the roles of $B_\tau^\star$ and τ reversed.

3.7. Changes of measure on pathspace: Girsanov and Cameron–Martin

Let M, N be two real-valued, jointly Gaussian random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. A direct calculation (e.g. using the characteristic function) shows

$$\mathbb{E}[g(M+c)] = \mathbb{E}[\rho g(M)]$$

for all bounded measurable functions g and where

$$\rho := \exp\left(N - \frac{1}{2}\operatorname{Var}[N] - \mathbb{E}[N]\right) \text{ and } c := \operatorname{Cov}(M, N).$$

Since $\mathbb{E}[\rho] = 1$ we can define a new probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ by $d\mathbb{Q} = \rho d\mathbb{P}$, that is

$$\mathbb{Q}(A) := \mathbb{E}_{\mathbb{P}}[\rho 1_A] \text{ for } A \in \mathcal{F}$$

Then $\mathbb{P} \sim \mathbb{Q}$, ρ is the density $\rho = \frac{d\mathbb{P}}{d\mathbb{Q}}$ and above can be rewritten as

$$\mathbb{E}_{\mathbb{P}}[g(M+c)] = \mathbb{E}_{\mathbb{Q}}[g(M)].$$

To sum up, *an additive term can be equivalent to an absolutely continuous change of measure* on $(\Omega, \mathcal{F})$. The corresponding density ρ takes an explicit form. The aim of this section is to develop a similar result in the infinite-dimensional case, that is when real-valued random variables are replaced by path-valued random variables, namely local martingales. We first recall the Radon–Nikodym theorem:

THEOREM 3.24. *Let $(\Omega, \mathcal{A}, \mu)$ be a separable measure space and μ and ν be finite measures on $\mathcal{A}$. TFAE*

(1) *For every $A \in \mathcal{A}$, $\mu(A) = 0$ implies $\nu(A) = 0$ (we say ν is absolutely continuous wrt to μ and write $\mu \gg \nu$)*
(2) *There exists a $\rho \in L^1(\Omega, \mathcal{A}, \mu)$ such that $k \geq 0$ and $\nu(A) = \int_A \rho d\mu$ (we call ρ the density of ν wrt μ and write $\rho = \frac{d\nu}{d\mu}$).*

If $\mu \gg \nu$ and $\nu \gg \mu$ we say μ and ν are equivalent and denote this as $\mu \sim \nu$. We are interested in the case when the measures are probablity measures on a pathspace: that is we have additionally a filtration, thus given a filtered probabilty space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ and a positive random variable $\rho_\infty > 0$ a.s. with $\mathbb{E}_{\mathbb{P}}[\rho_\infty] = 1$ we define a new probability measure

$$d\mathbb{Q} := \rho_\infty d\mathbb{P}$$

and the above implies $\mathbb{Q} \gg \mathbb{P}$, $\mathbb{P} \gg \mathbb{Q}$ and $\rho_\infty$is a density $\rho_\infty = \frac{d\mathbb{Q}}{d\mathbb{P}}$. Naturally, we are interested in the process given by conditioning ρ_∞ on $(\mathcal{F}_t)$.

LEMMA 3.25. *Let $\mathbb{Q} \sim \mathbb{P}$ and denote$\rho_\infty := \frac{d\mathbb{Q}}{d\mathbb{P}}$. Then*

(1) ***(conditional Bayes)*** $\mathbb{E}_{\mathbb{Q}}[Z|\mathcal{F}_t] = \frac{\mathbb{E}_{\mathbb{P}}[\rho_\infty Z|\mathcal{F}_t]}{\mathbb{E}_{\mathbb{P}}[\rho_\infty|\mathcal{F}_t]}$ *for any real-valued, bounded random variable Z,*
(2) *the process[1] $\rho = (\rho_t)_{t\geq 0} := (\mathbb{E}_{\mathbb{P}}[\rho_\infty|\mathcal{F}_t])_{t\geq 0}$ is a non-negative, u.i. $\mathbb{P}$-martingale,*
(3) *the process $\frac{1}{\rho} := \left(\frac{1}{\rho_t}\right)$ is a u.i. $\mathbb{Q}$-martingale and fulfills $\frac{1}{\rho_t} = \mathbb{E}_{\mathbb{Q}}\left[\frac{1}{\rho_\infty}|\mathcal{F}_t\right]$.*

PROOF. Since $\frac{\mathbb{E}_{\mathbb{P}}[\rho_\infty Z|\mathcal{F}_t]}{\mathbb{E}_{\mathbb{P}}[\rho_\infty|\mathcal{F}_t]}$ is $(\mathcal{F}_t)$-measurable

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}\left[\rho_\infty \frac{\mathbb{E}_{\mathbb{P}}[\rho_\infty Z|\mathcal{F}_t]}{\mathbb{E}_{\mathbb{P}}[\rho_\infty|\mathcal{F}_t]}|\mathcal{F}_t\right] &= \mathbb{E}_{\mathbb{P}}[\rho_\infty|\mathcal{F}_t]\frac{\mathbb{E}_{\mathbb{P}}[\rho_\infty Z|\mathcal{F}_t]}{\mathbb{E}_{\mathbb{P}}[\rho_\infty|\mathcal{F}_t]} \\
&= \mathbb{E}_{\mathbb{P}}[\rho_\infty Z|\mathcal{F}_t]
\end{aligned}$$

or if we write for brevity $C_t := \frac{\mathbb{E}_{\mathbb{P}}[\rho_\infty Z|\mathcal{F}_t]}{\mathbb{E}_{\mathbb{P}}[\rho_\infty|\mathcal{F}_t]}$ this reads

$$\mathbb{E}_{\mathbb{P}}[\rho_\infty C_t|\mathcal{F}_t] = \mathbb{E}_{\mathbb{P}}[\rho_\infty Z|\mathcal{F}_t].$$

[1]Recall that conditional expectation is only well-defined up to equivalence class. By convention 3, we always work with the cadlag version of a martingale.

Hence for $A \in \mathcal{F}_t$,

$$\mathbb{E}_{\mathbb{Q}}[1_A C] = \mathbb{E}_{\mathbb{P}}[\rho_\infty 1_A C] = \mathbb{E}_{\mathbb{P}}[1_A \mathbb{E}_{\mathbb{P}}[\rho_\infty Z|\mathcal{F}_t]] = \mathbb{E}_{\mathbb{P}}[1_A \rho_\infty Z] = \mathbb{E}_{\mathbb{Q}}[1_A Z]$$

This characterizes the conditional expection, that is $C = \mathbb{E}_{\mathbb{Q}}[X|\mathcal{F}_t]$.

The second point follows directly from the definition resp. Radon–Nikodym theorem. For the last statement, note that $\frac{d\mathbb{P}}{d\mathbb{Q}} = \frac{1}{\frac{d\mathbb{Q}}{d\mathbb{P}}} = \frac{1}{\rho_\infty}$ and by point (1) applied with $Z = 1$,

$$\mathbb{E}_{\mathbb{Q}}\left[\frac{d\mathbb{P}}{d\mathbb{Q}}|\mathcal{F}_t\right] = \frac{\mathbb{E}[1|\mathcal{F}_t]}{\mathbb{E}\left[\frac{d\mathbb{Q}}{d\mathbb{P}}|\mathcal{F}_t\right]} = \frac{1}{\rho_t}.$$

Especially $\frac{1}{\rho_t}$ is given as a conditional expectation, thus $\frac{1}{\rho}$ is a u.i. martingale. □

Proposition 3.26. *Let $\mathbb{Q} \sim \mathbb{P}$ and $\rho = (\rho_t)$ be as in Lemma 3.25. If X is an adapted process on $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ then TFAE*

(1) *X is a $\mathbb{Q}$-local martingale,*
(2) *$\rho X = (\rho_t X_t)_{t\geq 0}$ is $\mathbb{P}$-local martingale,*

Proof. Wlog $X_0 = 0$. We first prove it for martingales instead of local martingales. Since

$$|X_t|_{L^1(\mathbb{Q})} = \mathbb{E}_{\mathbb{P}}[\rho_\infty |X_t|] = \mathbb{E}_{\mathbb{P}}[|X_t| \mathbb{E}_{\mathbb{P}}[\rho_\infty|\mathcal{F}_t]] = |\rho_t X_t|_{L^1(\mathbb{P})}$$

we have$|X_t|_{L^1(\mathbb{Q})} < \infty$ iff $|\rho_t X_t|_{L^1(\mathbb{P})} < \infty$. By Lemma 3.25, point (1)

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}[\rho_t X_t|\mathcal{F}_s] &= \mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\mathbb{P}}[\rho_\infty X_t|\mathcal{F}_t]|\mathcal{F}_s] \\ &= \mathbb{E}_{\mathbb{P}}[\rho_\infty X_t|\mathcal{F}_s] \\ &= \rho_s \mathbb{E}_{\mathbb{Q}}[X_t|\mathcal{F}_s] \end{aligned}$$

hence $\mathbb{E}_{\mathbb{P}}[\rho_t X_t|\mathcal{F}_s] = \rho_s X_s$ iff $\mathbb{E}_{\mathbb{Q}}[X_t|\mathcal{F}_s] = X_s$.

To extend to local martingales we need to prove that given a stopping time τ, X^τ is $\mathbb{Q}$-martingale iff $(\rho X)^\tau$ is a $\mathbb{P}$-martingale. We have just seen that X^τ being a $\mathbb{Q}$-martingale is equivalent to ρX^τ being a $\mathbb{P}$-martingale. It therefore remains to show that $(\rho X)^\tau$ is a $\mathbb{P}$-martingale iff ρX^τ is a $\mathbb{P}$-martingale. Since ρ is a nonnegative u.i. martingale (Proposition 3.8 point 4.) we use OST to see $\mathbb{E}_{\mathbb{P}}[\rho_t|\mathcal{F}_\tau] = \rho_{t\wedge\tau}$ and

$$\mathbb{E}_{\mathbb{P}}[|\rho_t X_{t\wedge\tau}|] = \mathbb{E}_{\mathbb{P}}[|X_{t\wedge\tau}| \mathbb{E}[\rho_t|\mathcal{F}_\tau]] = \mathbb{E}_{\mathbb{P}}[|\rho_{t\wedge\tau} X_{t\wedge\tau}|] = \mathbb{E}_{\mathbb{P}}\left[\left|(\rho X)_t^\tau\right|\right].$$

Hence, $\left|\rho_t X_t^\tau\right|_{L^1(\mathbb{P})} < \infty$ iff $\left|(\rho X)_t^\tau\right|_{L^1(\mathbb{P})} < \infty$ and in this case $M := \rho X^\tau - (\rho X)^\tau$ is a martingale:

$$\begin{aligned} \mathbb{E}[M_t|\mathcal{F}_s] &= \mathbb{E}[(\rho_t - \rho_{t\wedge\tau}) X_{t\wedge\tau}|\mathcal{F}_s] \\ &= \mathbb{E}\left[\mathbb{E}\left[\rho_t - \rho_{t\wedge\tau}|\mathcal{F}_{t\wedge(s\vee\tau)}\right] X_{t\wedge\tau}|\mathcal{F}_s\right] \\ &= \mathbb{E}\left[\left(\rho_{\mathcal{F}_{t\wedge(s\vee\tau)}} - \rho_{t\wedge\tau}\right) X_{t\wedge\tau}|\mathcal{F}_s\right] \\ &= (\rho_s - \rho_{s\wedge\tau}) X_{s\wedge\tau} = M_s. \end{aligned}$$

Thus ρX^τ is a martingale iff $(\rho X)^\tau$ is a martingale. □

Below we show that such results hold in great generality: given a local martingale on our filtered probabilty space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ we can remove/add a bounded variation process by a change to an absolutely continuous measure.

Theorem 3.27. *Let $\mathbb{P} \sim \mathbb{Q}$ on $(\Omega, \mathcal{F}, (\mathcal{F}_t))$. Assume $\rho = (\rho_t)_{t\geq 0} := (\mathbb{E}_{\mathbb{P}}[\rho_\infty|\mathcal{F}_t])_{t\geq 0}$ is continuous. If X is a local continuous $\mathbb{P}$-martingale then there exists a local continous $\mathbb{Q}$-martingale Y such that*

$$(X_t)_{t\geq 0} = \left(Y_t + \int_0^t \rho_s^{-1} d\langle \rho, X\rangle_s\right)_{t\geq 0}$$

holds with probability one.

Proof. For brevity define the bounded variation process $C_t := \int_0^t \rho_s^{-1} d\langle \rho, X\rangle_s$ and set $Y := X - C$. By Lemma 3.26 it is enough to show that $\rho Y = (\rho_t Y_t)_{t\geq 0}$ is a local continuous $\mathbb{P}$-martingale. By definition of Y

$$\rho Y = \rho(X - C) = \rho X - \langle \rho, X\rangle - (\rho C - \langle \rho, X\rangle). \tag{3.7.1}$$

The first term $\rho X - \langle \rho, X\rangle$ is a continuous local $\mathbb{P}$-martingale by characterization of the bracket. By Ito and using that $\langle \rho, C\rangle = 0$ due to the finite variation of C,

$$\begin{aligned} d(\rho C)_t &= \rho_t \rho_t^{-1} d\langle \rho, X\rangle_t + C_t d\rho_t + \frac{1}{2}\langle \rho, C\rangle_t \\ &= \langle \rho, X\rangle_t + C_t d\rho_t. \end{aligned}$$

The local martingale is preserved under stochastic integration, hence $C \bullet \rho$ is a continuous local $\mathbb{P}$-martingale and therefore $\rho C - \langle \rho, X\rangle$ is also continuous local $\mathbb{P}$-martingale. By (3.7.1), ρY is a continuous local $\mathbb{P}$-martingale. □

Remark 3.28.

- Above proof immediately extends to non-continuous local martingales and ρ being cadlag if the Ito-formula is proven for non-continuous processes (which we have not done).
- All results in this section carry over to the multidimensional case in a straightforward fashion.

How to construct an equivalent probability measure? Given $\mathbb{P}$ and a positive u.i. martingale ρ we can define an equivalent measure by choosing a positive random variable $\rho_\infty := \lim_{t\to\infty} \rho_t$ and take $d\mathbb{Q} := \rho_\infty d\mathbb{P}$. A natural candidate for ρ is the stochastic exponential. In this case above reduces to Girsanov's theorem for which the finite variation term takes an especially simple form.

Theorem 3.29 (Girsanov). *Let X be a continous, local $\mathbb{P}$-martingale and $K \in L^2(X)$. If*

$$\rho := \mathcal{E}(K \bullet X) \text{ is a continuous u.i. martingale}$$

then

1. *ρ_∞ exists, $\mathbb{E}[\rho_\infty] = 1$ and $d\mathbb{Q} := \rho_\infty d\mathbb{P}$ fulfills $\mathbb{Q} \sim \mathbb{P}$,*
2. *There exists a $\mathbb{Q}$-local martingale Y such that*

$$(X_t)_{t\geq 0} = \left(Y_t + \int_0^t K_s d\langle X\rangle_s\right)_{t\geq 0}$$

with probability one.

Proof. By assumption, ρ is u.i. and therefore $\rho_\infty = \mathcal{E}(K \bullet X)_\infty$ exists, $\rho_t = \mathbb{E}_{\mathbb{P}}[\rho_\infty|\mathcal{F}_t]$ is positive. Thus $d\mathbb{Q} := \rho_\infty d\mathbb{P}$ defines an equivalent measure. We now apply Theorem 3.27: the stochastic exponential fulfills

$$d\rho_t = \rho_t d(K \bullet X)_t = \rho_t K_t dX_t$$

and using the characterization of the stochastic integral, Theorem 2.31,

$$\langle \rho, X\rangle = \langle (\rho K) \bullet X, X\rangle = \rho K \langle X\rangle.$$

Hence, $\int \rho_s^{-1} d\langle \rho, X\rangle_s = \int K_s d\langle X\rangle_s$. □

If we specialize to Brownian motion, this transformation gives another Brownian motion with drift.

Theorem 3.30 (Cameron–Martin). *Let B be a $\mathbb{P}$-Brownian motion and $K \in L^2(B)$. If*

$$\rho := \mathcal{E}(K \bullet B) \text{ is a continuous u.i. martingale}$$

then

1. *ρ_∞ exists, $\mathbb{E}[\rho_\infty] = 1$ and $d\mathbb{Q} := \rho_\infty d\mathbb{P}$ fulfills $\mathbb{Q} \sim \mathbb{P}$,*

(2) *There exists a $\mathbb{Q}$-Brownian motion W such that with probability one,*

$$(B_t)_{t\geq 0} = \left(W_t + \int_0^t K_s ds\right)_{t\geq 0} \tag{3.7.2}$$

Proof. We apply Theorem 3.29 and note that $\langle B \rangle = t$ a.s. To see that $W := B - \int K ds$ must be a Brownian motion, note that $\langle W \rangle = \left\langle B - \int_0^\cdot K_s ds \right\rangle = \langle B \rangle$ since finite variation processes have zero qudratic variation. Further, W is continuous, hence the result follows by Levy's characterization, Theorem 3.16. □

Remark 3.31.

(1) We just saw that the question if a stochastic exponential is a martingale (and not just a local martingale) arises naturally. In general this is a tough question but recall that Proposition 3.10 guve us sufficient criteria (Novikov and Kazamaki) that cover many examples; see also Proposition 3.8.
(2) The Cameron–Martin theorem can be strengthened: not only does Brownian motion transform to a Brownian motion plus a drift term under an equilavent measure change using the stochastic exponential, but every equivalent change of measure is of the type (3.7.2)! The Brownian (also called Wiener) measure can be thought of as the natural Gaussian measure on pathspace and this property of equivalence under addition is generic for Gaussian measures in finite as well as infinite dimensions. This phenomenon runs under the name "quasi-invariance"; the measures are not invariant under addition/translation but "quasi" invariant since null sets do not change and they remain Gaussian. We refer the motivated reader to the (very challenging, graduate level) book [**Malliavin, 1997**] for more details.
(3) Above results give us an elegant way to transform certain semimartingales into local martingales.

CHAPTER 4

Stochastic Differential Equations

We have already seen an examples of an SDEs of the type

$$dX_t = \mu(t, X_t)\,dt + \sigma(t, X_t)\,dB_t \tag{4.0.1}$$

and found explicitly a solution X by using Ito's formula. For example, $\mu(t,x) = 0$ and $\sigma(t,x) = x$ gives the stochastic exponential $\mathcal{E}(B)$. As in the case of ordinary differential equations ($\sigma = 0$) we cannot hope to find a solution in explicit form for general coefficients μ,σ. Instead we now study the general well-posedness of (4.0.1), that is excistence and uniqueness. More generally, we treat the multidimensional case that is we are interested in finding a e-dimensional process X such that

$$X_t^i = X_0^i + \int_0^t \mu_i(t, X_t)\,dt + \sum_{j=1}^d \int_0^t \sigma_{i,j}(t, X_t)\,dB_t^i,\ j = 1, \dots, e$$

with $B = \left(B_t^1, \dots, B_t^d\right)_{t\geq 0}$ a d-dimensional Brownian motion, $\mu = (\mu_i)$ is a vector-valued function, $\sigma = \left(\sigma_{i,j}\right)$ is a matrix-valued function. We also write the above in differential notation as

$$dX_t = \mu_t(t, X_t)\,dt + \sigma_t(t, X_t)\,dB_t.$$

As a first step we have to make precise what we exactly mean by a solution.

Definition 4.1. Let $\mu = (\mu_i) : [0,\infty)\times\mathbb{R}^e \to \mathbb{R}^e$ and $\sigma = \left(\sigma_{i,j}\right) : [0,\infty)\times\mathbb{R}^e \to \mathbb{R}^{e\times d}$ be Borel-measurable functions and $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ a filtered probability space. We call a pair of $(\mathcal{F}_t)$-adapted processes (X, B) defined on $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ a *solution of the stochastic differential equation* $\mathrm{e}(\mu, \sigma)$ *on* $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ *with initial condition* X_0 if

- B is a $(\mathcal{F}_t)$-standard BM in $\mathbb{R}^d$
- for every $i = 1, \dots, e$,

$$X_t^i = X_0^i + \int_0^t \mu_i(s, X)\,ds + \sum_{j=1}^d \int_0^t \sigma_{ij}(s, X)\,dB_s^j \tag{4.0.2}$$

We use the notation $\mathrm{e}_x(\mu, \sigma)$ to impose the constraint $X_0 = x$ a.s., $x \in \mathbb{R}^e$.

Intuitively we want that the only "source of randomness" that X reacts to is B, that is X is adapted to to the subfiltration of $(\mathcal{F}_t)$ that is generated by the Brownian motion.

Definition 4.2. We say that a solution (X, B) defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ is a *strong solution on* $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ if X is adapted to the completed filtration generated by the Brownian motion B. If (X, B) is a solution that is not a strong solution then we say that (X, B) is a *weak solution.*

Strong solvabilty implies weak solvability and looking at (4.0.2), it might be at first sight counterintuitve that SDEs exist that have a weak solution but no strong solution. A famous example due to Tanaka shows that this can already happen in rather simple situations.

Example 4.3 (Tanaka's SDE). Set $\sigma(x) := \mathrm{sgn}(x)$ and consider a solution (X, B) of $\mathrm{e}_0(0, \sigma)$.

- There exists a weak solution but not a strong solution

Similarly, the issue of uniqueness is also more subtle than in the deterministic case: do we just take into account distributional properties of two different solutions or do we care about pathwise behaviour (same trajectories for a.e. ω)?.

Definition 4.4.

(1) We say that *pathwise uniqueness* holds for $\mathrm{e}(\mu,\sigma)$, if (X,B) and (X',B) are two solutions defined on the same filtered probability space, then $X_0 = X_0'$ a.s. implies that $X = X'$ a.s.
(2) We say that *uniqueness in law* holds for $\mathrm{e}(\mu,\sigma)$ if (X,B) and (X',B') are two solutions, defined on possibly different filtered probability spaces, then$X_0 = X_0'$ in distribution implies that X equals X' in distribution

Example 4.5. Consider again Tanaka's SDE $\mathrm{e}_0(0,\sigma)$. If (X,B) is a solution then

$$X = \mathrm{sgn}(X) \bullet B$$

and therefore $X_0 = 0$, $X \in \mathcal{M}_{\mathrm{c,loc}}$ so by Levy's characterization X is a Brownian motion. Moreover,

- uniqueness in law holds: by the above X must be a Brownian motion
- pathwise uniqueness does not hold,
- By the above there exists a weak solution but no strong solution

We state the following theorem without proof.

Theorem 4.6 (Yameda–Watanabe). *If pathwise uniqueness holds for* $\mathrm{SDE}(\mu,\sigma)$ *then*

(1) *uniqueness in law holds for* $\mathrm{SDE}(\mu,\sigma)$,
(2) *every solution to* $\mathrm{SDE}(\mu,\sigma)$ *is a strong solution.*

4.1. Strong solutions

We proceed as in the ODE case: use the integral formulation and show that a Picard iteration is a contraction. To achieve this we need Lipschitz regularity of the coefficient.

Theorem 4.7. *Assume μ,σ are both Lipschitz continuous and of linear growth in space uniformly over time, that is there exists a $c > 0$ such that $\forall x, y \in \mathbb{R}^e, \forall t \geq 0$*

$$|\mu(t,x) - \mu(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq c\,|x-y|$$

and

$$|\sigma(t,x)| + |\mu(t,x)| \leq c\,(1+|x|)\,.$$

Assume there exists a filtered probability space $(\Omega,\mathcal{F},(\mathcal{F}_t),\mathbb{P})$ on which a $\mathcal{F}_0$-measurable $\mathbb{R}^e$-valued random variable ξ with $\mathbb{E}\left[|\xi|^2\right] < \infty$ and a d-dimensional Brownian motion are defined. Then the SDE $\mathrm{e}(\mu,\sigma)$ has strong solution with initial condition $X_0 = \xi$ and pathwise unique holds.

Proof. For brevity of notation we give the proof for $d = e = 1$, $\mu = 0$, $\xi = x$ for $x \in \mathbb{R}^e$; the modifications for the general are straightforward.

Existence. Take a filtered probability space $(\Omega,\mathcal{F},(\mathcal{F}_t),\mathbb{P})$ that carries a Brownian motion B where $(\mathcal{F}_t)$ denotes the filtration generated by B (as usual, completed to fufill the usual conditions). Fix $T > 0$ and note that $B^T = (B_{t\wedge T}) \in H^2$. We define the Picard iteration as

$$X^{n+1} := \left(x + \int_0^t \sigma(r, X_r^n)\, dB_r^T\right)_{t\geq 0} \tag{4.1.1}$$

The sequence (X^n) given by (4.1.1) is well defined: $B^T \in H^2$ and if $X^n \in H^2$ then $\sigma(r,X^n) \in L^2\left(B^T\right)$ since the linear growth implies

$$\left|(\sigma(t,X_t))_{t\geq 0}\right|_{L^2(B^T)} = \mathbb{E}\left[\int_0^\infty |\sigma(r,X_r)|^2\, d\left\langle B^T\right\rangle_r\right] = \mathbb{E}\left[\int_0^T |\sigma(r,X_r)|^2\, dr\right] \leq c^2\mathbb{E}\left[\int_0^T (1+|X_r^n|)^2\, dr\right] < \infty$$

Thus $X^n \in H^2$ implies $X^{n+1} \in H^2$ since stochastic integration preserves H^2, Theorem 2.28. Since $X^0 = x \in H^2$ we get by induction $(X^n)_{n\geq 0} \subset H^2$.

For brevity, set

$$M_{n,t} := \left(X^n - X^{n-1}\right)^\star_T$$

Claim 1: $\left|M_{n,t}\right|_{L^2} \to_n 0$.

Apply Doob's inequality, Ito's isometry and the Lipschitz assumption to get

$$\begin{aligned}
\left|M_{n+1,T}\right|^2_{L^2} &= \mathbb{E}\left[\sup_{s\in[0,T]}\left|\int_0^s \left(\sigma(r,X^n_r) - \sigma\left(r,X^{n-1}_r\right)\right) dB_r\right|^2\right] \\
\text{(Doob)} \quad &\leq 4\mathbb{E}\left[\left(\int_0^T \sigma(r,X^n_r) - \sigma\left(r,X^{n-1}_r\right) dB_r\right)^2\right] \\
\text{(Ito isometry)} \quad &\leq 4\mathbb{E}\left[\int_0^T \left(\sigma(r,X^n) - \sigma\left(r,X^{n-1}\right)\right)^2 dr\right] \\
\text{(Lipschitz)} \quad &\leq 4c^2\mathbb{E}\left[\int_0^T \left(X^n_r - X^{n-1}_r\right)^2 dr\right] \\
&\leq 4c^2\mathbb{E}\left[\int_0^T M^2_{n,t}\, dt\right].
\end{aligned} \tag{4.1.2}$$

Iteration yields

$$\begin{aligned}
\left|M_{n+1,T}\right|^2_{L^2} &\leq \left(4c^2\right)^n \int_0^T \int_0^{t_{n-1}} \cdots \int_0^{t_2} \left|M_{1,t_1}\right|^2_{L^2} dt_1 \cdots dt_n \\
&\leq \frac{\left(4c^2T\right)^n}{n!}\left|M_{1,T}\right|^2_{L^2}.
\end{aligned}$$

Now, $\sum_{n\geq 1}\left|M_{n,T}\right|^2_{L^2} \leq e^{4c^2T}\left|M_{1,T}\right|^2_{L^2} < \infty$, hence we must have $\left|M_{n,T}\right|^2_{L^2} \to 0$.

Claim 2: There exists a continuous, adapted process $X = (X_t)_{t\geq 0}$ on $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ such that

$$(X^n - X)^\star_T \quad \to_n \quad 0 \text{ a.s.}$$

To see this, apply Chebycheff's inequality and then above estimate to get

$$\begin{aligned}
\mathbb{P}\left(M_{n+1,T} > 2^{-(n+1)}\right) &\leq \frac{\mathbb{E}\left[M^2_{n+1,T}\right]}{2^{2(n+1)}} \\
&\leq 4\frac{\left(4^2c^2T\right)^n}{n!}\left|M_{1,T}\right|^2_{L^2}
\end{aligned}$$

Summing up shows $\sum_{n\geq 0}\mathbb{P}\left(M_{n+1,T} > 2^{-(n+1)}\right) < \infty$. By Borell-Cantelli it follows that

$$\mathbb{P}\left(M_{n+1,T} \leq 2^{-(n+1)}\text{for infinitely many } n \in \mathbb{N}\right) = 0.$$

Hence, there exists a random variable N that is finite a.s., such that for a.e. ω

$$M_{n+1,T}(\omega) \leq 2^{-(n+1)} \text{ for all } n \geq N(\omega).$$

We use this and the triangle inequality, to conclude that for any $k \geq 1$,

$$\begin{aligned}
\left(X^{n+k} - X^n\right)^\star_T(\omega) &\leq \left(X^{n+k} - X^{n+k-1}\right)^\star_T(\omega) + \cdots + \left(X^{n+1} - X^n\right)^\star_T(\omega) \\
&\leq \sum_{i=n+1}^{n+k} 2^{-i} = 2^{-(n+1)}\sum_{i\geq 0} 2^{-i} = 2^{-n}
\end{aligned}$$

holds for all $n > N(\omega)$. Since the right hand side does not dependend on k,

$$\sup_{k\geq 1}\left(X^{n+k}-X^{n}\right)_{T}^{\star}(\omega)\leq 2^{-n}\text{ for all } n\geq N(\omega).$$

Therefore, we have shown $(X^n(\omega))$ is almost surely a Cauchy sequence in $(C([0,T],\mathbb{R}^e),|.|_\infty)$. By completeness of $(C([0,T],\mathbb{R}),|.|_\infty)$ there exists for a.e. ω a $X(\omega)\in C([0,T],\mathbb{R})$ such that

$$(X^n-X)_T^{\star}(\omega)\rightarrow 0.$$

Since the above argument holds for every $T>0$, X_t is well defined for every $t\geq 0$. Moreover, it is adapted by adaptedness of (X^n).

Claim 3. (X,B) is a solution of $\mathrm{e}_x(\sigma,\mu)$.

We estimate as above

$$\begin{aligned}(4.1.3)\qquad \mathbb{E}\left[\left|\int_0^T\sigma(r,X^n)-\sigma(r,X)\,dB_r\right|^2\right] &= \mathbb{E}\left[\int_0^T|\sigma(r,X^n)-\sigma(r,X)|^2\,dr\right]\\ &\leq c^2T\mathbb{E}\left[\sup_{r\in[0,T]}|X_r^n-X_r|^2\right].\end{aligned}$$

It remains to show the convergence as $n\rightarrow\infty$ on the right hand side. Therefore use the triangle inequality to write

$$\sqrt{\mathbb{E}\left[\left|(X^m-X^n)_T^{\star}\right|^2\right]}\leq\sum_{k=m}^{n-1}\sqrt{\mathbb{E}\left[\left|\left(X^{k+1}-X^k\right)_T^{\star}\right|^2\right]}=\sum_{k=m}^{n-1}\left|M_{k+1,T}\right|_{L^2}<\infty.$$

Let first $n\rightarrow\infty$, (using Fatou's lemma) shows $\sqrt{\mathbb{E}\left[\sup_{t\in[0,T]}\left|X_t^m-X_t\right|^2\right]}\leq\sum_{k\geq m}\left|M_{k+1,T}\right|_{L^2}$ and using $|(a_n)|_{\ell^2}=\sqrt{\sum a_n^2}\leq\sum|a_n|=|(a_n)|_{\ell^1}$ we get $\sqrt{\mathbb{E}\left[\sup_{t\in[0,T]}\left|X_t^m-X_t\right|^2\right]}\leq\sqrt{\sum_{k\geq m}\left|M_{k+1,T}\right|_{L^2}^2}$ and by the results in (claim 1), the right term converges to 0 as $m\rightarrow\infty$. Thus the L^2 convergence of (4.1.3) follows. Since L^2-convergence implies a.s. convergence along a subsequence, we have shown the a.s. convergence of

$$X_t^{n+1}=x+\int_0^t\sigma\left(s,X_s^n\right)dB_s$$

to

$$X_t=x+\int_0^t\sigma(s,X_s)\,dB_s.$$

Thus (X,B) is a strong solution.

Pathwise uniqueness: follows from Lemma (4.9) by letting $r\rightarrow\infty$. □

We first recall Gronwall's Lemma

LEMMA 4.8. *Let $\varphi:[0,\infty)\rightarrow[0,\infty)$ Borel measurable and locally bounded. If there exists $a,b\in\mathbb{R}$, $b\geq 0$ such that*

$$\varphi(t)\leq a+b\int_0^t\varphi(r)\,dr\text{ for all }t\geq 0$$

then

$$\varphi(t)\leq ae^{bt}.$$

If $a=0$, it follows that $\varphi=0$.

LEMMA 4.9. *Let μ,σ be as above and X and X' be two continuous adapted processes adapted on $(\Omega,\mathcal{F},(\mathcal{F}_t),\mathbb{P})$ and B a Brownian motion. For $r>0$ set $\tau_r := \inf\{t>0 : |X_t|\vee|Y_t|>r\}$. If X and Y satisfy*

$$\begin{aligned} X_t &= x+\int_0^t \mu(t,X_r)\,dr+\int_0^t \sigma(r,X_r)\,dB_r \\ Y_t &= x+\int_0^t \mu(t,Y_r)\,dr+\int_0^t \sigma(r,Y_r)\,dB_r \end{aligned}$$

on $[0,\tau_r]$ then $X=Y$ on $[0,\tau_r]$.

PROOF. Again wlog $\mu=0$ and let $\varphi(t) := \left|(X-Y)^\star_{t\wedge\tau_r}\right|^2_{L^2}$. The same estimates (Doob,Ito isometry, Lipschitz) as used above give

$$\begin{aligned} \varphi(t) &\le c^2\mathbb{E}\left[\int_0^{\tau_r\wedge t}|X_s-Y_s|^2\,ds\right] \\ &= c^2\mathbb{E}\left[\int_0^t \left|X_{s\wedge\tau_r}-Y_{s\wedge\tau_r}\right|^2 ds\right] \\ &\le c^2\mathbb{E}\left[\int_0^t \left|(X-Y)^\star_{s\wedge\tau_r}\right|^2 ds\right] = c^2\int_0^t \varphi(r)\,dr. \end{aligned}$$

Thus by Gronwall's lemma, $\varphi=0$. □

REMARK 4.10. As in the ODE case:

- Lipschitzness guarentees uniqueness, e.g. $\mu(t,x)=\sqrt{x}$, $\sigma(t,x)=0$ has infinitely many solutions, $X_t=\frac{1}{4}\left(t^2+2at+b\right)$ for a,b constant.
- Linear growth guarantees global existence, e.g. $\mu(t,x)=x^2$, $\sigma(t,x)=0$ has as solution $X_t=\frac{1}{\frac{1}{x}-t}$ and $\lim_{t\to\frac{1}{x}} X_t=\infty$.
- These conditions are sufficient but not necessary.

REMARK 4.11. Sometimes one is only interested in non-global strong solutions, that is process that are strong solutions only up to a stopping time τ. Note Lemma 4.9 is formulated general enough to also give pathwise uniqueness in this setting.

THEOREM 4.12. *Let μ,σ be as in Theorem 4.7 and additionally bounded. Then there exists a jointly continuous process $X=\left(X_t^x\right)_{t\ge0,x\in\mathbb{R}^e}$ such that*

$$X_t^x = x+\int_0^t \mu(r,X_r^x)\,dr+\int_0^t \sigma(r,X_r^x)\,dB_r.$$

.

PROOF. Again wlog $\mu=0$. Fix $p>2$, $t>0$. Since $|a+b|^p\le 2^{p-1}(|a|^p+|b|^p)$ we have

$$\begin{aligned} \sup_{s\in[0,t]}\left|X_s^x-X_s^y\right|^p &\le 2^{p-1}\Bigg(|x-y|^p \\ &\quad + \sup_{s\in[0,t]}\left|\int_0^s \left(\sigma(r,X_r^x)-\sigma\left(r,X_r^y\right)\right)dB_r\right|^p\Bigg) \end{aligned}$$

Using BDG for the last term gives

$$\mathbb{E}\left[\sup_{s\in[0,t]}\left|\int_0^s \left(\sigma(r,X_r^x)-\sigma\left(r,X_r^y\right)\right)dB_r\right|^p\right] \le c_p\mathbb{E}\left[\left|\int_0^t \left(\sigma(r,X_r^x)-\sigma\left(r,X_r^y\right)\right)^2 dr\right|^{p/2}\right]$$

Now apply Holder ($|fg|_{L^1} \le |f|_{L^q} |g|_{L^r}$ with $q^{-1}+r^{-1}=1$ applied with $r=\frac{p}{2}$, $q=\frac{p}{p-2}$ and $f=1$,$g_r = \left(\sigma(r,X_r^x)-\sigma\left(r,X_r^y\right)\right)^2$) and Lipschitzness to further estimate

$$\begin{aligned}\mathbb{E}\left[\sup_{s\in[0,t]}\left|\int_0^s\left(\sigma(r,X_r^x)-\sigma\left(r,X_r^y\right)\right)dB_r\right|^p\right] &\le c_p t^{\frac{p-2}{2}}\mathbb{E}\left[\int_0^t\left|\sigma(r,X_r^x)-\sigma\left(r,X_r^y\right)\right|^p dr\right]\\ &\le c_p t^{\frac{p-2}{2}} c_{Lip}\mathbb{E}\left[\int_0^t\left|X_r^x-X_r^y\right|^p dr\right]\\ &\le c_p t^{\frac{p-2}{2}} c_{Lip}\mathbb{E}\left[\int_0^t\left|(X^x-X^y)_r^\star\right|^p dr\right]\end{aligned}$$

The boundedness of σ ensures that $\langle X^x-X^y\rangle_t=\int_0^t\left(\sigma\left(s,X_s^x\right)-\sigma\left(s,X_s^y\right)\right)^2 ds < ct$, thus by BDG

$$\mathbb{E}\left[\left|(X^x-X^y)_t^\star\right|^p\right] \le \mathbb{E}\left[\left|\langle X^x-X^y\rangle_t\right|^{p/2}\right] < \infty.$$

Putting everthing together, shows there exists constants c_1,c_2 s.t.

$$\varphi(t) \le c_1^p|x-y|^p + c_2^p\int_0^t\varphi_r dr \text{ with } \varphi(t) := \left|(X^x-X^y)_t^\star\right|_{L^p}^{p/2}.$$

By Gronwall, there exists a $c_0=c_0(p,t)$ s.t. $\varphi(t)\le c_0^p|x-y|^p$. Using Kolmogorov's Theorem, there exists a modification of X that is continuous in both t and x. □

4.2. Weak solutions

We now are going to see that the assumptions on the coefficients μ,σ can be significantly weakened if we are only interested in in the well posedness of the SDE $\mathrm{e}(\mu,\sigma)$ in a weak sense, that is the existence of a weak solution (X,B) that is unique in law.

4.2.1. Weak solutions by change of measure: modifying drift. The Girsanov theorem gives an easy way to construct weak solutions: let μ,σ be measurable and assume we have found (X,B) on some filtered probability space on $(\Omega,\mathcal{F},(\mathcal{F}_t),\mathbb{P})$ such that

$$dX_t = \mu(t,X_t)dt + \sigma(t,X_t)dB_t. \tag{4.2.1}$$

If $\mathcal{E}(M)$ with $M_t=\int_0^t\beta(r,X_r)dr$ is a u.i. martingale (and this is in general a strong assumption) then we can define the equivalent measure

$$d\mathbb{Q} := \mathcal{E}(M)_\infty d\mathbb{P}.$$

Under $\mathbb{Q}$, the process B is a Brownian motion with drift, $B=W+\int_0^t\beta(r,X_r)dr$, hence

$$dX_t = (\mu(t,X_t)+\beta(t,X_t))dt + \sigma(t,X_t)dW_t.$$

To sum up,

$$e_x(\mu,\sigma) \text{ on } (\Omega,\mathcal{F},(\mathcal{F}_t),\mathbb{P}) \text{ is equivalent to } e_x(\mu+\beta,\sigma)\ (\Omega,\mathcal{F},(\mathcal{F}_t),\mathbb{Q}).$$

For example, if $\sigma(t,x)=1$ and $\beta:=-\mu$ this reduces the existence of a weak solution of $\mathrm{e}_x(\mu,1)$ to the existence of a filtered probability space $(\Omega,\mathcal{F},(\mathcal{F}_t),\mathbb{Q})$ that carries a Brownian motion (which is obviously true). Unfortunately it relies on the assumption that $\mathcal{E}(M)$ is a u.i. martingale and this is too strong for many applications (e.g. it does not hold if $\beta(t,x)=1$). There are essentially two ways to deal with this: either work on a finite time interval $[0,T]$ or alternatively carry out a "localized version of Girsanov" (that is construct a measure $\mathbb{Q}_T$ on every $[0,T]$ and then argue that there exists a $\mathbb{Q}$ such that $\mathbb{Q}|_{\mathcal{F}_t}=\mathbb{Q}_t$). The latter option involves some technicaliites but we refer the interested reader to xxx.

Proposition 4.13. *Fix $T>0$ and let $\mu:\mathbb{R}_+\times\mathbb{R}^e\to\mathbb{R}^e$ be bounded and measurable. Then the SDE $\mathrm{e}_x(\mu,1)$ has a weak solution on $[0,T]$.*

PROOF. Wlog $e=1$. Take a Brownian motion X on some probability space $(\Omega,\mathcal{F},(\mathcal{F}_t),\mathbb{Q})$. If the process $\mathcal{E}(M)$, $M_t := \int_0^t \mu(t,X_t)\,dX_t^T$, is a u.i. martingale then we can use Theorem 3.29 and use above construction using the change of measure $\frac{d\mathbb{Q}}{d\mathbb{P}} := \mathcal{E}_\infty(M)$ on $\left(\Omega,\mathcal{F}_T,(\mathcal{F}_t)_{t\in[0,T]}\right)$, that is under $\mathbb{P}$, X is a solution of $\mathrm{e}(\mu,1)$ on $[0,T]$. Therefore note that

$$\langle M\rangle_\infty = \left\langle \int_0^\cdot \mu(t,X_t)\,dX_t^T \right\rangle_\infty = \int_0^T \mu^2(r,X_r)\,dr \le \sup_{t,x}|\mu(t,x)|^2 T$$

thus $\mathbb{E}\left[\exp\left(\frac{1}{2}\langle M\rangle_\infty\right)\right] < \infty$ and by the Novikov condition $\mathcal{E}(M)$ is a u.i. martingale. □

4.3. Local time

Given $X\in\mathcal{S}_c$ and $t\ge 0$ and a set $A\subset\mathcal{B}(\mathbb{R})$ we wan to measure the time the process has spend on the interval $[0,t]$ in A. It is natural to do this with respect to the "inner clock" $\langle X\rangle$ of X, hence we define the so-called occupation measure at time t as

$$\nu_t(A) := \int_0^t 1_A(X_s)\,d\langle X\rangle_s .$$

For fixed t, ν_t is a random measure on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ and we can ask if it has a density denoted L_t^a with respect to Lebesgue measure, that is if

$$\nu_t(da) = L_t^a\,da.$$

In this case it should hold that

$$\frac{\nu_t([a,a+\epsilon))}{\epsilon} = \frac{1}{\epsilon}\int_0^t 1_{[a,a+\epsilon)}(X_s)\,d\langle X\rangle_s \rightarrow_{\epsilon\to 0} L_t^a$$

Provided the above can made rigorous (that is especially the existence of L_t^a) we see that informally the process $L=\left(L_t^a\right)_{t\ge 0, a\in\mathbb{R}^e}$ should fulfill

$$L_t^a = \int_0^t \delta_a(X_s)\,d\langle X\rangle_s = \lim_\epsilon \frac{1}{\epsilon}\int_0^t 1_{[a,a+\epsilon)}(X_s)\,d\langle X\rangle_s .$$

This explains why we will call L_t^a the local time of X on the interval $[0,t]$ in the point $a\in\mathbb{R}$. Above is obviously not rigourous (for example why should a density exist?) but we immediately see heuristcally a surprising connection with Ito's formula: if we could apply Ito to the non-smooth(!) function $|X_t-a|$ then

$$\begin{aligned} |X_t-a| &= |X_0-a| + \int_0^t \operatorname{sgn}(X_s-a)\,ds + \int_0^t \delta_a(X_s)\,d\langle X\rangle_s \\ &= |X_0-a| + \int_0^t \operatorname{sgn}(X_s-a)\,ds + L_t^a \end{aligned} \tag{4.3.1}$$

thus the local time appears as the second derivative in above generalized "Ito formula". We now make all this rigorous: we first show Ito's formula generalizes to convex functions, then take use this to define L via 4.3.1 and study its properties as a stochastic process. Finally we arrive at great generalization of Ito's formula, namely the Meyer–Tanaka formula.

We start with a lemma about convex functions: it tells us that convex functions are essentially twice differentiable if we are willing to think about the second derivative in a distributional sense, that is as a measure. Moreover, they remain convex under mollficiation.

LEMMA 4.14. *Let $f:\mathbb{R}\to\mathbb{R}$ be convex, that is for all $x,y\in\mathbb{R}$ and $\lambda\in[0,1]$*

$$f(\lambda x+(1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y).$$

Then

(1) *f has a left derivative f'_- and a right derivative f'_+, that is for all $x \in \mathbb{R}$ the limits $f'_-(x)$ and $f'_+(x)$*

$$\frac{f(x)-f(y)}{y-x} \nearrow f'_-(x) \text{ as } y \nearrow x,$$
$$\frac{f(y)-f(x)}{y-x} \searrow f'_+(x) \text{ as } y \searrow x,$$

exist in $\mathbb{R}$.

(2) *The left derivative $x \mapsto f'_-(x)$ is left continuous and increasing, the right derivative $x \mapsto f'_+(x)$ is right continuous and increasing and for $x < y$*

$$f'_+(x) \le \frac{f(y)-f(x)}{y-x} \le f'_-(y).$$

(3) *A fundamental theorem of calculus holds: $f(b)-f(a) = \int_a^b f'_-(x)\,dx = \int_a^b f'_+(x)\,dx$.*

(4) *$\mu([a,b)) := f'_-(b) - f'_-(a)$ defines a locally finite measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. We call μ the second derivative of f in distributional sense.*

(5) *Let $g \in C_c^\infty(\mathbb{R},\mathbb{R})$ with $\int_{\mathbb{R}} g(r)\,dr = 1$ and*

$$f_n(x) := (f \star g_n)(x) := \int_{\mathbb{R}} f(x-y)\,g_n(y)\,dy$$

with $g_n(x) := ng(nx)$. Then $f_n \in C^\infty(\mathbb{R},\mathbb{R})$ and for all compact intervals $[a,b]$,

$$\sup_{x\in[a,b]} |f(x) - f_n(x)| \quad \to \quad 0 \text{ as } n \to \infty$$

Proof. Consider $x_1 < x_2 < x_3$ and set $\lambda := \frac{x_2-x_1}{x_3-x_1}$ so that

$$x_2 = \lambda x_3 + (1-\lambda)x_1.$$

Set $y_2 := \lambda f(x_3) + (1-\lambda) f(x_1)$. By convexity $y_2 \ge f(x_2)$, hence we have

$$\frac{f(x_3)-f(x_2)}{x_3-x_2} \ge \frac{f(x_3)-y_2}{x_3-x_2} = \frac{f(x_3)-f(x_1)}{x_3-x_1} = \frac{y_2-f(x_1)}{x_2-x_1} \ge \frac{f(x_2)-f(x_1)}{x_2-x_1}$$

(draw a picture to see the equality in the middle). Therefore $x \mapsto \frac{f(x)-f(x_1)}{x_1-x}$ decreases as $x \searrow x$, thus the limit exists in $\mathbb{R} \cup \{\infty\}$ at x_1. However, it must be finite since by adding a point x_0 above argument applied to $x_0 < x_1 < x_2$ we can repeat the same argument. □

If f is twice differentiable then $f''(x)\,dx = \mu(dx)$, which justifies why we call μ the second derivative of f in distribution. Moreover, many properties of the second derivative transfer to μ, for eample integration by parts

$$\int_{\mathbb{R}} g(x)\,\mu(dx) = -\int_{\mathbb{R}} g'(x)\,f'_-(x)\,dx \text{ for } g \in C_c^\infty(\mathbb{R},\mathbb{R}).$$

Lemma 4.15. *Let $f : \mathbb{R} \to \mathbb{R}$ be convex and $X \in \mathcal{S}_c$. Then $f(X) \in \mathcal{S}_c$ and there exists a continuous, increasing process K^f such that*

$$f(X_t) = f(X_0) + \int_0^t f'_-(X_s)\,dX_s + K_t^f$$

where f'_- denotes the left derivative of f, $f_-(x) := \lim_{\epsilon\searrow 0} \frac{f(x)-f(x-\epsilon)}{\epsilon}$.

Proof. Denote the semimartingale decomposition $X = M + A$, $M \in \mathcal{M}_{c,loc}$, $A \in \mathrm{BV}$. By stopping at $\tau_n := \inf\{t : |X_t| \vee |A_t| \vee \langle M\rangle_t > n\}$ and subsequently letting $n \to \infty$ we can assume wlog that $|X_t| \le n, |A_t| \le n$, $\langle M\rangle_t \le n$ for all $t \ge 0$. Fix $g \in C_c^\infty(\mathbb{R},\mathbb{R})$, $\int g(x)\,dx = 1$ with support in $(-\infty, 0]$ and define

$$f_n(x) := (f \star g_n)(x) := \int_{\mathbb{R}} f(x-y)\,g_n(y)\,dy = \int_{-\infty}^0 f\left(x - \frac{y}{n}\right) g(y)\,dy \text{ with } g_n(x) := ng(nx).$$

By definition, f_n is twice differentiable and applying Ito yields

$$f_n(X_t) = f_n(X_0) + \int_0^t f_n'(X_s)\,dM_s + \int_0^t f_n'(X_s)\,dA_s + \frac{1}{2}\int_0^t f_n''(X_s)\,d\langle M\rangle_s . \tag{4.3.2}$$

Set

$$\begin{aligned} I^n &:= f_n'(X) \bullet M \text{ and } I := f_-'(X) \bullet M, \\ J^n &:= f_n'(X) \bullet A \text{ and } J := f_-'(X) \bullet A. \end{aligned}$$

Moreover, by Lemma 4.14 and since by assumption X is bounded we have

$$(f_n(X) - f(X))^\star_\infty \to 0 \text{ a.s.}$$

By definition of f_n it follows that

$$f_n'(x) \nearrow f(x)$$

We now show that $I^n \to I$, $J^n \to J$ uniformly which will imply via (4.3.2) that $\frac{1}{2}\int f_n''(X)\,d\langle M\rangle$ has to converge as well. Now $M \in H^2$ (recall Proposition $M \in H^2$ iff $\mathbb{E}[\langle M\rangle_\infty] < \infty$) and since X is bounded, $f_n'(X), f'(X) \in L^2(M)$, hence $I, I^n \in H^2$. Applying Doob and the Ito-isometry yields

$$\mathbb{E}\left[\left|(I_t^n - I_t)^\star_\infty\right|^2\right] \le 4\mathbb{E}\left[(I_\infty^n - I_\infty)^2\right] = \mathbb{E}\left[\int_0^\infty \left(f_n'(X_t) - f_-'(X_t)\right)^2 d\langle M\rangle_t\right].$$

By dominated convergence the rhs goes to 0 as $n \to \infty$. Since L^2-convergence implies a.s. convergence along a subsequence we have

$$(I^n - I)^\star_\infty \to 0 \text{ a.s.}$$

along some subequence. Similarly, with $J_n := \left(f_n'(X) - f_-'(X)\right) \bullet A$ we have again by dominated convergence that

$$(J_n - J)^\star_\infty \le \int_0^\infty \left|f_n'(X_s) - f_-'\left(X_s'\right)\right| dA_s \to 0 \ a.s.$$

To sum up, by switching to subsequence (henceforth again denoted with n) we have shown that the increasing, adapted, continuous process

$$\frac{1}{2}\int_0^t f_n''(X_s)\,d\langle X\rangle = f_n(X_t) - f_n(X_t) - \int_0^t f_n'(X_s)\,dM_s - \int_0^t f_n'(X_s)\,dA_s$$

converges uniformly in t as $n \to \infty$. We denote the limit by K^f and by uniform convergence it is again an increasing, adapted, continuous process. □

Above Lemma applied to $f(x) = |x - a|$ shows $f_-'(x) = \operatorname{sgn}(x)$ and $f'' = \delta_a$ in distributional sense. Hence we have given meaning to L and (4.3.1).

Definition 4.16. Let $X \in \mathcal{S}_c$ and $a \in \mathbb{R}$. We call the adapted, increasing process given by Lemma 4.15 K^f applied with $f(x) = |x - a|$ the *local time of X at a*. We also denote it as L^a.

Above tells us that

$$L_t^a = |X_t - a| - |X_0 - a| - \int_0^t \operatorname{sgn}(X_s)\,dX_s$$

where $\operatorname{sgn}(x) = 1_{x>0} - 1_{x\le 0}$.

Proposition 4.17. *Let $X \in \mathcal{S}_c(X)$, $a \in \mathbb{R}$.*

1. *If Lemma 4.15 is applied*[1] *with $f(x) = (x-a)^+$ or $f(x) = (x-a)^-$ then $K^f = \frac{1}{2}L^a$.*
2. *The local time L^a increases only at the time points $\{t : X_t = a\}$.*

[1] As usual $(x-a)^+ := (x-a)1_{x-a>0}$, $(x-a)^- := -(x-a)1_{x-a<0}$

PROOF. To see the first point, set $f_1(x) = (x-a)^+$, $f_2(x) = (x-a)^-$. Then $|x-a| = f_1(x) + f_2(x)$, hence $K^{f_1} + K^{f_2} = K^{f_1+f_2}$. Also $x - a = f_1(x) - f_2(x)$, hence by Ito applied to $X - a$, we have $X_t = X_0 + \int_0^t dX_s$, thus $K^{f_1} - K^{f_2} = K^{f_1-f_2} = 0$. Hence,

$$K^{f_1} = K^{f_2} = \frac{1}{2}K^{f_1+f_2} = \frac{1}{2}L^a.$$

For the second point, let σ,τ be two stopping times $\sigma \leq \tau$. By Lemma 4.15 applied to $f(x) = (x-a)^+$ we have

$$(X_\tau - a)^+ - (X_\sigma - a)^+ = \int_\sigma^\tau 1_{X_s > a} dX_s + \frac{1}{2}\left(L_\tau^a - L_\sigma^a\right).$$

If σ,τ are such that $[\sigma,\tau] \subset \{t : X_t < a\}$ then the lhs and the integral equals 0, hence $L_\tau^a = L_\sigma^a$. Especially this applies to

$$\sigma = \inf\left\{t > q : X_t \leq a - \frac{1}{m}\right\} \text{ and } \tau = \inf\left\{t > \sigma : X_t \geq a - \frac{1}{n}\right\}$$

for any choice of $m, n \in \mathbb{N}$, $m > n$ and $q \in \mathbb{Q}$. Taking the countable union over (m, n, q) shows shows that $t \mapsto L_t^a$ is constant on $\{t : X_t < a\}$. The same argument for $(x-a)^-$ shows that $t \mapsto L_t^a$ is constant on $\{t : X_t > a\}$. □

THEOREM 4.18 (MEYER–TANAKA). *Let $f : \mathbb{R} \to \mathbb{R}$ be the difference of two convex functions and $X \in \mathcal{S}_c$. Then $f(X) \in \mathcal{S}_c$ and*

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)\, dX_s + \frac{1}{2}\int_{\mathbb{R}} L_t^a \mu(da)$$

where f' denotes the left derivative of f and μ is the second derivative of f in distributional sense.

PROOF. Again by stopping we can assume wlog that $|X_t| \leq n$, $\langle X \rangle_t \leq n$ for $t \geq 0$. By linearity, it is sufficient to prove it for f a convex function. Set

$$g(x) := \frac{1}{2}\int_{-n}^n |x-a|\, \mu(da)$$

and note that g is convex with

$$g'(x) = \frac{1}{2}\int_{-n}^n \operatorname{sgn}(x-a)\, \mu(da) \tag{4.3.3}$$

and $g''(x)\, dx = \mu(dx)$. Hence $(f-g)'' = 0$ and

$$f(x) = g(x) + ax + b.$$

for some $a, b \in \mathbb{R}$. Again by linearity, it is sufficient to prove the formula separately for g and for $x \mapsto ax + b$. It is trivially true for the linear function, $aX_t + b = aX_0 + b + a\int_0^t dX_t$, so it just remains to show it for g. By definition of the local time L^a,

$$|X_t - a| - |X_0 - 0| = \int_0^t \operatorname{sgn}(X_s - a)\, dX_s + L_t^a. \tag{4.3.4}$$

By integrating (4.3.4) against μ over the interval $[-n, n]$, we have by (4.3.3) that

$$2g(X_t) - 2g(X_0) = \int_{-n}^n \int_0^t \operatorname{sgn}(X_s - a)\, dX_s \mu(da) + \int_{-n}^n L_t^a \mu(da).$$

We divide by 2 and apply the Fubini-theorem for semimartingales, see for example [**Durrett, 1996**, Chapter 2, Theorem 11.6], to interchange the order of integration to get

$$\begin{aligned} g(X_t) - g(X_0) &= \int_0^t \frac{1}{2}\int_{-n}^{n} \operatorname{sgn}(X_s - a)\,\mu(da)\,dX_s + \frac{1}{2}\int_{-n}^{n} L_t^a \mu(da) \\ &= \int_0^t g'(X_s)\,dX_s + \frac{1}{2}\int_{-n}^{n} L_t^a \mu(da). \end{aligned}$$

□

PROPOSITION 4.19 (OCCUPATION FORMULA). *Let $X \in \mathcal{S}_c$. Then for every bounded, measurable function $g : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ we have for all $t \geq 0$*

$$\int_{\mathbb{R}} g(x) L_t^x dx = \int_0^t g(X_s)\, d\langle X\rangle_s.$$

PROOF. By the monotone class theorem it is enough to show the formula for continuous g. But in the case of g is continuous, then above follows by comparing the Ito formula with the Meyer–Tanaka formula applied to the function f given by $f'' = g$. □

We think about the local time $L = \left(L_t^a\right)_{t\geq 0, a\in\mathbb{R}}$ as a stochastic process indexed by time and space. The propositions below shows that L is a continuous process.

PROPOSITION 4.20. *Let $M \in \mathcal{M}_{c,loc}$ and $L = \left(L_t^a\right)_{t\geq 0, a\in\mathbb{R}}$ its local time. Then there exists a modification of L such that $(t, a) \mapsto L_t^a$ is continuous.*

PROOF. As usual, wlog we assume that $M_t \leq n$ and $\langle M\rangle \leq n$. By Proposition 4.17

$$\frac{1}{2} L_t^a = (X_t - a)^+ - (X_0 - a)^+ - \int_0^t 1_{X_s > a} dX_s$$

The continuity of the first two terms is clear, so we just have to show that $(t,a) \mapsto I_t^a := \int_0^t 1_{X_s>a} dX_s$ has a continuous modification. To do so we fix $T > 0$ and apply Kolmogorov's criterion by regarding $a \mapsto I^a$ as map from $\mathbb{R} \to C([0,T], \mathbb{R})$. Therefore let $a < b$ and apply BDG to get

$$\begin{aligned} \mathbb{E}\left[\left|\left(I^a - I^b\right)_T^\star\right|^4\right] &\leq c_4 \mathbb{E}\left[\left|\int_0^T \left(1_{X_s>a} - 1_{X_s>b}\right) d\langle X\rangle_s\right|^2\right] \\ &= c_4 \mathbb{E}\left[\left|\int_0^T 1_{a<X_s\leq b}\, d\langle X\rangle_s\right|^2\right]. \end{aligned}$$

We now use the occupation formula, Proposition 4.19, Cauchy–Schwarz and Fubini

$$\begin{aligned} \mathbb{E}\left[\left|\int_0^T 1_{a<X_s\leq b}\, d\langle X\rangle_s\right|^2\right] &= \mathbb{E}\left[\left(\int_a^b L_T^x dx\right)^2\right] \\ &\leq (b-a)\mathbb{E}\int_a^b \left(L_T^x\right)^2 dx \\ &= (b-a)\int_a^b \mathbb{E}\left[\left|L_T^x\right|^2\right] dx \\ &\leq (b-a)^2 \sup_{x\in[a,b]} \mathbb{E}\left[\left|L_T^x\right|^2\right]. \end{aligned}$$

To show that $\sup_{x\in[a,b]} \mathbb{E}\left[\left|L_T^x\right|^2\right] < \infty$, recall that by definition of the local time L

$$L_T^x = |X_T - x| - |X_0 - x| - \int_0^T \operatorname{sgn}(X_s)\, dX_s.$$

By the reverse triangle, $||X_T - x| - |X_0 - x|| \leq |X_T - X_0|$. Together with $(a+b)^2 \leq 2\left(a^2+b^2\right)$, the above gives

$$\left|L_T^x\right|^2 \quad \leq \quad 2\left(|X_T - X_0|^2 + \left(\int_0^T \operatorname{sgn}(X_s)\, dX_s\right)^2\right).$$

Taking expectation, using the Ito-isometry and the boundedness assumption on X and $\langle X\rangle$ gives

$$\begin{aligned}\mathbb{E}\left[\left|L_T^x\right|^2\right] &\leq 2\left(\mathbb{E}\left[|X_T - X_0|^2\right] + \mathbb{E}\left[\left|\int_0^T \operatorname{sgn}(X_s)\, dX_s\right|^2\right]\right) \\ &\leq 2\left((2n)^2 + 2\mathbb{E}\left[\langle X\rangle_t^2\right]\right) \leq 8\left(n^2+n\right).\end{aligned}$$

□

EXAMPLE 4.21. Above is not true for all continuous semimartingales, for example $X = |B_t|$; in this case

$$L^a(X) = \begin{cases} L^a(B) + L^{-a}(B) & \text{if } a > 0, \\ 0 & \text{else.}\end{cases}$$

Since $L_t^0(B) \neq 0$ for fixed $t > 0$, the process $a \mapsto L_t^a$ must have a discontinuity at $a = 0$. However, for $X \in \mathcal{S}_c$ one can still show that a modification exists such that $(t,a) \mapsto L_t^a$ is continuous in t and cadlag in a (using a slight variation of above proof).

COROLLARY 4.22. *Let $M \in \mathcal{M}_{c,loc}$ and $L = \left(L_t^a\right)_{t\geq 0, a\in\mathbb{R}}$ its local time. Then*

$$L_t^a = \lim_{\epsilon\to 0} \frac{1}{2\epsilon}\int_0^t 1_{(a-\epsilon,a+\epsilon)}(M_s)\, d\langle M\rangle_s .$$

Due the importance of Brownian motion and its appearance as noise in the SDEs we now have closer look at Brownian local time. First note that for BM and $a = 0$ above reads

$$L_t^0(B) = \lim_{\epsilon\to 0}\frac{1}{2\epsilon}\int_0^t 1_{|B_s|<\epsilon}\, ds, \tag{4.3.5}$$

thus the Brownian local time is adapted to the filtration generated by $|B|$. The following observation becomes very useful

LEMMA 4.23 (SKOROKHOD'S LEMMA). *Let $f : [0,\infty) \to \mathbb{R}$ be continuous and $f(0) \geq 0$. Then there exists a unique pair $g, a : [0,\infty) \to \mathbb{R}$ such that*

(1) *$f = g - a$,*
(2) *$g(x) \geq 0$ for all x,*
(3) *a is non-decreasing, continuous and $a(0) = 0$. The measure da has its support in $\{t : g = 0\}$, that is $\int 1_{g(x)>0}\, da(x) = 0$.*

Moreover, a is uniquely determined as

$$a(t) = \sup_{s\leq t}(-f(s)) \vee 0.$$

PROOF. Existence follows by taking the pair $a(t) := \sup_{s\leq t} f(x)^-$ and $g := f + a$. To see uniqueness let $\overline{g}, \overline{a}$ be another pair and note that the function $g - \overline{g} = a - \overline{a}$ is of bounded variation. Using first integration by parts, and then point 3 and (2) shows

$$0 \leq (g(t) - \overline{g}(t))^2 = 2\int_0^t (g(s) - \overline{g}(s))\, d(a(s) - \overline{a}(s)) = -2\int_0^t g(s)\, d\overline{a}(s) - 2\int_0^t \overline{g}(s)\, da(s) \leq 0.$$

□

Corollary 4.24. *Let B be a Brownian motion. The process $X_t := \int_0^t \operatorname{sgn}(B_s)\,dB_s$ is a Brownian motion and the filtration generated by X is the same as the filtration generated by $|B|$. Moreover,*

$$L_t^0(B) = -\sup_{s\leq t} X_s \text{ for all } t \geq 0, a.s. \tag{4.3.6}$$

Proof. By Levy's characterization X is a Brownian motion, and by Tanaka's formula $|B_t| = X_t + L_t^0(B)$. Hence with $f(t) := X_t(\omega)$, the pair $g(t) = |B_t(\omega)|$, $a(t) = L_t^0(B)$ fulfills $f = g - a$. But by Lemma 4.23 $a(t) = \sup_{s\leq t}(-X_t(\omega) \vee 0) = -\sup_{s\leq t} X_s(\omega)$.

If we denote by $\mathcal{F}_t^{|B|}$,$\mathcal{F}_t^X$,$\mathcal{F}_t^{L^0(B)}$ the filtrations generated by $|B|$,X and $L^0(B)$ on $[0,t]$, then by Tanaka's formula and 4.3.6 $\mathcal{F}_t^{|B|} \subset \mathcal{F}_t^X$. However, by 4.3.5 $\mathcal{F}_t^{L^0(B)} \subset \mathcal{F}_t^{|B|}$ and therefore also $\mathcal{F}_t^X \subset \mathcal{F}_t^{|B|}$. □

We can now make Tanaka's SDE rigorous.

Corollary 4.25. *Set $\sigma(x) = \operatorname{sgn}(x)$. The SDE $\mathrm{e}_x(0,\sigma)$ has*

(1) *a weak solution but no strong solution,*
(2) *uniquness in law holds but not pathwise uniqueness*

Proof. Exercise sheet. □

Corollary 4.26. *Let B be a Brownian motion and $S_t := \sup_{s\leq t} B_s$. The two-dimensional processes*

$$\left(|B|, L^0(B)\right) \text{ and } (S - B, S)$$

have the same distribution.

Proof. Define $\overline{B}_t := -\int_0^t \operatorname{sgn}(B_s)\,dB_s$ and $\overline{S}_t := \sup_{s\leq t} \overline{B}_s$. By above Lemma 4.3.6, $L^0(B) = \overline{S}$ and using this in Tanaka's formula gives

$$|B| = L^0(B) - \overline{B} = \overline{S} - \overline{B}.$$

But by Levy's characterization $\overline{B}$ is a Brownian motion, hence (B, S) and $\left(\overline{B}, \overline{S}\right)$ have the same distribution and the result follows. □

4.4. One-dimensional SDEs: pathwise uniqueness via local times

We now use local time to revisit the question of uniqueness of solutions. The Yameda–Watanabe Theorem shows that pathwise uniqueness implies uniqueness in law. We now show that sometimes the reverse implication holds. We focus on the one-dimensional case which allows to use the local time and generalize from Lipschitz to Holder regular coefficients. The key result is that by using local time one can show under certain conditions the inverse of the Yameda–Watanabe theorem, Theorem **??**, that is we will see conditions such that uniqueness in law implies pathwise uniqueness (recall that by Yameda–Watanbe a solution must be a strong solution if pathwise uniqueness holds).

Proposition 4.27. *Assume uniqueness in law holds for $\mathrm{e}_x(\sigma,\mu)$. If for any two solutions (X^1, B), (X^2, B) of $\mathrm{e}_x(\sigma,\mu)$ on the same filtered probability space we have*

$$L^0\left(X^1 - X^2\right) = 0$$

then pathwise uniqueness holds for $\mathrm{e}_x(\sigma,\mu)$.

Proof. By Lemma 4.28, X^1 and $X^1 \vee X^2$ are solutions, thus have the same distribution. Especially for every t, $X_t^1 \geq X_t^2$ a.s. Same argument with X^2 and $X^1 \vee X^2$ shows $X_t^2 \geq X_t^1$ a.s., thus by continuity of trajectories $X^1 = X^2$ a.s. □

Above proposition follows from the key lemma below: the supremum of two ODE solutions is again an ODE solution but his is not true for SDE solutions. However, local time gives a sufficient and necessary criteria.

LEMMA 4.28. *Let $(X^1,B),(X^2,B)$ be two solutions of* $\mathrm{e}_x(\sigma,\mu)$ *on the same filtered probability space. Then $(X^1\vee X^2,B)$ is a solution of* $\mathrm{e}_x(\sigma,\mu)$ *if and only if* $L^0(X^2-X^1)=0$.

PROOF. By Tankas formula applied to $(X^2-X^1)^+$ we have

$$X_t^1\vee X_t^2=X_t^1+\left(X_t^2-X_t^1\right)^+=X_t^1+\int_0^t 1_{X_s^2>X_s^1}d\left(X^2-X^1\right)_s+\frac{1}{2}L_t^0\left(X^2-X^1\right).$$

Using that $dX_t^i=\mu\left(t,X_t^i\right)dt+\sigma\left(t,X_t^i\right)dB_t$ for $i=1,2$ above becomes

$$\begin{aligned}
X_t^1\vee X_t^2 &= \int_0^t\mu\left(t,X_t^1\right)dt+\int_0^t\sigma\left(t,X_t^1\right)dB_t\\
&\quad+\int_0^t\left(\mu\left(t,X_t^2\right)-\mu\left(t,X_t^1\right)\right)1_{X_s^2>X_s^1}dt+\int_0^t\left(\sigma\left(t,X_t^2\right)-\sigma\left(t,X_t^1\right)\right)1_{X_s^2>X_s^1}dB_t\\
&\quad+\frac{1}{2}L_t^0\left(X^1-X^2\right)\\
&= \int_0^t\mu\left(t,X_t^1\vee X_t^2\right)dt+\int_0^t\sigma\left(t,X_t^1\vee X_t^2\right)dB_t+\frac{1}{2}L_t^0\left(X^1-X^2\right)
\end{aligned}$$

Hence $X^1\vee X^2$ is a solution of $\mathrm{e}(\sigma,\mu)$ iff $L_t^0\left(X^1-X^2\right)=0$. □

We now need to find criteria that imply $L^0\left(X^1-X^2\right)=0$. A useful Lemma for this is

LEMMA 4.29. *Let $\rho:(0,\infty)\to(0,\infty)$ be Borel measurable and such that*

$$\int_{0+}^{\epsilon}\frac{1}{\rho(a)}da=\infty \tag{4.4.1}$$

for an $\epsilon>0$. If $X\in\mathcal{S}_c$ such that

$$\int_0^t 1_{0<X_s\le\epsilon}\frac{1}{\rho(X_s)}d\langle X\rangle_s<\infty \text{ for all } t\ge 0 \tag{4.4.2}$$

then $L^0(X)=0$.

PROOF. By the Occupation formula, Proposition 4.19 above integral equals

$$\int_0^{\epsilon}L_t^a(X)\frac{1}{\rho(a)}da.$$

As ϵ goes to 0, a approaches 0 and L_t^a converges to L_t^0. But by 4.4.1 this implies that $L^0(X)=0$, otherwise (4.4.2) is ∞ with positive probability. □

Putting all the above together we now have a very general criteria for pathwise uniquness in one dimension.

THEOREM 4.30 (LE GALL). *Let $\mu,\sigma:\mathbb{R}\to\mathbb{R}$ be Borel measurable,*

(1) *μ Lipschitz,*
(2) *there exists a $\rho:(0,\infty)\to(0,\infty)$ be Borel measurable, $\int_{0+}^{\epsilon}\frac{1}{\rho(a)}da=\infty$ $\forall\epsilon>0$ such that,*

$$|\sigma(x)-\sigma(y)|^2\le\rho(|x-y|)\ \ \forall x,y\in\mathbb{R}. \tag{4.4.3}$$

Then pathwise uniqueness holds for $\mathrm{e}_x(\mu,\sigma)$.

PROOF. Let (X^1,B) and (X^2,B) be two solutions. Set $Y:=X^1-X^2$ and use (4.4.3) to estimate

$$\begin{aligned}
\int_0^t\frac{1}{\rho(Y_s)}1_{0<Y\le\epsilon}d\langle Y\rangle_s &= \int_0^t\frac{1}{\rho\left(X_s^1-X_s^2\right)}\left(\sigma\left(X^1\right)-\sigma\left(X^2\right)\right)^2 1_{0<Y_s<\epsilon}ds\\
&\le t<\infty
\end{aligned}$$

Hence, Lemma 4.29 implies that $L^0\left(X^1-X^2\right)=0$ and so by Tanaka's formula

$$\left|X_t^1-X_t^2\right|=\int_0^t \operatorname{sgn}\left(X_s^1-X_s^2\right)\left(\sigma\left(X_s^1\right)-\sigma\left(X_s^2\right)\right)dB_s+\int_0^t \operatorname{sgn}\left(X_s^1-X_s^2\right)\left(\mu\left(X_s^1\right)-\mu\left(X_s^2\right)\right)ds.$$

We can assume wlog that σ is bounded (by localization) and therefore assume that $\int_0^t \operatorname{sgn}\left(X_s^1-X_s^2\right)\left(\sigma\left(X_s^1\right)-\sigma\left(X_s^2\right)\right)dB_s$ is a continuous martingale. Thus taking expectations and using the Lipschitzness of μ gives

$$\begin{aligned}\mathbb{E}\left[\left|X_t^1-X_t^2\right|\right] &\leq \mathbb{E}\left[\int_0^t\left|\mu\left(X_s^1\right)-\mu\left(X_s^2\right)\right|ds\right]\\ &\leq c\int_0^t \mathbb{E}\left[\left|X_s^1-X_s^2\right|\right]ds.\end{aligned}$$

Using Gronwall's lemma,$\mathbb{E}\left[\left|X_t^1-X_t^2\right|\right]=0$ and we conlude $X_t^1=X_t^2$ a.s., and by continuity $X^1=X^2$ a.s. □

There are many other statements of above type with different conditions on the coefficients. We state the following theorem without proof.

THEOREM 4.31 (NAKAO). *Let $\mu,\sigma:\mathbb{R}\to\mathbb{R}$ be Borel measurable and assume that*

(1) *μ is bounded,*
(2) *there exists an increasing, Borel measurable function $f:\mathbb{R}\to\mathbb{R}$ and a $\epsilon>0$ such that for all $x,y\in\mathbb{R}$*

$$|\sigma(x)-\sigma(y)|^2\leq|f(x)-f(y)|,\sigma(x)>\epsilon.$$

Then pathwise uniqueness holds for SDE $\mathrm{e}_x(\mu,\sigma)$.

EXAMPLE 4.32. Consider $\mathrm{e}_x(\mu,\sigma)$ with $\mu(x)=\sqrt{|x|}$ and $\sigma(x)=\epsilon$ and fix $T>0$. By Proposition **??** it has a weak solution on $[0,T]$, by above Theorem 4.31 this solution is pathwise unique and by Yameda–Watanabe, Theorem **??** we conclude that this solution the pathwise unique strong solution. Note that $\sigma(x)>0$ is essential since the ODE $dy_t=\mu(y_t)dt$ has infinitely many solutions. This is an example of the regularising effect of the quadratic variation of Brownian motion.

Using local times, we can compare SDEs driven by the same noise but different drift coefficients.

THEOREM 4.33 (SDE COMPARISON). *Let $\mu_1,\mu_2,\sigma:\mathbb{R}\to\mathbb{R}$ be measurable with*

(1) *σ as in Theorem 4.30,*
(2) *$\mu_1(x)\geq\mu_2(x)$ for all $x\in\mathbb{R}$ and at least one of μ_1,μ_2 Lipschitz.*

If $\left(X^1,B\right)$ is a solution of $e_{x^1}(\mu,\sigma)$ and $\left(X^2,B\right)$ a solution of $e_{x^2}(\mu,\sigma)$ on the same filtered probabilty space and $x^1\geq x^2$ then

$$X_t^2\geq X_t^1 \text{ for all } t\geq0$$

holds a.s.

PROOF. As in Theorem 4.30 we use the assumption on σ to show $L\left(X^2-X^1\right)=0$. Therefore

$$\begin{aligned}\left(X^2-X^1\right)^+ &= \int_0^t 1_{X_s^2>X_s^1}d\left(X^2-X^1\right)_s\\ &= \int_0^t 1_{X_s^2>X_s^1}\left(\mu_2\left(X_s^2\right)-\mu_1\left(X_s^1\right)\right)ds\\ &\quad+\int_0^t 1_{X_s^2>X_s^1}\left(\sigma\left(X_s^2\right)-\sigma\left(X_s^1\right)\right)dB_s.\end{aligned}$$

As in Theorem 4.30, wlog σ is bounded (by localization) so taking expectation in above shows

$$\begin{aligned}\phi(t) &:= \mathbb{E}\left[\left(X_t^2 - X_t^1\right)^+\right] \\ &= \mathbb{E}\left[\int_0^t 1_{X_s^2 > X_s^1}\left(\mu_2\left(X_s^2\right) - \mu_1\left(X_s^1\right)\right) ds\right].\end{aligned}$$

Now if μ_1 is Lipschitz, we estimate

$$\begin{aligned}\mathbb{E}\left[\int_0^t 1_{X_s^2 > X_s^1}\left(\mu_2\left(X_s^2\right) - \mu_1\left(X_s^1\right)\right) ds\right] &\leq \mathbb{E}\left[\int_0^t 1_{X_s^2 > X_s^1}\left(\mu_1\left(X_s^2\right) - \mu_1\left(X_s^1\right)\right) ds\right] \\ &\leq c\mathbb{E}\left[\int_0^t \left|X_s^2 - X_s^1\right| 1_{X_s^2 > X_s^1}\right] ds \leq c\phi(s)\end{aligned}$$

and the result follows by Gronwall's lemma. If μ_2 is Lipschitz use that $(a-b) \leq |a-c| + c - b$ to estimate

$$\begin{aligned}\mathbb{E}\left[\int_0^t 1_{X_s^2 > X_s^1}\left(\mu_2\left(X_s^2\right) - \mu_1\left(X_s^1\right)\right) ds\right] &\leq \mathbb{E}\left[\int_0^t 1_{X_s^2 > X_s^1}\left|\mu_2\left(X_s^2\right) - \mu_2\left(X_s^1\right)\right| ds\right] \\ &\quad + \mathbb{E}\left[\int_0^t 1_{X_s^2 > X_s^1}\left(\mu_2\left(X_s^1\right) - \mu_1\left(X_s^1\right)\right) ds\right] \\ &\leq \mathbb{E}\left[\int_0^t 1_{X_s^2 > X_s^1}\left|\mu_2\left(X_s^2\right) - \mu_2\left(X_s^1\right)\right| ds\right]\end{aligned}$$

where the last estimate follows since $\mu_1 \geq \mu_2$. The result now follows from the Lipschitzness of μ_2 and Gronwall's lemma. □

APPENDIX A

Reading list

Below I give some pointers to the literature for topics covered in the course. It should NOT be assumed that the list below constitutes either necessary or sufficient material for the exam. However, you might find the references useful for exam preparation and revision of material

Your first reference for background material should be your notes and the lectures notes for B8.1 and B8.2. If you get one additional book, then either [**Revuz and Yor, 1999**] or [**Karatzas and Shreve, 1991**] are both excellent choices and cover all the material of this course and much more in concise way; if you prefer a somewhat slower pace, [**Durrett, 1996**] gives great motivation. A classic reference that I highly recommend is [**Rogers and Williams, 2000a, Rogers and Williams, 2000b**]. Yet another addition to all the material is the blog by George Lowther `https://almostsure.wordpress.com`.

(1) **Introduction**
   (a) Motivation and big picture: [**Gardiner, 2009, Gardiner et al., 1985**] applications in natural sciences; [**Øksendal, 2003**] applications in finance, signal filtering, PDEs
   (b) Recall probability: [**Malliavin, 1995**] concise and elegant
   (c) Recall stochastic processes, filtrations, stopping times: [**Karatzas and Shreve, 1991**, Section 1.1 - 1.2] and [**Revuz and Yor, 1999**, Section 1.1-1.4]

(2) **Martingale theory in continuous time**
   Main references are the lecture notes for B8.2 [**Obloj, 2015**] and [**Revuz and Yor, 1999**].
   (a) Brownian motion & applications: [**Rogers and Williams, 2000a**, I.1-I.16] short overview of BM (highly recommended); [**Mörters and Peres, 2010**] readable and modern account of BM
   (b) Martingale inequalites, regularity, optional stopping: [**Revuz and Yor, 1999**, Chapter 2]
   (c) Stochastic integration**:** [**Durrett, 1996**] and [**Rogers and Williams, 2000b**, Chapter IV, Section 4] intuitive, step-by-step; [**Revuz and Yor, 1999**, Chapter 4] and [**Karatzas and Shreve, 1991**, Chapter 3] concise presentation of standard L^2 martingale theory; [**Protter, 2005**, Chapter 2] presents alternative approach: semimartingales defined as "good integrators", forces to prove the Bichteler–Delacherie–Meyer theorem that shows equivalence with the standard definition of semimartingales; [**McKean, 2005**] densely written but highly recommended as a second reference to any of the above.

(3) **Ito's stochastic calculus and applications**
   (a) Ito's formula: any book about stochastic calculus.
   (b) Stochastic exponential, Levy's characterization, and Dambis–Dubins–Schwarz: these were already treated in [**Obloj, 2015**]; we follow closely [**Revuz and Yor, 1999**]. There is much more to say about the stochastic exponential, for example applications to complex valued martingales and conformal mappings. Similarly, time changes of can be studied more generally see [**Revuz and Yor, 1999**]
   (c) Change of measure: for basic measure theory background [**Malliavin, 1995**, Chapter 6]; the blog `https://almostsure.wordpress.com/2010/05/03/girsanov-transformations/` has a very nice presentation which we follow partly; the assumption of $\mathcal{E}(M)$ being u.i. resp. $\mathbb{P}$ and $\mathbb{Q}$ being equivalent is often too strong and finer estimates are possible see

[**Karatzas and Shreve, 1991**, Chapter 3.5]; [**Revuz and Yor, 1999**] presents a very general theory of Girsanov transformations. For Gaussian quasi-invariance see [**Malliavin, 1997**].

(4) **Stochastic differential equations**
   (a) The standard reference for SDE is [**Ikeda and Watanabe, 1989**].

Bibliography

[Durrett, 1996] Durrett, R. (1996). *Stochastic calculus: a practical introduction*, volume 6. CRC press.

[Gardiner, 2009] Gardiner, C. (2009). *Stochastic methods*. Springer Berlin.

[Gardiner et al., 1985] Gardiner, C. W. et al. (1985). *Handbook of stochastic methods*, volume 4. Springer Berlin.

[Hida, 1980] Hida, T. (1980). *Brownian motion*. Springer.

[Ikeda and Watanabe, 1989] Ikeda, N. and Watanabe, S. (1989). *Stochastic differential equations and diffusion processes*. North-Holland Publishing Co., Amsterdam, second edition.

[Karatzas and Shreve, 1991] Karatzas, I. and Shreve, S. E. (1991). *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition.

[Malliavin, 1995] Malliavin, P. (1995). *Integration and probability*, volume 157. Springer Science & Business Media.

[Malliavin, 1997] Malliavin, P. (1997). *Stochastic analysis*, volume 313 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin.

[McKean, 2005] McKean, H. P. (2005). *Stochastic integrals*. AMS Chelsea Publishing, Providence, RI. Reprint of the 1969 edition, with errata.

[Mörters and Peres, 2010] Mörters, P. and Peres, Y. (2010). *Brownian motion*, volume 30. Cambridge University Press.

[Obloj, 2015] Obloj, J. (2015). *Continuous martingales and stochastic calculus*. Lecture notes B8.2.

[Øksendal, 2003] Øksendal, B. (2003). *Stochastic differential equations*. Springer.

[Protter, 2004] Protter, P. E. (2004). *Stochastic integration and differential equations*, volume 21 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, second edition. Stochastic Modelling and Applied Probability.

[Protter, 2005] Protter, P. E. (2005). *Stochastic integration and differential equations*, volume 21 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin. Second edition. Version 2.1, Corrected third printing.

[Revuz and Yor, 1999] Revuz, D. and Yor, M. (1999). *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition.

[Rogers and Williams, 2000a] Rogers, L. C. G. and Williams, D. (2000a). *Diffusions, Markov processes, and martingales. Vol. 1*. Cambridge Mathematical Library. Cambridge University Press, Cambridge. Foundations, Reprint of the second (1994) edition.

[Rogers and Williams, 2000b] Rogers, L. C. G. and Williams, D. (2000b). *Diffusions, Markov processes, and martingales. Vol. 2*. Cambridge Mathematical Library. Cambridge University Press, Cambridge. Ito calculus, Reprint of the second (1994) edition.

Index