Stochastic Differential Equations

Lecture Notes
SummerTerm 2012

by Markus Kunze

Preface

The present manuscript contains the notes for a lecture given at the University of Ulm in the summer term 2012. It was my goal to give an overview of existence and uniqueness results for stochastic differential equations. Moreover, I wanted to give a presentation of the results which is more or less self-contained, thus I wanted to avoid merely quoting results, even if the results are somewhat technical.

As prerequisites, I assumed basic knowledge from measure theory, probability theory and functional analysis, as well as some familarity with ordinary differential equations. Some more advanced results are recalled and (with the exception of Prokhorov's theorem) also proved in the Appendices.

In the preparation of this manuscript I used the following monographs which I also recommend for further reading:

The book by Kallenberg [**4**] gives an overview of all of probability and is a source of concise and elegant proofs. The books of Karatzas and Shreve [**5**], Øksendal [**6**] and Revuz and Yor [**7**] are standard intruductions to the topic, with the book by Øksendal maybe being the most "student-friendly". The books by Stroock and Varadhan [**8**] and Ethier Kurtz [**2**] are more focussed around the martingale problem, with [**8**] more focussed on diffusion processes (and thus partial differential equations) whereas [**2**] also treats more general Markov processes (which not necessarily continuous paths).

The current manuscript is a preliminary version. It might be changed during the semester. If you find any mistakes, please let me know.

M.K.

Contents

Preface iii

Chapter 1. A First Glance at Stochastic Integration 1
1.1. Brownian Motion 1
1.2. The Wiener Integral 2
1.3. Itô's Integral 5
1.4. Exercises 8

Chapter 2. Continuous Local Martingales 11
2.1. Martingales: Basic results 11
2.2. Quadratic Variation 15
2.3. Covariation 19
2.4. Exercises 21

Chapter 3. Stochastic Calculus 23
3.1. The Itô Integral 23
3.2. Itô's Formula 28
3.3. First applications of Itô's Formula 30
3.4. Exercises 33

Chapter 4. Stochastic Differential Equations with Locally Lipschitz Coefficients 37
4.1. Solutions via Banach's Fixed Point Theorem 37
4.2. Extension to locally Lipschitz Coefficients 39
4.3. Examples 41
4.4. Exercises 44

Chapter 5. Yamada-Watanabe Theory 47
5.1. Different notions of existence and uniqueness 47
5.2. On strong and weak uniqueness 49
5.3. Pathwise uniqueness for some one-dimensional equations 52
5.4. Exercises 54

Chapter 6. Martingale Problems 57
6.1. The Martingale Problem associated to an SDE 57
6.2. Existence of weak solutions 63
6.3. Uniqueness of solutions 66
6.4. Exercises 68

Appendix A. Continuity of Paths for Brownian Motion 69

Appendix B. Stochastic Processes as Random Elements 73

Appendix C. Stieltjes Integrals 77

Appendix D. Measures on Topolgical Spaces 81

Bibliography 83

CHAPTER 1

A First Glance at Stochastic Integration

1.1. Brownian Motion

DEFINITION 1.1.1. Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space, $I \subset \mathbb{R}$. A *stochastic process* is a family $(X(t))_{t\in I}$ of random variables $X(t) : \Omega \to \mathbb{R}$.

DEFINITION 1.1.2. Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space. A *Brownian motion* or *Wiener process* is a stochastic process $(W(t))_{t\geq 0}$ such that

(1) $W(0) = 0$ almost surely.
(2) $W(t+s) - W(t)$ is independent of $\sigma(W(r) : 0 \leq r \leq t)$ for all $t, s \geq 0$.
(3) $W(t+s) - W(t)$ has distribution $\mathscr{N}(0, s)$, Gaussian distribution with mean 0 and variance s.

Given a *filtration* $\mathbb{F} = (\mathscr{F}_t)_{t\geq 0}$, i.e. a family of sub σ-algebras $\mathscr{F}_t$ with $\mathscr{F}_s \subset \mathscr{F}_t$ for $t \geq s$, we will say that a process $(W(t))_{t\geq 0}$ is an $\mathbb{F}$-Brownian motion, if it is *adapted*, i.e. $W(t)$ is $\mathscr{F}_t$ measurable and (1), (2') and (3) hold, where

(2') $W(t) - W(s)$ is independent of $\mathscr{F}_s$ for all $0 \leq t < s$.

REMARK 1.1.3. A Brownian motion $(W(t))_{t\geq 0}$ is an $\mathbb{F}^W$-Brownian motion where $\mathbb{F}^W$ is the σ-algebra generated by $(W(t))_{t\geq 0}$, i.e. $\mathscr{F}_t = \sigma(W(s) : s \leq t)$.

To construct Brownian motion, we make use of so-called *isonormal Gaussian processes.*

DEFINITION 1.1.4. Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space, H be a Hilbert space. An *H-isonormal Gaussian process* is a map $\mathscr{W} : H \to L^2(\Omega, \Sigma, \mathbb{P})$ such that

(1) $\mathscr{W}(h)$ is a (centered) Gaussian random variable for all $h \in H$, i.e. for some $q \in [0, \infty)$ we have $\varphi_{\mathscr{W}(h)}(t) := \mathbb{E}e^{it\mathscr{W}(h)} = e^{-\frac{q}{2}t^2}$. Thus, $\mathscr{W}(h)$ is either constantly zero $(q = 0)$ or has distribution $\mathscr{N}(0, q)$ $(q > 0)$.
(2) $(h_1 \,|\, h_2)_H = (\mathscr{W}(h_1) \,|\, \mathscr{W}(h_2))_{L^2(\Omega)}$ for all $h_1, h_2 \in H$.

REMARK 1.1.5. If X is centered Gaussian with $\varphi_X = e^{-\frac{q}{2}t^2}$, then

$$\varphi_X'(t) = -qte^{-\frac{q}{2}t^2} \qquad \varphi_X''(t) = (-q + q^2t^2)e^{-\frac{q}{2}t^2}$$

$$\varphi_X^{(3)}(t) = (3q^2t - q^3t^3)e^{-\frac{q}{2}t^2} \qquad \varphi_X^{(4)}(t) = (3q^2 - 6q^3t^2 + q^4t^4)e^{-\frac{q}{2}t^2}$$

Hence, for the first and third moment of X we have $\mathbb{E}X = i\varphi_X'(0) = 0 = i^3\varphi_X^{(3)}(0) = \mathbb{E}X^3$, for the second moment we obtain $\mathbb{E}X^2 = i^2\varphi_X''(0) = q$ and for the forth moment $\mathbb{E}X^4 = i^4\varphi_X^{(4)}(0) = 3q^2$.

LEMMA 1.1.6. *Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space, H be a Hilbert space and $\mathscr{W} : H \to L^2(\Omega, \Sigma, \mathbb{P})$ be an H-isonormal Gaussian process. Then $\mathscr{W}$ is linear.*

PROOF. For $\alpha, \beta \in \mathbb{R}$ and $h_1, h_2 \in H$ we have

$$\|\mathscr{W}(\alpha h_1 + \beta h_2) - \alpha\mathscr{W}(h_1) + \beta\mathscr{W}(h_2)\|_{L^2(\Omega)}^2 = 0$$

which proves that $\mathscr{W}$ is linear. $\square$

PROPOSITION 1.1.7. *Let H be a separable Hilbert space, $(\Omega, \Sigma, \mathbb{P})$ be a probability space on which a sequence of independent standard Gaussian random variables $(\gamma_n)_{n\in\mathbb{N}}$ is defined (e.g. $(\prod_{k\in\mathbb{N}}\mathbb{R}, \bigotimes_{k\in\mathbb{N}}\mathscr{B}(\mathbb{R}), \bigotimes_{n\in\mathbb{N}}\gamma)$ with $\gamma_k((x_n)_{n\in\mathbb{N}}) = x_k$).*

Then there exists an H-isonormal Gaussian process $\mathscr{W} : H \to L^2(\Omega)$.

PROOF. Let $(e_n)_{n\in\mathbb{N}}$ be an orthonormal basis of H and define

$$\mathscr{W}(h) := \sum_{n\in\mathbb{N}} \gamma_n (h \,|\, e_n)\,.$$

Then $\mathscr{W}$ is an H-isonormal Gaussian process.

Indeed, by independence, $\sum_{k=1}^N \gamma_k (h\,|\,e_k)$ is a Gaussian random variable with variance $\sum_{k=1}^N |(h\,|\,e_k)|^2$ for every $N \in \mathbb{N}$. Upon $N\to\infty$, we see that $\mathscr{W}(h)$ is Gaussian with variance $\sum_{k=1}^\infty |(h\,|\,e_k)|^2 = \|h\|_H^2$. ☐

THEOREM 1.1.8. *There exists a Brownian motion.*

PROOF. Let $\mathscr{W}$ be an $L^2([0,\infty))$-isonormal Gaussian process and put $W(t) := \mathscr{W}(\mathbb{1}_{[0,t)})$.

Then $\|W(0)\|_{L^2(\Omega)} = \|0\|_{L^2([0,\infty))} = 0$, hence $W(0) = 0$ almost surely. Given $0 \le t_1 < t_2 < \cdots t_n = t < t+s$, observe that the vectors $\mathbb{1}_{[t_1,t_2)}, \ldots, \mathbb{1}_{[t_{n-1},t_n)}, \mathbb{1}_{[t,t+s)}$ are orthogonal in $L^2([0,\infty))$. Since $\mathscr{W}$ is isometric, $W(t_2) - W(t_1) = \mathscr{W}(\mathbb{1}_{[t_1,t_2)}), \ldots W(t_n) - W(t_{n-1}) = \mathscr{W}(\mathbb{1}_{[t_{n-1},t_n)}), W(t+s) - W(t) = \mathscr{W}(\mathbb{1}_{[t,t+s)})$ are orthogonal in $L^2(\Omega)$; since the latter random variables are jointly Gaussian, they are independent. Since $\sigma(W(r) : 0 \le r \le t) = \sigma(W(r) - W(q) : 0 \le q \le r \le t)$, it follows that $W(t+s) - W(s)$ is independent of $\sigma(W(r) : 0 \le r \le t)$. Finally, $W(t+s) - W(t)$ is Gaussian with mean 0 and covariance $\|\mathbb{1}_{[t,t+s)}\|_{L^2([0,\infty))} = s$. ☐

In Appendix A, we prove that every Brownian motion $(W(t))_{t\ge0}$ has a continuous version, i.e. there exists a family $\tilde{W}(t) : \Omega \to \mathbb{R}$ of random variables such that

(1) $t \mapsto \tilde{W}(t,\omega)$ is continuous for all $\omega \in \Omega$.
(2) $\mathbb{P}(W(t) = \tilde{W}(t)) = 1$ for all $t \ge 0$.

Note that $(\tilde{W}(t))_{t\ge0}$ is also a Brownian motion. Indeed, (1) and (3) are clear since $\tilde{W}(0) = W(0)$ and $\tilde{W}(t+s) - \tilde{W}(s) = W(t+s) - W(s)$ almost surely for all $t, s \ge 0$. As for (2), note that if $t_1 < t_2 < \cdots t_n = t < t+s$, then

$$\begin{aligned}&(W(t+s) - W(t), W(t_n) - W(t_{n-1}), \ldots, W(t_2) - W(t_1))\\ &= (\tilde{W}(t+s) - \tilde{W}(t), \tilde{W}(t_n) - \tilde{W}(t_{n-1}), \ldots, \tilde{W}(t_2) - \tilde{W}(t_1))\end{aligned}$$

almost surely. Hence these vectors are identically distributed. Now (2) follows as in the proof of Theorem 1.1.8 From now on, we will always use Brownian motions with continuous paths.

1.2. The Wiener Integral

DEFINITION 1.2.1. Let $\mathscr{W}$ be an $L^2([0,\infty))$-isonormal Gaussian process and $(W(t))_{t\ge0}$ be the Brownian motion constructed from this isonormal process as in the proof of Theorem 1.1.8. For $\phi \in L^2([0,\infty))$, the *Wiener integral* of ϕ is defined as

$$\int_0^\infty \phi(s)\, dW(s) := \mathscr{W}(\phi)\,.$$

For $t > 0$, we define $\int_0^t \phi(s)\,dW(s) := \int_0^\infty \phi(s)\mathbb{1}_{[0,t)}(s)\,dW(s)$.

REMARK 1.2.2. Let $t > 0$ and $0 = t_0 < t_1 < \cdots < t_n = t$ be a partition of $[0,t]$. If $\phi = \sum_{k=1}^n a_k \mathbb{1}_{[t_{k-1},t_k)}$, then

$$\int_0^t \phi(s)\,dW(s) = \mathscr{W}\Big(\sum_{k=1}^n a_k \mathbb{1}_{[t_{k-1},t_k)}\Big) = \sum_{k=1}^n a_k \mathscr{W}(\mathbb{1}_{t_{k-1},t_k)}) = \sum_{k=1}^n a_k (W(t_k) - W(t_{k-1}))\,.$$

Definition 1.2.3. (Stochastic differential equations with additive noise)
Let $f : \mathbb{R} \to \mathbb{R}$ be measurable and bounded on bounded subsets of $\mathbb{R}$ and $\sigma > 0$. A *solution* of the *stochastic differential equation*

$$(SDE)\begin{cases} dX(t) &= f(X(t))dt + \sigma dW(t) \\ X(0) &= x_0 \end{cases}$$

is a stochastic process $(X(t))_{t\geq 0}$ with continuous paths such that for all $t \geq 0$

$$X(t) = x_0 + \int_0^t f(X(s))\, ds + \int_0^t \sigma\, dW(s)$$

almost surely.

Remark 1.2.4. Note that the deterministic integral above is well-defined pathwise. Indeed, X has continuous paths, hence $f(X(\cdot))$ is a bounded, measurable function on $[0, t]$ for all $t > 0$. Also note that $\int_0^t \sigma dW(s) = \sigma W(t)$, hence X is a solution whenever almost surely $X(t) = x_0 + \int_0^t f(X(s))\, ds + \sigma W(t)$ for all $t \geq 0$. We have chosen the above notation to be consistent with more general equations appearing later on.

It is a natural question, how to construct solutions to stochastic differential equations.

Theorem 1.2.5. *Let f be Lipschitz continuous and $\sigma \geq 0$. Then there exists a unique solution of (SDE).*

Proof. Fix $T > 0$. For $\omega \in \Omega$, define $\Phi = \Phi(\omega) : C([0, T]) \to C([0, T])$ by

$$\big[\Phi(u)\big](t) := x_0 + \int_0^t f(u(s))\, ds + \sigma W(t, \omega)\, .$$

Here, as always in what follows, we have assumed that W has continuous paths. A standard application of Banach's fixed point theorem yields that if f is Lipschitz continuous, $\Phi(\omega)$ has a unique fixed point $X^T(\omega) \in C([0, T])$. We put $X^T(t, \omega) := [X^T(\omega)](t)$.

Note that $X^T : \Omega \to C([0, T])$ is measurable. Indeed, as $\Psi : C([0, T]) \to C([0, T])$, defined by $[\Psi u](t) = x_0 + \int_0^t f(u(s))\, ds$ is continuous, it is measurable. Hence with Y also $\Phi(Y) = \Psi(Y) + W$ is measurable. It follows that all iterates $\Phi^n(Y)$, hence also their limit X^T is measurable.

We now put $X(t) := X^T(t)$ for some $T > t$. This is well-defined by uniqueness. Obviously, we have

$$X(t) = x_0 + \int_0^t f(X(s))\, ds + \sigma B(t)$$

for all $t \geq 0$ and $\omega \in \Omega$. □

Example 1.2.6. (Ornstein-Uhlenbeck process)
Let $a \in \mathbb{R}$ and $\sigma > 0$. Then, for every $x_0 \in \mathbb{R}$, there exists a unique solution of the stochastic differential equation

$$(SDE)\begin{cases} dX(t) &= aX(s)dt + \sigma dW(t) \\ X(0) &= x_0 \end{cases}$$

This follows directly from Theorem 1.2.5, noting that $f(t) := at$ is Lipschitz continuous. The solutions of this equation are called *Ornstein-Uhlenbeck processes*.

With the help of the fixed point iteration, we can approximate the solution X:
Let us start with $X_0 := x_0 + \sigma W(t)$. Then

$$X_1(t) = x_0 + \int_0^t ax_0 + a\sigma W(s)\, ds + \sigma W(t) = x_0 + ax_0 t + \int_0^t a\sigma W(s)\, ds + \sigma W(t)\, .$$

With Fubini's theorem, we obtain

$$\begin{aligned} X_2(t) &= x_0 + \int_0^t \Big[a x_0 + a^2 x s + a\sigma W(s) + \int_0^s a^2 \sigma W(r)\,dr \Big]\,ds + \sigma W(t) \\ &= \sum_{k=0}^{2} \frac{a^k t^k}{k!} x_0 + \int_0^t \sum_{k=0}^{1} \sigma \frac{a^{k+1}(t-s)^k}{k!} W(s)\,ds + \sigma W(t)\,. \end{aligned}$$

Inductively,

$$X_n(t) = \sum_{k=0}^{n} \frac{a^k t^k}{k!} x_0 + \int_0^t \sum_{k=0}^{n-1} \sigma \frac{a^{k+1}(t-s)^k}{k!} W(s)\,ds + \sigma W(t)\,.$$

As $n \to \infty$, this converges to

$$X(t) = e^{at} x_0 + \int_0^t a\sigma e^{a(t-s)} W(s)\,ds + \sigma W(t)\,.$$

Let us compare the situation with *ordinary* differential equations. We consider the ODE $u'(t) = au(t) + f(t)$. Writing $u'(t) = \frac{du(t)}{dt}$ and multiplying with dt, we could equivalently have written

$$du(t) = au(t)dt + f(t)dt.$$

Then the solution of the inhomogeneous problem is obtained from the solution of the homogeneous problem via the *variation of constants formula*

$$u(t) = e^{at} x_0 + \int_0^t e^{a(t-s)} f(s)\,ds\,.$$

Using $f(t) = \sigma \frac{dW(t)}{dt}$ as inhomogenity and boldly using integration by parts, we indeed obtain

$$X(t) = e^{at} X_0 + \int_0^t \sigma e^{a(t-s)} W'(s)\,ds = e^{at} X_0 + \int_0^t \sigma a e^{a(t-s)} W(s)\,ds + \sigma W(t)\,.$$

We may wonder, whether we have a corresponding formula involving the Wiener integral, i.e.

$$X(t) = e^{at} x_0 + \int_0^t \sigma e^{a(t-s)}\,dW(s)$$

This is indeed the case and follows from

LEMMA 1.2.7. *(Integration by parts)*
For $\phi \in C^1([0,T])$, we have, almost surely,

$$\int_0^T \phi'(s) W(s)\,ds = \phi(T)W(T) - \int_0^T \phi(s)\,dW(s)$$

PROOF. Let $\pi_n := (0 = t_0^{(n)} < t_1^{(n)} < \cdots < t_{k_n}^{(n)} = T)$ be a sequence of partitions of $[0,T]$ such that the mesh size $|\pi_n| := \max\{|t_j^{(n)} - t_{j-1}^{(n)}| : 1 \le j \le k_n\} \to 0$. We define

$$\phi_n(t) := \sum_{j=1}^{k_n} \phi(t_{j-1}^{(n)}) \mathbb{1}_{[t_{j-1}^{(n)}, t_j^{(n)})}\,.$$

Then $\phi_n \to \phi$ in $L^2((0,T))$ and hence, by continuity of $\mathscr{W}$,

$$\int_0^T \phi_n(s)\,dW(s) \to \int_0^T \phi(s)\,dW(s) \quad \text{in } L^2(\Omega)\,. \tag{1.1}$$

On the other hand, by Abel partial summation, noting that $W(0) = 0$, we obtain

$$\int_0^T \phi_n(s)\,dW(s) \;=\; \sum_{j=1}^{k_n} \phi(t_{j-1}^{(n)}) \big(W(t_j^{(n)}) - W(t_{j-1}^{(n)}) \big)$$

$$\begin{aligned}
&= \phi(T)W(t) + \sum_{j=1}^{k_n} \left[\phi(t_{j-1}^{(n)}) - \phi(t_j^{(n)})\right] W(t_j^{(n)}) \\
&= \phi(T)W(T) + \sum_{j=1}^{k_n} -\phi'(\xi_j^{(n)}) W(t_j^{(n)})(t_j^{(n)} - t_{j-1}^{(n)}) \\
&= \phi(T)W(T) + \sum_{j=1}^{k_n} -\phi'(t_j^{(n)}) W(t_j^{(n)})(t_j^{(n)} - t_{j-1}^{(n)}) \\
&\quad + \int_0^T \rho_n(t,\omega)\, dt.
\end{aligned}$$

Here, we have used the mean value theorem, so $\xi_j^{(n)}$ is a suitable element of $[t_{j-1}^{(n)}, t_j^{(n)}]$; moreover,

$$\rho_n(t,\omega) = \sum_{j=1}^{k_n} \left[\phi'(t_j^{(n)}) - \phi'(\xi_j^{(n)})\right] W(t_j^{(n)}, \omega) \mathbb{1}_{[t_{j-1}^{(n)}, t_j^{(n)})}\,.$$

Noting that $|\rho_n(t,\omega)| \leq 2\|\phi'\|_\infty \|W(\cdot,\omega)\|_\infty$ and that $\rho_n(t,\omega) \to 0$ as $n \to \infty$, which is easy to see using that ϕ' is uniformly continuous on the compact set $[0,T]$ and that $|t_j^{(n)} - \xi_j^{(n)}| \leq |\pi_n| \to 0$ as $n \to \infty$ for all j, it follows from dominated convergence that $\int_0^T \rho_n(t,\omega)\, dt \to 0$ as $n \to \infty$ for (almost) all ω.

Moreover,

$$\sum_{j=1}^{k_n} -\phi'(t_j^{(n)}) W(t_j^{(n)})(t_j^{(n)} - t_{j-1}^{(n)}) \to \int_0^T -\phi'(t)W(t)\, dt$$

almost surely, as the former are Riemannian sums of the latter integral. Altogether, we have

$$\int_0^T \phi_n(s)\, dW(s) \to \Phi(T)W(t) - \int_0^T \phi'(t)W(t)\, dt$$

almost surely. Passing to a subsequence in (1.1) which converges almost surely we see that indeed

$$\int_0^T \phi'(s)W(s)\, ds = \phi(T)W(T) - \int_0^T \phi(s)\, dW(s)$$

almost surely. □

1.3. Itô's Integral

> **Question:** How can we extend the stochastic integral so that we can integrate *random processes* $\phi : [0,\infty) \times \Omega \to \mathbb{R}$ with respect to $(W(t))_{t\geq 0}$?

The naive approach to work pathwise, i.e. to define the integral ω by ω does not work since Brownian motion does not have *finite variation* (as we will see in Lemma 2.2.2). The basic idea behind Itô's integral is to look back to Remark 1.2.2:

When given a stochastic process of the form

$$\phi(t,\omega) = \sum_{k=1}^{n} \eta_k(\omega) \mathbb{1}_{[t_{k-1}, t_k)}(t)$$

where $0 = t_0 < t_1 < \cdots < t_n = T$ is a partition of some finite interval and $\eta_k : \Omega \to \mathbb{R}$ is a random variable for all $1 \leq k \leq n$, it is rather natural to define

$$\int_0^T \phi(s)\, dW(s) := \sum_{k=1}^{n} \eta_k \cdot \left[W(t_k) - W(t_{k-1})\right]. \tag{1.2}$$

In a way, what allowed us to extend Wiener's integral beyond step functions was the isometry $\|\mathscr{W}(h)\|_{L^2(\Omega)} = \|h\|_{L^2([0,\infty))}$. Hence, if we had a similar isometry here, we could also extend the above definition beyond step*processes*. It depends on the measurability of the η_k's whether or not we have such an isometry. Before proceeding, let us give an economic interpretation of the preliminary Itô-Integral defined above. This will also give some intuition about the measurability we will require.

Suppose that $W(t)$ gives the (random) value of an asset at time t, which we also allow to be negative. Assuming we hold $\eta_k(\omega)$ of this asset over the interval $[t_{k-1}, t_k)$, then we *pay* $\eta_k(\omega)W(t_{k-1},\omega)$ at time t_{k-1} to buy and *get* $\eta_k(\omega)W(t_k,\omega)$ at time t_k when selling the asset. Thus, if $\phi(t,\omega) = \sum_{k=1}^n \eta_k(\omega)\mathbb{1}_{[t_{k-1},t_k)}$ is a plan of holding the asset in question (i.e., and *investment plan*), then

$$\int_0^{t_n} \phi(s)\,dW(s) = \sum_{k=1}^n \eta_k(\omega)(W(t_k) - W(t_{k-1}))$$

is our total gain/loss over the time period $[0, t_n]$ when following this investment plan.

We note that the investment plan ϕ may well depend on ω, e.g. it may incorporate rules such as buying at time t_k an amount of the asset depending on the value of that asset at time t_k; however, it makes no sense economically to allow the amount to be dependent on the value of the asset at a future time such as t_{k+1}, as we cannot look into the future.

We will again denote the filtration generated by the Brownian motion $(W(t))_{t\geq 0}$ by $\mathbb{F}^W = (\mathscr{F}_t^W)_{t\geq 0}$.

Theorem 1.3.1. *Let $0 = t_0 < t_1 < \cdots < t_n = T$ be a partition of $[0,T]$ and $\eta_k \in L^2(\Omega, \mathscr{F}_{t_{k-1}}, \mathbb{P})$ for $1 \leq k \leq n$. For $\phi := \sum_{k=1}^n \eta_k \mathbb{1}_{[t_{k-1},t_k)}$, we define $\int_0^T \phi(t)\,dW(t)$ by (1.2). Then*

$$\mathbb{E}\Big|\int_0^T \phi(t)\,dW(t)\Big|^2 = \mathbb{E}\int_0^T |\phi(t)|^2\,dt\,.$$

Proof. Let us start with some preliminary observations. If $\eta \in L^1(\Omega, \mathscr{F}_t^W, \mathbb{P})$, then for all $s > 0$ we have

$$\mathbb{E}\big[\eta[W(t+s)-W(t)]^2\big] = \mathbb{E}\big[\mathbb{E}[\eta[W(t+s)-W(t)]^2|\mathscr{F}_t^W]\big] = \mathbb{E}\big[\eta\mathbb{E}[W(t+s)-W(t)]^2\big] = s\mathbb{E}\eta\,.$$

Here, we have used the $\mathscr{F}_t^W$-measurability of η and the independence of $W(t+s) - W(t)$ from $\mathscr{F}_t^W$ which yields that $\mathbb{E}\big[[W(t+s) - W(t)]^2\big|\mathscr{F}_t^W\big] = \mathbb{E}(W(t+s) - W(t))^2 = s$.

Similarly, for $s_1 < t_1 < s_2 < t_2$ and $\mathscr{F}_{s_2}^W$-measurable η, we obtain

$$\begin{aligned}\mathbb{E}\big[\eta(W(t_1) - W(s_1))(W(t_2) - W(s_2)\big] &= \mathbb{E}\big[\mathbb{E}[\eta(W(t_1) - W(s_1))(W(t_2) - W(s_2))|\mathscr{F}_{s_2}^W]\big]\\ &= \mathbb{E}\big[\eta(W(t_1) - W(s_1))\mathbb{E}[W(t_s) - W(s_2)]\big] = 0\,.\end{aligned}$$

Using these two facts, we obtain for $\phi = \sum_{k=1}^n \eta_k \mathbb{1}_{[t_{k-1},t_k)}$, where η_k is $\mathscr{F}_{t_{k-1}}^W$-measurable

$$\begin{aligned}\mathbb{E}\Big|\int_0^T \phi(t)\,dW(t)\Big|^2 &= \mathbb{E}\Big(\sum_{k=1}^n \eta_k(W(t_k) - W(t_{k-1}))\Big)^2\\ &= \mathbb{E}\sum_{k=1}^n \eta_k^2(W(t_k) - W(t_{k-1}))^2\\ &\quad +2\mathbb{E}\sum_{k=1}^n\sum_{j=k+1}^n \eta_k\eta_j(W(t_k) - W(t_{k-1}))(W(t_j) - W(t_{j-1}))\\ &= \mathbb{E}\sum_{k=1}^n \eta_k^2(t_k - t_{k-1}) + 0\end{aligned}$$

$$= \mathbb{E}\int_0^T |\phi(t)|^2\,dt\,.$$

□

By Theorem 1.3.1, the integral defined by (1.2) gives a partial isometry between a subspace of $L^2(\Omega\times[0,\infty),\mathbb{P}\otimes\lambda)$, namely that of "predictable step processes", and a subspace of $L^2(\Omega,\mathbb{P})$. Hence it can be uniquely extended to the closure of that space. This extension is exactly the Itô integral. We do not go into details here, since we will do this construction in more generality in Chapter 3. Instead, let us use this extension to compute $\int_0^T W(t)\,dW(t)$.

Example 1.3.2. Fix $T>0$. For $n\in\mathbb{N}$, put

$$\phi_n := \sum_{k=1}^n W\Big(\frac{(k-1)T}{n}\Big)\mathbb{1}_{[\frac{(k-1)T}{n},\frac{(kT}{n})}\,.$$

Then ϕ_n is an adapted step process. Since W has continuous paths, $\phi_n\to W$ pointwise $\mathbb{P}\otimes\lambda$-a.e. on $\Omega\times[0,T]$. Moreover, $|\phi_n(t,\omega)|\le\sup_{0\le t\le T}|W(t)|$. We will see in Corollary 2.1.9 that the latter is square integrable. Hence, by dominated convergence and the above,

$$\int_0^T W(t)\,dW(t) = \lim_{n\to\infty}\int_0^T\phi_n(t)\,dW(t)\,.$$

Let us abbreviate $W_{k,n}:=W(\frac{kT}{n})$. Noting that

$$W_{k,n}^2-W_{k-1,n}^2 = (W_{k,n}-W_{k-1,n})^2+2W_{k-1,n}(W_{k,n}-W_{k-1,n})$$

we obtain

$$\begin{aligned}\int_0^T\phi_n(t)\,dW(t) &= \sum_{k=1}^n W_{k-1,n}(W_{k,n}-W_{k-1,n})\\ &= \frac12\sum_{k=0}^{n-1}\big(W_{k,n}^2-W_{k-1,n}^2\big)-\frac12\sum_{k=0}^{n-1}(W_{k,n}-W_{k-1,n})^2\\ &= \frac12(W_{n,n}^2-W_{0,n}^2)-\frac12\sum_{k=0}^{n-1}(W_{k,n}-W_{k-1,n})^2\\ &\to \frac12 W(T)^2-\frac12 T\end{aligned}$$

in $L^2(\Omega)$, since $\sum_{k=1}^n(W_{k,n}-W_{k-1,n})^2\to T$ in $L^2(\Omega)$ which follows from Lemma 1.3.3 below.

Hence,

$$\int_0^T W(t)\,dW(t) = \frac12 W(T)^2-\frac12 T\,.$$

Lemma 1.3.3. *Let $T>0$ and $\pi:=(0=t_0<t_1<\cdots<t_n=T)$ be a partition of $[0,T]$. Let us put*

$$V^2(W;\pi,T):=\sum_{k=1}^n|W(t_k)-W(t_{k-1})|^2$$

Then $\mathbb{E}|V^2(W;\pi,T)-T|^2\le 2T|\pi|$. In particular, if π_n is a sequence of partitions with $|\pi_n|\to0$ as $n\to\infty$, then $V^2(W;\pi_n,T)\to T$ in $L^2(\Omega)$.

Proof. Let us put $D_k:=W(t_k)-W(t_{k-1})$ and $X_k=D_k^2-(t_k-t_{k-1})$. Then we have $V^2(W;\pi,T)-T=\sum_{k=1}^n X_k$. Consequently,

$$\mathbb{E}|V^2(W;\pi,T)-T|^2 = \mathbb{E}\sum_{k=1}^n X_k^2+2\mathbb{E}\sum_{k=1}^n\sum_{j=k+1}^n X_kX_j = \mathbb{E}\sum_{k=1}^n X_k^2$$

since $\mathbb{E}(X_kX_j) = 0$ for $k \neq j$ since X_k and X_j are independent random variables with mean 0.

Now note that

$$\begin{aligned}\mathbb{E}X_k^2 &= \mathbb{E}\big[D_k^4 - 2D_k^2(t_k - t_{k-1}) + (t_k - t_{k-1})^2\big]\\ &= 3(t_k - t_{k-1})^2 - 2(t_k - t_{k-1})^2 + (t_k - t_{k-1})^2 = 2(t_k - t_{k-1})^2\end{aligned}$$

since D_k, being Gaussian with Variance $t_k - t_{k-1} =: \sigma_k^2$, satisfies $\mathbb{E}D_k^4 = 3\sigma_k^4$. Consequently,

$$\mathbb{E}|V^2(W;\pi,T) - T|^2 = 2\sum_{k=1}^n (t_k - t_{k-1})^2 \leq 2|\pi|\sum_{k=1}^n (t_k - t_{k-1}) = 2T|\pi|\,. \qquad \square$$

1.4. Exercises

(1) A *Gaussian process* is a stochastic process $(X(t))_{t\geq 0}$ such that for every choice of points $t_1, \dots, t_n$ in $[0,\infty)$ the random vector $(X(t_1), \dots, X(t_n))$ is Gaussian, i.e. for all $\alpha \in \mathbb{R}^n$ the random variable $\sum_{k=1}^n \alpha_k X(t_k)$ is Gaussian.

Clearly, if $(h_t)_{t\geq 0}$ is a collection of functions in $L^2([0,\infty))$, then $X(t) = \mathscr{W}(h_t)$ is a Gaussian process, as $\mathscr{W}$ is linear. In particular, Brownian motion is a Gaussian process.

The *covariance function* of a Gaussian process is the function $C : [0,\infty) \to \mathbb{R}$, given by $C(s,t) = \mathbb{E}\big[X(t)X(s)\big]$.

Show that a Gaussian process is a Brownian motion if and only if its covariance function is $s \wedge t$. Conclude that if $(W(t))_{t\geq 0}$ is a Brownian motion, then $B(t) := tW(t^{-1})$ for $t > 0$ and $B(0) = 0$ is a Brownian motion.

(2) Let us look again at Example 1.2.6. We have proved existence and uniqueness of solutions to the Ornstein-Uhlenbeck equation. By slightly changing the setting, we can also allow initial datums x_0 which are random variables in their own right: if $\mathscr{G}$ is a σ-algebra independent of $\mathscr{F}_t$ for all $t \geq 0$, then $\mathbb{F}$, defined by $\mathscr{F}_t = \sigma(\mathscr{G} \cup \mathscr{F}_t^W)$ is a σ-algebra and $(W(t))_{t\geq 0}$ is an $\mathbb{F}$-Brownian motion. Now basically the same proof shows that for $\mathscr{F}_0$-measurable initial datums x_0 the unique solution of the Ornstein Uhlenbeck equation is given by $X(t) = e^{at}x_0 + \int_0^t \sigma e^{a(t-s)}\,dW(s)$.

Suppose that x_0 has normal distribution with mean 0 and variance q. Determine the law of $X(t)$ and try to find a q such that all random variables $X(t)$ have the same distribution. Such a distribution is called *stationary*.

(3) In Example 1.3.2, we have seen that $\int_0^t W(s)\,dW(s) = \frac{1}{2}W(t)^2 - \frac{1}{2}t$. One should compare this with the formula $\int_0^t x\,dx = \frac{1}{2}t^2$ or, for those familiar with Stieltjes integrals, with the formula $\int_0^t g(t)\,dg(t) = \frac{1}{2}g(t)^2$ for a function g of bounded variation with $g(0) = 0$. In the formula for the stochastic integral, we obtain the *Itô correction term* $\frac{1}{2}t$. This is due to the fact that Brownian motion has unbounded variation.

Let us try a different approximation of the stochastic integral $\int_0^T W(t)dW(t)$. Namely, for a given parameter $\theta \in [0,1]$ and a partition $\pi = (0 = t_0 < t_1 < \cdots < t_n = T)$, we use

$$\phi_\pi^\theta := \sum_{k=1}^n \big[(1-\theta)W(t_{k-1}) + \theta W(t_k)\big]\mathbb{1}_{[t_{k-1},t_k)}$$

as an approximation for W. We could then define

$$I_\theta := \sum_{k=1}^n \big[(1-\theta)W(t_{k-1}) + \theta W(t_k)\big](W(t_k) - W(t_{k-1}))$$

as an approximation for the stochastic integral. Here $\theta = 0$ corresponds to the approximation in Itô's integral, $\theta = 1$ leads to the so-called *backward Itô integral* and $\theta = \frac{1}{2}$ to the *Stratonovich integral.*

Show that for as $|\pi| \to 0$ we have $I_\theta \to \frac{1}{2}W(T)^2 + (\theta - \frac{1}{2})T$.

(4) Give an example of a step function $\phi = \sum_{k=1}^{n} a_k \mathbb{1}_{[t_{k-1},t_k)}$ which is *not* adapted, i.e. even though all a_k are square integrable, not all a_k are $\mathscr{F}_{t_{k-1}}$-measurable, such that

$$\mathbb{E}\Big|\sum_{k=1}^{n} a_k \cdot [W(t_k) - W(t_{k-1})]\Big|^2 \neq \mathbb{E}\int_0^T |\phi(t)|^2 \, dt\,.$$

CHAPTER 2

Continuous Local Martingales

2.1. Martingales: Basic results

Throughout, we are given a filtered probability space $(\Omega, \Sigma, \mathbb{F}, \mathbb{P})$, i.e. $\mathbb{F} = (\mathscr{F}_t)_{t\in I}$ is a filtration on the probability space $(\Omega, \Sigma, \mathbb{P})$. Here $I \subset \mathbb{R} \cup \{\infty\}$ is a directed set. We are mainly interested in the cases $I = \mathbb{N}$ (discrete time), $I = [0, \infty)$ (continuous time) and $I = [0, T]$ (continuous time with finite time horizon T). A filtered probability space is also called a *stochastic basis*.

A stochastic process $(X(t))_{t\in I}$ is called *adapted* to $\mathbb{F}$, if $X(t)$ is $\mathscr{F}_t$-measurable for all $t \in I$. The smallest filtration with this property is given by $\mathscr{F}_t^X := \sigma(X_r : r \le t)$ is called the filtration generated by the process $(X(t))_{t\in I}$, cf. the filtration $\mathbb{F}^W$ from the previous chapter.

Definition 2.1.1. A *martingale*, more precisely, an $\mathbb{F}$-martingale, is an adapted process $(X(t))_{t\in I}$ of integrable random variables such that for all $t, s \in I$ with $s \le t$ we have

$$\mathbb{E}[X(t)|\mathscr{F}_s] = X(s) \quad \mathbb{P}\text{-}a.s.$$

If for such t and s we merely have $\mathbb{E}[X(t)|\mathscr{F}_s] \ge X(s)$ almost surely, then $(X(t))_{t\in I}$ is called a *submartingale*.

Example 2.1.2. Brownian motion is a martingale with respect to its natural filtration $\mathbb{F}^W$. Indeed, since $W(t) = W(t) - W(s) + W(s)$ where $W(t) - W(s)$ has mean 0 and is independent of $\mathscr{F}_s^W$ and $W(s)$ is $\mathscr{F}_s^W$-measurable, we obtain $\mathbb{E}[W(t)|\mathscr{F}_s^W] = W(s)$ for all $s \le t$. Similarly, an $\mathbb{F}$-Brownian motion is an $\mathbb{F}$-martingale.

Lemma 2.1.3. *Let $(X(t))_{t\in I}$ be a martingale and $\varphi : \mathbb{R} \to \mathbb{R}$ be convex. Then $(\varphi(X(t)))_{t\in I}$ is a submartingale. In particular, $(X(t)^2)_{t\in I}$ and $(|X(t)|)_{t\in I}$ are submartingales. If φ is convex and non increasing, then also for submartingales $(X(t))_{t\in I}$ the process $(\varphi(X(t)))_{t\in I}$ is a submartingale.*

Proof. We have $\mathbb{E}[\varphi(X(t))|\mathscr{F}_s] \ge \varphi\big(\mathbb{E}[X(t)|\mathscr{F}_s]\big) = \varphi(X(s))$ almost surely, by Jensen's inequality for conditional expectation. For the addendum, note that if $(X(t))_{t\in I}$ is merely a submartingale but φ is non increasing, then $\varphi\big(\mathbb{E}[X(t)|\mathscr{F}_s]\big) \ge \varphi(X(s))$ almost surely. ☐

A *stopping time*, more precisely, an $\mathbb{F}$-stopping time, is a a map $\tau : \Omega \to I$ such that $\{\tau \le t\} := \tau^{-1}\big(\{s \in I : s \le t\}\big) \in \mathscr{F}_t$. Every stopping time τ induces a σ-algebra $\mathscr{F}_\tau$ via

$$\mathscr{F}_\tau = \{A \in \Sigma : A \cap \{\tau \le t\} \in \mathscr{F}_t,\ \forall\, t \in I\}\,.$$

We collect some basic properties of stopping times, omitting the easy proofs which may be found in every probability book containing a chapter on martingales. In order to get used to the concept of stopping times, we suggest the reader to prove the results themselves.

Lemma 2.1.4. *Let τ, σ be stopping times. Then*

1. *$\tau \wedge \sigma$ and $\tau \vee \sigma$ are stopping times.*
2. *τ is $\mathscr{F}_\tau$-measurable.*
3. *if I is countable and $(X(t))_{t\in I}$ is adapted, then $X(\tau)$ is $\mathscr{F}_\tau$-measurable.*
4. *$\mathscr{F}_\tau = \mathscr{F}_t$ on $\{\tau = t\}$ for all $t \in I$, i.e. the induced σ-algebras on the set $\{\tau = t\}$ agree.*

(5) $\mathscr{F}_\sigma \cap \{\sigma \leq \tau\} \subset \mathscr{F}_{\sigma\wedge\tau} = \mathscr{F}_\sigma \cap \mathscr{F}_\tau$.
(6) *if* $\sigma \leq \tau$ *then* $\mathscr{F}_\sigma \subset \mathscr{F}_\tau$.

We now proceed to important results for martingales, including optional sampling and Doob's maximal inequality. As is typical, such results are proved first for finite index sets I and then extended via approximation. We do not strike for greatest possible generality and only establish results in the extend needed in what follows.

THEOREM 2.1.5. *(Optional sampling)*
Let I be a finite set, $(X(t))_{t\in I}$ be a submartingale and τ, σ be stopping times. Then

$$\mathbb{E}[X(\tau)|\mathscr{F}_\sigma] \geq X(\tau\wedge\sigma).$$

If $(X(t))_{t\in I}$ is even a martingale then the inequality above is in fact an equality.

PROOF. We assume that $I = \{t_1, \dots, t_n\}$, where $t_1 < \cdots < t_n$. We have to prove that

$$\int_A X(\tau)\, d\mathbb{P} \geq \int_A X(\tau\wedge\sigma)\, d\mathbb{P}$$

for all $A \in \mathscr{F}_\sigma$. Writing A as disjoint union of $A \cap \{\sigma = t_k\}$, it suffices to prove

$$\int_{A\cap\{\sigma=t_k\}} X(\tau)\, d\mathbb{P} \geq \int_{A\cap\{\sigma=t_k\}} X(\tau\wedge\sigma)\, d\mathbb{P} = \int_{A\cap\{\sigma=t_k\}} X(\tau\wedge t_k)\, d\mathbb{P}$$

Since $\mathscr{F}_\sigma$ and $\mathscr{F}_{t_k}$ agree on $\{\sigma = t_k\}$, this is equivalent with

$$\mathbb{E}[X(\tau)|\mathscr{F}_{t_k}] \geq X(\tau\wedge t_k)\,. \tag{2.1}$$

Observing that

$$\begin{aligned}\mathbb{E}\big[X(\tau\wedge t_k)|\mathscr{F}_{t_{k-1}}\big] &= \mathbb{E}\big[X(t_k)\mathbb{1}_{\{\tau>t_{k-1}\}} + X(\tau\wedge t_{k-1})\mathbb{1}_{\{\tau\leq t_{k-1}\}}\big|\mathscr{F}_{t_{k-1}}\big]\\ &= \mathbb{E}[X(t_k)|\mathscr{F}_{t_{k-1}}]\mathbb{1}_{\{\tau>t_{k-1}\}} + X(\tau\wedge t_{k-1})\mathbb{1}_{\{\tau\leq t_{k-1}\}}\\ &\geq X(t_{k-1})\mathbb{1}_{\{\tau>t_{k-1}\}} + X(\tau\wedge t_{k-1})\mathbb{1}_{\{\tau\leq t_{k-1}\}} = X(\tau\wedge t_{k-1})\end{aligned}$$

(2.1) follows inductively, starting with $k = n$, observing that $\tau\wedge t_n = \tau$.
In the case of martingales, observe that all inequalities are in fact equalities. □

Actually, a weaker statement than in Theorem 2.1.5 characterizes martingales.

PROPOSITION 2.1.6. *Let I be an arbitrary index set with minimal element* 0 *and $(X(t))_{t\in I}$ be an adapted process with $\mathbb{E}|X(t)| < \infty$ for all $t \in I$. Then X is a martingale if and only if $\mathbb{E}X(\tau) = \mathbb{E}X(0)$ for all stopping times τ that take at most two values.*

PROOF. For $s, t \in I$ with $s < t$ and $A \in \mathscr{F}_s$, put $\tau = s\mathbb{1}_A + t\mathbb{1}_{A^c}$. Then τ is a stopping time. Thus, by hypothesis

$$\mathbb{E}X(0) = \mathbb{E}X(\tau) = \mathbb{E}X(s)\mathbb{1}_A + \mathbb{E}X(t)\mathbb{1}_{A^c}$$

On the other hand, t is a stopping time whence $\mathbb{E}(0) = \mathbb{E}X(t) = \mathbb{E}X(t)\mathbb{1}_A + \mathbb{E}X(t)\mathbb{1}_{A^c}$. Subtracting both equations, it follows that $\mathbb{E}(X(t)-X(s))\mathbb{1}_A = 0$. As $A \in \mathscr{F}_s$ was arbitrary, $\mathbb{E}[X(t)|\mathscr{F}_s] = X(s)$. The converse follows from Theorem 2.1.5, noting that $(X(t))_{t\in I}$ is a martingale if and only if $(X(t))_{t\in\tilde I}$ is a martingale for all $\tilde I \subset I$ with at most two elements. □

LEMMA 2.1.7. *Let I be finite and $(X(t))_{t\in I}$ be an adapted submartingale and $T \in I$. Then, for every $r > 0$, we have*

$$\mathbb{P}\big(\sup_{t\leq T} X(t) > r\big) \leq r^{-1}\int_{\{\sup_{t\leq T} X(t)>r\}} X(T)\, d\mathbb{P} \leq r^{-1}\mathbb{E}[X^+(T)]\,.$$

Proof. Define the stopping time $\sigma := \min\{t \leq T : X(t) > r\}$ where $\min\emptyset := T$. Moreover, let $A = \{\max_{t\leq T} X(t) > r\}$. Since $\{\sigma = t\} \cap A = \{X(s) \leq r : \forall\, s < t\} \cap \{X(t) > r\} \in \mathscr{F}_t$ by adaptedness, we easily obtain that $A \in \mathscr{F}_\sigma$.

Since $X(\sigma) > r$ on A we obtain, using Theorem 2.1.5 with $\tau = T$, that

$$r\mathbb{P}(A) \leq \int_A X(\sigma)\, d\mathbb{P} \leq \int_A \mathbb{E}[X(T)|\mathscr{F}_\sigma]\, d\mathbb{P} = \int_A X(T)\, d\mathbb{P} \leq \int_A X^+(T)\, d\mathbb{P} \leq \mathbb{E}X^+(T)$$

which is equivalent with the assertion. □

We now easily obtain Doob's maximal inequality.

Theorem 2.1.8. *(Doob's maximal inequality)*
Let I be countable and let $(X(t))_{t\in I}$ be a martingale. Put $X^*(t) := \sup_{s\leq t}|X(t)|$. Then, for $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$, we have $\|X^*(t)\|_p \leq q\|X(t)\|_p$, for all $t \in I$.

Proof. Let us first assume that I is finite.

Since $(X(t))_{t\in I}$ is a martingale, $(|X(t)|)_{t\in I}$ is a submartingale by Lemma 2.1.3. Thus, by Lemma 2.1.7,

$$r\mathbb{P}(X^*(t) > r) \leq \int_{\{X^*(t)>r\}} |X(t)|\, d\mathbb{P}$$

for all $t \in I$.

Consequently, using Fubini's theorem and the Hölder inequality,

$$\begin{aligned}
\|X^*(t)\|_p^p &= p\int_0^\infty \mathbb{P}(X^*(t) > r) r^{p-1}\, dr \\
&\leq p\int_0^\infty \int_{\{X^*(t)>r\}} |X(t)|\, d\mathbb{P}\, r^{p-2}\, dr \\
&= p\int_\Omega |X(t)| \int_0^{X^*(t)} r^{p-2}\, dr\, d\mathbb{P} \\
&= \frac{p}{p-1}\mathbb{E}\big(|X(t)||X^*(t)|^{p-1}\big) \\
&\leq \frac{p}{p-1}\|X(t)\|_p\|X^*(t)^{p-1}\|_q = q\|X(t)\|_p\|X^*(t)\|_p^{p-1}\,.
\end{aligned}$$

This implies that $\|X^*(t)\|_p \leq q\|X(t)\|_p$.

In the case where I is infinite, we write $I = \bigcup I_n$, where I_n is a finite index set, increasing in n. If we put $X_n^*(t) := \sup_{s\leq t, s\in I_n}|X(s)|$ then $X_n^*(t) \uparrow X^*(t)$, almost surely. Moreover, $\|X_n^*(t)\|_p \leq q\|X(t)\|_p$ for all $n \in \mathbb{N}$ by the above. By monotone convergence, $\|X^*(t)\|_p \leq q\|X(t)\|_p$. □

Doob's maximal inequality generalizes to continuous time martingales with regular paths. Note that this in particular yields the square-integrability of $\sup_{0\leq s\leq t}|W(s)|$ needed in Example 1.3.2.

Corollary 2.1.9. *Let $(\Omega, \Sigma, \mathbb{F}, \mathbb{P})$ be a filtered probability space where $\mathbb{F} = (\mathscr{F}_t)_{t\geq 0}$. Moreover, let $(X(t))_{t\geq 0}$ be a martingale with right-continuous paths and define $X^*(t) := \sup_{s\leq t}|X(t)|$. Then, for $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$, we have $\|X^*(t)\|_p \leq q\|X(t)\|_p$, for all $t \in I$.*

Moreover, we have $\mathbb{P}(X^* > r) \leq r^{-1}\mathbb{E}X(t)^+$.

Proof. Since $\sup_{0\leq s\leq t}|X(s)| = \sup_{s\in[0,t]\cap\mathbb{Q}}|X(s)|$, the first assertion follows immediately from Theorem 2.1.8. For the second assertion, write $\{t\} \cup \mathbb{Q} \cap [0, t] = \bigcup I_n$, where I_n is an increasing sequence of finite sets. Then $\{\sup r \in I_n > r\} \uparrow \{X^*(t) > r\}$. Thus the second assertion follows from Lemma 2.1.7. □

We now also extend the optional sampling theorem to show that a stopped continuous time martingale is again a martingale. Let us start with a lemma.

Lemma 2.1.10. *Let $\mathbb{F} = (\mathscr{F})_{t\in I}$, where $I \subset [0,\infty)$ be a filtration on the probability space $(\Omega, \Sigma, \mathbb{P})$ and let τ be an $\mathbb{F}$ stopping time. Then there exists a sequence τ_n of $\mathbb{F}$ stopping times which take only countably many values and decreases to τ.*

Proof. Put

$$\tau_n := \sum_{k\in\mathbb{N}} \frac{k}{2^n} \mathbb{1}_{\{\frac{k-1}{2^n} < \tau \le \frac{k}{2^n}\}}$$

Then clearly, τ_n takes only countably many values and decreases to τ. Moreover, τ_n is a stopping time, since

$$\{\tau_n \le t\} = \bigcup_{k:\,\frac{k}{2^n} \le t} \{\tau \le \frac{k}{2^n}\} \in \mathscr{F}_t$$

since $\{\tau < \frac{k}{2^n}\} = \bigcup_m \{\tau \le \frac{k}{2^n} - \frac{1}{m}\}$ where $\{\tau \le \frac{k}{2^n} - \frac{1}{m}\} \in \mathscr{F}_{\frac{k}{2^n} - \frac{1}{m}} \subset \mathscr{F}_t$. ☐

Proposition 2.1.11. *Let $(X(t))_{t\in I}$, where I is an interval in $[0,\infty)$ be a continuous martingale on the filtered probability space $(\Omega, \Sigma, \mathbb{F}, \mathbb{P})$ and τ be an $\mathbb{F}$-stopping time. Then $X^\tau := (X(\tau \wedge t))_{t\in I}$ is a martingale.*

Proof. Let us first prove that $X(\tau \wedge t)$ is $\mathscr{F}_t$-measurable, i.e. the stopped process is again adapted.

To that end, first note that for every $t \in I$, the map $(s,\omega) \mapsto X(t,\omega)$ is $\mathscr{B}([0,t]) \otimes \mathscr{F}_t$-measurable (one says that X is *progressively measurable*). Indeed, for every $n \in \mathbb{N}$ the map $(s,\omega) \mapsto \sum_{k=1}^n X(\frac{(k-1)t}{n}, \omega)\mathbb{1}_{[\frac{(k-1)t}{n}, \frac{kt}{n})}(s) + X(t,\omega)\mathbb{1}_{\{t\}}(s)$ has the claimed measurability by adaptedness and they converge to the map above by continuity of the paths.

Moreover, the map $(s,\omega) \mapsto (\tau(\omega)\wedge s, \omega)$ is $\mathscr{B}([0,t])\otimes\mathscr{F}_t/\mathscr{B}([0,t])\otimes\mathscr{F}_t$-measurable, since τ is a stopping time. Hence also the composition, i.e. $(s,\omega) \mapsto X(s\wedge\tau,\omega)$ is $\mathscr{B}([0,t]) \otimes \mathscr{F}_t$-measurable.

Now, let τ_n be a sequence of stopping times decreasing to τ such that τ_n only takes countably many values. Note that the proof of Lemma 2.1.10 yields that we can choose τ_n such that $\tau_n \wedge r$ takes only finitely many values, say $t_1, \ldots, t_{k(r)}$. By continuity of the paths, $X(t \wedge \tau_n)$ converges to $X(t)$ for all $t \in I$. Fix $t, s \in I$ with $t > s$ and put $I_n := \{t, s\} \cup \{t_1, \ldots, t_{k(t)}\}$.

Applying Theorem 2.1.5 with $I = I_n$, $\tau = t \wedge \tau_n$ and $\sigma = s$, we obtain $\mathbb{E}[X(t\wedge\tau_n)|\mathscr{F}_s] = X(s \wedge \tau_n)$. We note that if $n \to \infty$, then $X(t \wedge \tau_n) \to X(t\wedge\tau)$ for all $t \in I$. Moreover, by optional sampling, $X(t \wedge \tau_n) = \mathbb{E}(X(t)|\mathscr{F}_{\tau_n})$, proving that the random variables $X(t \wedge \tau_n)$ are equi-integrable. Thus $X(t\wedge\tau_n)$ converges to $X(t\wedge\tau)$ in $L^1(\Omega)$ and the claim follows upon $n \to \infty$. ☐

In what follows, we will always write X^τ for the process $X(\cdot \wedge \tau)$.

For continuous martingales, Proposition 2.1.11 is most often applied with τ a certain *hitting time*.

Example 2.1.12. Let $(X(t))_{t\ge 0}$ be an adapted process with continuous paths and $C \subset \mathbb{R}$ be closed. Then $\tau := \inf\{t > 0 : X(t) \in C\}$ is a stopping time. Indeed, by continuity of the paths

$$\{\tau \le t\} = \bigcap_{n\in\mathbb{N}} \bigcup_{r\in\mathbb{Q}\cap[0,t]} \{\mathrm{dist}(X(r), C) \le \frac{1}{n}\} \in \mathscr{F}_t\,.$$

We end this section by introducing some spaces of martingales.

Definition 2.1.13. Let $I = [0,T]$. A *continuous, square integrable martingale* is a martingale $(X(t))_{t\in[0,T]}$ with continuous paths such that $\sup_{t\in[0,T]} \mathbb{E}|X(t)|^2 < \infty$ and $X(0) = 0$ almost surely. We write $\mathsf{M}_2([0,T]; \Omega, \Sigma, \mathbb{F}, \mathbb{P})$ or, if the stochastic basis is understood, $\mathsf{M}_2([0,T])$ for the space of all continuous, square integrable martingales with respect to $\mathbb{F}$.

A *continuous local martingale* is an adapted process $(X(t))_{t\in[0,T]}$ with continuous paths such that there exists a sequence of stopping times τ_n with $\tau_n \uparrow T$ almost surely such that X^{τ_n} is a martingale for all $n \in \mathbb{N}$. Such a sequence is called *localizing sequence.* We write $\mathsf{M}_{\mathrm{loc}}([0,T];\Omega,\Sigma,\mathbb{F},\mathbb{P})$ or, briefly, $\mathsf{M}_{\mathrm{loc}}([0,T])$ for the space of all continuous local martingales

Lemma 2.1.14. *Let $(X(t))_{t\in[0,T]}$ be a continuous, adapted process and put $\sigma_n := \inf\{t > 0 : |X(t)| \geq n\}$. Then X is a local martingale if and only if X^{σ_n} is a martingale for all $n \in \mathbb{N}$.*

Proof. As $\sigma_n \uparrow T$ almost surely by path continuity, it is clear that X is a local martingale if X^{σ_n} is a martingale for all $n \in \mathbb{N}$. For the converse, assume that τ_n is a sequence of stopping times with $\tau_n \uparrow \infty$ almost surely such that X^{τ_n} is a martingale for all $n \in \mathbb{N}$. By optional sampling, also $X^{\tau_n\wedge\sigma_n}$ is a martingale. Note that $|X(t\wedge\tau_n\wedge\sigma_m)| \leq m$ for all $n \in \mathbb{N}$. Thus, letting $n \to \infty$ we infer from dominated convergence that $X(\cdot \wedge \sigma_m)$ is a martingale for all $m \in \mathbb{N}$. □

Lemma 2.1.15. *$\mathsf{M}_2([0,T])$ is a closed, linear subspace of $L^2(\Omega; C([0,T]))$. On $\mathsf{M}_2([0,T])$, the expression $\|X\|^2 := \mathbb{E}|X(T)|^2$ defines an equivalent norm. Moreover, $\mathsf{M}_{\mathrm{loc}}([0,T])$ is a closed, linear subspace of $L^0(\Omega; C([0,T]))$.*

Proof. By Corollary 2.1.9 with $p = q = 2$, we have $\mathbb{E}|X(T)|^2 \leq \mathbb{E}\|X\|^2_{C([0,T])} \leq 4\mathbb{E}|X(T)|^2 < \infty$. It follows that $\mathsf{M}_2([0,T]) \subset L^2(\Omega; C([0,T]))$. For $0 \leq s < t \leq T$ and $\eta \in L^2(\Omega)$, let $\varphi_{s,t,\eta}(X) := \mathbb{E}\big[(X(s) - \mathbb{E}[X(t)|\mathscr{F}_s])\eta\big]$. Clearly, $\varphi_{s,t,\eta}$ is a bounded linear functional on $L^2(\Omega; C([0,T]))$. Moreover,

$$\mathsf{M}_2([0,T]) = \bigcap_{\eta\in L^2(\Omega)} \bigcap_{0\leq s<t\leq T} \ker\varphi_{s,t,\eta}\,.$$

This proves that $\mathsf{M}_2([0,T])$ is a closed, linear subspace of $L^2(\Omega; C([0,T]))$ and that $\mathbb{E}|X(T)|^2$ defines an equivalent norm on $\mathsf{M}_2([0,T])$.

For the second part first observe that if X and Y are continuous local martingales with localizing sequences τ_n resp. σ_n then $\alpha X+\beta Y$ is a continuous local martingale with localizing sequence $\tau_n\wedge\sigma_n$, as is easy to see. Now let X_n be a sequence of continuous local martingales converging to X in $L^0(\Omega, C([0,T]))$. Passing to a subsequence, we assume that we have convergence pointwise almost everywhere in $C([0,T])$. Put $\tau_n := \inf\{t \in [0,T] : |X(t)| \geq n\}$ and $\sigma_{n,k} := \tau_n\wedge\inf\{t \in [0,T] : |X_k(t)| \geq n+\frac{1}{k}\}$. Then $X_k^{\sigma_{n,k}}$ is a martingale for all $n, k \in \mathbb{N}$. Moreover, $X_k^{\sigma_{n,k}} \to X^{\tau_n}$ almost surely in $C([0,T])$. As $\|X_k^{\sigma_{n,k}}\|_\infty \leq n+1$ for all $k \in \mathbb{N}$, we infer that $X_k^{\sigma_{n,k}} \to X^{\tau_n}$ in $L^2(\Omega; C([0,T]))$. By the first part, X^{τ_n} is a martingale for all n, hence X is a local martingale. □

2.2. Quadratic Variation

Definition 2.2.1. Let $\varphi : [0,T] \to \mathbb{R}$ and $\pi := (0 = t_0 < t_1 < \cdots < t_n = T)$ be a partition of $[0,T]$. For $p \in [1,\infty)$, we put

$$V^p(\varphi,\pi,T) := \sum_{k=1}^n |\varphi(t_k) - \varphi(t_{k-1})|^p\,.$$

The function φ is said to be of *bounded variation* if $V(\varphi,T) := \sup_\pi V^1(\varphi,\pi,T) < \infty$.

If $\lim_{|\pi|\to0} V^2(\varphi,\pi,T)$ exists, we say that φ has finite quadratic variation. In this case, the limit $V^2(\varphi,T) := \lim_{|\pi|\to0} V^2(\varphi,\pi,T)$ is called the *quadratic variation.*

As is well-known, a function is of bounded variation if and only if it is the difference of two increasing functions. Moreover, a function of bounded variation is bounded and $\|\varphi\|_\infty \leq \|\varphi(0)\| + V(\varphi,T)$.

Functions of bounded variation play an important role in the theory of Riemann-Stieltjes integrals. Namely, if φ is of bounded variation and f is a continuous function, then, given a sequence of partitions $\pi_n := (0 = t_0^{(n)} < t_1^{(n)} < \cdots < t_{k_n}^{(n)} = T)$ with mesh size converging to 0, the limit

$$\int_0^T f(t)\, d\varphi(t) := \lim_{n\to\infty} \sum_{j=1}^{k_n} f(t_j^{(n)})\big[\varphi(t_j^{(n)}) - \varphi(t_{j-1}^{(n)})\big]$$

exists. As we have already mentioned, it would be tempting to construct stochastic integrals pathwise as Riemann-Stieltjes integrals. However, this approach does not work as Brownian motion has almost surely paths of unbounded variation. This is what we prove next.

LEMMA 2.2.2. *Let $\varphi : [0,T] \to \mathbb{R}$ be a continuous function. If φ is of bounded variation, then $V^2(\varphi, T) = 0$.*

PROOF. Let $\pi = (0 = t_0 < t_1 < \cdots < t_n = T)$ be a partition of $[0,T]$. We put $M(\varphi, \pi) := \sup_{1\le k\le n} |\varphi(t_k) - \varphi(t_{k-1})|$. Then

$$V^2(\varphi, \pi, T) := \sum_{k=1}^n |\varphi(t_k) - \varphi(t_{k-1})|^2 \le M(\varphi, \pi)V(\varphi, T)\,.$$

Since φ is continuous on the compact interval $[0,T]$, it is uniformly continuous, whence $M(\varphi, \pi) \to 0$ as $|\pi| \to 0$. Thus $V^2(\varphi, T) = 0$. □

EXAMPLE 2.2.3. It follows from Lemma 1.3.3 that, pathwise, $\langle W\rangle_t := V^2(W, t) = t$. By Lemma 2.2.2, the paths of Brownian motion are almost surely of unbounded variation.

In fact, only trivial continuous local martingales have bounded variation as Lemma 2.2.4 below shows. Before proceeding, let us make some preliminary observations, which generalize computations in Example 1.3.2.

Let $(X(t))_{t\in[0,T]}$ be a martingale with $X(0) = 0$ and let $\pi = (0 = t_0 < t_1 < \cdots < t_n = T)$ be a partition of $[0,T]$. Then

$$\sum_{k=1}^n \big[X(t_k) - X(t_{k-1})\big]^2 = X(T)^2 - 2\sum_{k=1}^n X(t_{k-1})\big[X(t_k) - X(t_{k-1})\big] \tag{2.2}$$

which follows by using the equality $(a - b)^2 = a^2 - b^2 - 2b(a - b)$ in every summand and canceling out in the telescoping sum which appears.

Moreover,

$$\mathbb{E}\sum_{k=1}^n \big[X(t_k) - X(t_{k-1})\big]^2 = \mathbb{E}X(T)^2 \tag{2.3}$$

as

$$\begin{aligned}
\mathbb{E}\sum_{k=1}^n \big[X(t_k) - X(t_{k-1})\big]^2 - \mathbb{E}X(T)^2 &= -2\mathbb{E}\sum_{k=1}^n X(t_{k-1})\big(X(t_k) - X(t_{k-1})\big)\\
&= -2\mathbb{E}\sum_{k=1}^n \mathbb{E}\big[X(t_{k-1})\big(X(t_k) - X(t_{k-1})\big)\big|\mathscr{F}_{t_{k-1}}\big]\\
&= -2\mathbb{E}\sum_{k=1}^n X(t_{k-1})\mathbb{E}\underbrace{\big[X(t_k) - X(t_{k-1})\big|\mathscr{F}_{t_{k-1}}\big]}_{=0}
\end{aligned}$$

by the martingale property.

LEMMA 2.2.4. *Let $(X(t))_{t\in[0,T]}$ be a continuous local martingale with (pathwise) bounded variation and $X(0) = 0$ a.s. Then $X = 0$ almost surely.*

Proof. Let $V(t)$ denote the total variation of X on $[0,t]$. Then $V(t)$ is a continuous, adapted process, whence $\tau_n := \inf\{t \in [0,T] : |V(t)| \geq n\}$, where $\inf \emptyset := T$, is a stopping time. Moreover by continuity of paths and assumption, $\Omega = \bigcup_{n\in\mathbb{N}}\{\tau_n = T\}$. If $(X(t))_{t\in I}$ is a continuous local martingale with bounded variation, then X^{τ_n} is a uniformly bounded, continuous martingale with uniformly bounded variation. Clearly, if the latter are 0 almost surely, then so is X.

Hence it suffices to consider continuous martingales X whose total variation is uniformly bounded, say by M. For $t \in [0,T]$, $n \in \mathbb{N}$ and $k = 0,\ldots,n$, put $X_{n,k} := X(\frac{kt}{n})$. By continuity of the paths,

$$Q_n := \sum_{k=1}^{n}(X_{n,k} - X_{n,k-1})^2 \leq M \sup_{1\leq k\leq n} |X_{n,k} - X_{n,k-1}| \to 0$$

almost surely and in L^1, as Q_n is bounded by M^2. Moreover, by (2.3), $\mathbb{E}X(t)^2 = \mathbb{E}Q_n \to 0$ as $n \to \infty$. Hence $X(t) = 0$ a.s. for all $t \in [0,T]$. □

In Example 1.3.2, we have seen that $\int_0^t W(s)\,dW(s) = \frac{1}{2}W(t) - \frac{1}{2}t$. Thus, the quadratic variation process $\langle w\rangle_t = t$ appears in the Itô correction term. As we shall see, this is not by accident. Our goal in this section is to prove that every continuous local martingale has a well-defined quadratic variation. We will see later on, that this process plays an important role in stochastic integration. We will only treat the case of a finite time-horizon. This is sufficient for our purposes and simplifies the exposition.

We now come to the main result of this section, namely the existence and uniqueness of the quadratic variation process.

Theorem 2.2.5. *Let $X \in \mathsf{M}_{\mathrm{loc}}([0,T])$. Then there exists a unique continuous, adapted and increasing process $(\langle X\rangle_t)_{t\in[0,T]}$ such that $X^2 - \langle X\rangle$ is a local martingale. Moreover, if $\pi_n := (0 = t_0^{(n)} < t_1^{(n)} < \cdots < t_{k_n}^{(n)} = T)$ is a sequence of partitions of $[0,T]$ with $|\pi_n| \to 0$ and $\pi_n \subset \pi_{n+1}$, then V_{π_n}, defined by*

$$V_{\pi_n}(t) := \sum_{j=1}^{k_n}(X(t_j^{(n)} \wedge t) - X(t_{j-1}^{(n)} \wedge t))^2$$

converges to $\langle X\rangle$ in $L^0(\Omega; C([0,T]))$. If $X \in \mathsf{M}_2([0,T])$, then $X^2 - \langle X\rangle$ is even a martingale.

Definition 2.2.6. We call $(\langle X\rangle_t)_{t\in[0,T]}$ the *quadratic variation process of X*.

Proof. *Uniqueness:* if $(A(t))_{t\in[0,T]}$ and $(B(t))_{t\in[0,T]}$ are continuous, adapted and increasing processes such that $X(t)^2 - A(t)$ and $X^2(t) - B(t)$ are local martingales, then their difference, $A(t) - B(t)$ is a continuous local martingale which is of bounded variation, as it is the difference of two increasing processes. By Lemma 2.2.4, $A = B$ almost surely.

Existence: We proceed in several steps. In Steps 1 and 2 we assume that X is a uniformly bounded martingale.

Step 1: Given a partition $\pi := (0 = t_0 < t_1 < \ldots < t_n = T)$, we define

$$\begin{aligned} V_\pi(t) &:= \sum_{k=1}^{n}(X(t_k \wedge t) - X(t_{k-1} \wedge t))^2 \\ &= X(t)^2 - 2\sum_{k=1}^{n} X(t_{k-1} \wedge t)(X(t_k \wedge t) - X(t_{k-1} \wedge t)) \end{aligned} \tag{2.4}$$

where the last equality is (2.2). We claim that if X is a martingale, then $V_\pi(t)$ is a continuous adapted process such that $X(t)^2 - V_\pi(t)$ is a martingale.

Continuity and adaptedness of V_π is obvious. For the martingale property, we use Proposition 2.1.6 and let a stopping time τ taking at most two values be given. Noting that $X(t_k\wedge\tau)$ is $\mathscr{F}_{t_k}$-measurable, we obtain

$$\begin{aligned}\mathbb{E}(X^2(\tau)-V_\pi(\tau)) &= 2\mathbb{E}\sum_{k=1}^n X(t_{k-1}\wedge\tau)(X(t_k\wedge\tau)-X(t_{k-1}\wedge\tau))\\ &= 2\mathbb{E}\sum_{k=1}^n \mathbb{E}\big[X(t_{k-1}\wedge\tau)(X(t_k\wedge\tau)-X(t_{k-1}\wedge\tau))\big|\mathscr{F}_{t_{k-1}}\big]\\ &= 2\mathbb{E}\sum_{k=1}^n X(t_{k-1}\wedge\tau)\mathbb{E}\big[(X(t_k\wedge\tau)-X(t_{k-1}\wedge\tau))\big|\mathscr{F}_{t_{k-1}}\big]=0,\end{aligned}$$

since X is a martingale.

Step 2: Construction of $\langle X\rangle$ for uniformly bounded X.

We pick a sequence π_n of partitions with $\pi_n\subset\pi_{n+1}$ and $|\pi_n|\to 0$. By Doob's maximal inequality,

$$\mathbb{E}\|V_{\pi_n}-V_{\pi_m}\|^2_{C([0,T])}=\mathbb{E}\|X^2-V_{\pi_n}-(X^2-V_{\pi_m})\|^2_{C([0,T])}\leq 4\mathbb{E}|V_{\pi_n}(T)-V_{\pi_m}(T)|^2\,.$$

Lemma 2.2.7 below shows that $V_{\pi_n}(T)$ is a Cauchy sequence in $L^2(\Omega)$ whence, by the above, V_{π_n} is a Cauchy sequence in $L^2(\Omega;C([0,T]))$. Thus, V_{π_n} converges to some $\langle X\rangle$ in $L^2(\Omega;C([0,T]))$. Hence also $X^2-V_{\pi_n}\to X^2-\langle X\rangle$ in $L^2(\Omega;C([0,T]))$. By closedness of $\mathsf{M}_2([0,T])$, the process $X^2-\langle X\rangle$ is a martingale; in particular, it is adapted whence also $\langle X\rangle$ is adapted. We now pass to a subsequence such that $V_{\pi_n}\to\langle X\rangle$ almost surely.

To see that $\langle X\rangle$ is increasing, put $D=\bigcup_{n\in\mathbb{N}}\pi_n$. Since $|\pi_n|\to 0$, D is dense in $[0,T]$. By continuity of paths it suffices to show $\langle X\rangle_t\geq\langle X\rangle_s$ for $t\geq s$ with $t,s\in D$.

However, for such t,s, we find n_0 such that $t,s\in\pi_n$ for all $n\geq n_0$. Obviously, $V_{\pi_n}(t)\geq V_{\pi_n}(s)$ for all $n\geq n_0$ hence also $\langle X\rangle_t\geq\langle X\rangle_s$.

Step 3: Extension to general $X\in\mathsf{M}_{\mathrm{loc}}([0,T])$.

Define the stopping times $\tau_n:=\inf\{t\in[0,T]:|X(t)|\geq n\}$ (with $\inf\emptyset=T$) and $X_n(t):=X(\tau_n\wedge t)$. By Step 2, there exist unique continuous, increasing adapted processes $\langle X_n\rangle$ such that $X_n^2-\langle X_n\rangle$ is a martingale.

For $n\geq m$, we have $X_n(t\wedge\tau_m)=X_m(t)$. Hence, since $X_n(t\wedge\tau_m)^2-\langle X_n\rangle_{\tau_m\wedge t}$ and $X_m^2-\langle X_m\rangle$ are martingales, the uniqueness assertion yields $\langle X_n\rangle=\langle X_m\rangle$ on $[0,\tau_m]$. As $\Omega=\bigcup\{\tau_k=T\}$, we find a continuous, increasing process $\langle X\rangle$ with $\langle X\rangle=\langle X_n\rangle$ on $\{\tau_n=T\}$. Thus $(X^2-\langle X\rangle)(t\wedge\tau_n)=X_n^2-\langle X_n\rangle$. As the latter are martingales, it follows that $X^2-\langle X\rangle$ is a local martingale.

We leave it to the reader to prove the assertion about convergence of V_{π_n} to $\langle X\rangle$.

Step 4: We prove the addendum concerning $X\in\mathsf{M}_2([0,T])$.

Let τ_n be as in Step 3. By monotone convergence,

$$\mathbb{E}\langle X\rangle_T=\lim_{n\to\infty}\mathbb{E}\langle X\rangle_{\tau_n}=\lim_{n\to\infty}\mathbb{E}X(\tau_n)^2\geq\mathbb{E}X(T)^2\,.$$

Here, we have used that $X(\cdot\wedge\tau_n)^2-\langle X_n\rangle=X(\cdot\wedge\tau_n)^2-\langle X\rangle_{\cdot\wedge\tau_n}$ is a martingale with expectation 0 in the second equality and Fatou's lemma in the last inequality. Since X is a martingale, X^2 is a submartingale whence, by optional sampling $\mathbb{E}X(\tau_n)^2\leq\mathbb{E}X(T)^2$. It follows that $\mathbb{E}\langle X\rangle_T=\mathbb{E}X(T)^2$. In particular, $X^2-\langle X\rangle_T$ is integrable. However, a continuous local martingale $(M(t))_{t\in[0,T]}$ such that $M(T)$ is integrable is a continuous martingale, which follows immediately from Corollary 2.1.9 as in the proof of Proposition 2.1.11. □

LEMMA 2.2.7. *In the situation of Theorem 2.2.5, if X is uniformly bounded, say by M, then $V_{\pi_n}(T)$ is a Cauchy sequence in $L^2(\Omega)$.*

PROOF. We let $m \geq n$ and denote $\pi_n = (0 = t_0 < t_1 < \cdots < t_n = T)$ and $\pi_m = (0 = s_0 < \cdots < s_N = T)$. Then $\pi_n \subset \pi_m$. Let us moreover define $\tilde{X}(t) := X(t_{k-1})$ for $t_{k-1} \leq t < t_k$.

Using (2.2), we see that

$$V_{\pi_n}(T) - V_{\pi_m}(T) = 2\sum_{j=1}^{N} \big[X(s_{j-1}) - \tilde{X}(s_{j-1})\big](X(s_j) - X(s_{j-1}))\,.$$

Thus,

$$\mathbb{E}|V_{\pi_n}(T) - V_{\pi_m}(T)|^2 = 4\mathbb{E}\sum_{j=1}^{N}(X(s_{j-1}) - \tilde{X}(s_{j-1}))^2(X(s_j) - X(s_{j-1}))^2\,,$$

noting that all mixed terms vanish by the martingale property, cf. the proof of (2.3).

Next put $\Delta(n,m) = \sup_j\{|X(s_{j-1}) - \tilde{X}(s_{j-1})|^2\}$ and note that $\Delta(n,m) \to 0$ as $n, m \to \infty$ pointwise almost surely by continuity of the paths. Since $|\Delta(n,m)| \leq 4M^2$, it follows that $\Delta(n,m) \to 0$ in $L^2(\Omega)$. By the Cauchy-Schwarz inequality,

$$\mathbb{E}|V_{\pi_n}(T) - V_{\pi_m}(T)|^2 \leq \|\Delta(n,m)\|_2 \Big\|\sum_{j=1}^{N}(X(s_j) - X(s_{j-1})^2\Big\|_2\,.$$

Hence, to finish the proof it remains to show that $\big\|\sum_{j=1}^{N}(X(s_j) - X(s_{j-1}))^2\big\|_2$ is uniformly bounded, independently of n and m.

To that end, for ease of notation, let us write $X_j := X(s_j)$. Then

$$\begin{aligned}
\mathbb{E}\Big(\sum_{j=1}^{N}(X_j - X_{j-1})^2\Big)^2 &= \mathbb{E}\sum_{j=1}^{n}(X_j - X_{j-1})^4 + 2\mathbb{E}\sum_{j=1}^{n}\sum_{l=j+1}^{N}(X_j - X_{j-1})^2(X_l - X_{l-1})^2 \\
&\leq 2M^2\mathbb{E}\sum_{j=1}^{N}(X_j - X_{j-1})^2 + 2\mathbb{E}\sum_{j=1}^{N}(X_j - X_{j-1})^2\sum_{l=j+1}^{N}\mathbb{E}\big[(X_l - X_{l-1})^2|\mathscr{F}_{s_{l-1}}\big] \\
&= 2M^2\mathbb{E}X_N^2 + 2\mathbb{E}\sum_{j=1}^{N}(X_j - X_{j-1})^2\sum_{l=j+1}^{N}X_l^2 - X_{l-1}^2 \quad \text{by (2.3)} \\
&= 2M^2\mathbb{E}X_N^2 + 2\mathbb{E}\sum_{j=1}^{N}(X_j - X_{j-1})^2(X_N^2 - X_j^2) \\
&\leq 2M^2\mathbb{E}X_N^2 + 4M^2\mathbb{E}\sum_{j=1}^{N}(X_j - X_{j-1})^2 \\
&= 6M^2\mathbb{E}X_N^2\,.
\end{aligned}$$

Here, we have used that $\mathbb{E}\sum_{j=1}^{N}(X_j - X_{j-1})^2 = \mathbb{E}X_N^2 = \mathbb{E}X(T)^2$. □

2.3. Covariation

We now extend the quadratic variation to products of processes via "polarization". Note that if $X, Y \in \mathsf{M}_{\text{loc}}([0,T])$, then both $(X+Y)^2 - \langle X+Y\rangle$ and $(X-Y)^2 - \langle X-Y\rangle$ are local martingales. Thus, so is their difference $4XY - \langle X+Y\rangle + \langle X-Y\rangle$

DEFINITION 2.3.1. For $X, Y \in \mathsf{M}_{\text{loc}}([0,T])$, we put

$$\langle X, Y\rangle := \frac{1}{4}(\langle X+Y\rangle - \langle X-Y\rangle)$$

Note that $\langle X,Y\rangle$ has paths of bounded variation, whence Lemma 2.2.4 shows that $\langle X,Y\rangle$ is the unique continuous, adapted process with paths of bounded variation such that $XY - \langle X,Y\rangle$ is a local martingale.

We start with a result about stopping covariations.

LEMMA 2.3.2. *Let $X,Y \in \mathsf{M}_{\mathrm{loc}}([0,T])$ and τ be a stopping time. Then*

$$\langle X,Y\rangle^\tau = \langle X^\tau, Y^\tau\rangle = \langle X, Y^\tau\rangle = \langle X^\tau, Y\rangle\,.$$

PROOF. Let us first assume that $X,Y \in \mathsf{M}_2([0,T])$.

By assumption, $\langle X,Y\rangle$ has paths of bounded variation and $XY - \langle X,Y\rangle$ is a local martingale. By optional sampling, $X^\tau Y^\tau - \langle X,Y\rangle^\tau$ is a local martingale. As $\langle X,Y\rangle^\tau$ has paths of bounded variation, $\langle X^\tau,Y^\tau\rangle = \langle X,Y\rangle^\tau$.

For the second equality (the last is proved similarly) it suffices to show that $(X-X^\tau)Y^\tau$ is a local martingale, for in this case $XY^\tau - \langle X,Y\rangle^\tau = X^\tau Y^\tau - \langle X,Y\rangle^\tau + (X-X^\tau)Y^\tau$ is also a local martingale, whence $\langle X,Y^\tau\rangle = \langle X,Y\rangle^\tau$.

To see that $(X-X^\tau)Y^\tau$ is a martingale, first let τ take only finitely many values and let σ be a stopping time which takes at most two values. Then

$$\mathbb{E}((X(\sigma) - X^\tau(\sigma))Y^\tau(\sigma)) = \mathbb{E}\big(Y(\tau\wedge\sigma)\mathbb{E}[X(\sigma) - X(\tau\wedge\sigma)|\mathscr{F}_{\tau\wedge\sigma}]\big) = 0$$

since, by optional sampling $\mathbb{E}[X(\sigma)-X(\tau\wedge\sigma)|\mathscr{F}_{\tau\wedge\sigma}] = 0$. Thus $(X-X^\tau)Y^\tau$ is a martingale by Proposition 2.1.6. By approximation arguments, using that the processes have continuous paths, it follows that the same is true for arbitrary stopping times τ.

In the general case of local martingales, one simply replaces τ with $\tau\wedge\sigma_n\wedge\rho_n$, where σ_n is a sequence of stopping times such that $\sigma_n \uparrow T$ almost surely such that $X^{\sigma_n} \in \mathsf{M}_2([0,T])$ and ρ_n is a corresponding sequence for Y. ☐

The covariation bracket $\langle\cdot,\cdot\rangle$ shares many properties of an inner product.

LEMMA 2.3.3. *For $X,Y,Z \in \mathsf{M}_{\mathrm{loc}}([0,T])$ and $\alpha,\beta\in\mathbb{R}$, we have*

(1) $\langle \alpha X + \beta Y, Z\rangle = \alpha\langle X,Z\rangle + \beta\langle Y,Z\rangle$.
(2) $\langle X,Y\rangle = \langle Y,X\rangle$.
(3) $|\langle X,Y\rangle|^2 \le \langle X\rangle\langle Y\rangle$.

PROOF. Clearly, $\alpha\langle X,Z\rangle + \beta\langle Y,Z\rangle$ is an adapted finite variation process. Moreover

$$(\alpha X+\beta Y)Z - \alpha\langle X,Z\rangle - \beta\langle Y,Z\rangle = \alpha(XZ - \langle X,Z\rangle) + \beta(YZ - \langle Y,Z\rangle)$$

is a local martingale, whence (1). Similarly, one proves (2). As for (3), first note that $\langle X,X\rangle = \frac{1}{4}\langle 2X\rangle\langle X\rangle \ge 0$ almost surely, as the latter is an increasing process starting a.s. at 0. Hence, for $\lambda\in\mathbb{R}$, by (1)

$$0 \le \langle X+\lambda Y\rangle = \langle X\rangle + 2\lambda\langle X,Y\rangle + \lambda^2\langle Y\rangle$$

almost surely. Fixing versions of these processes, the exceptional set can be chosen independently of t and λ. For $t\in[0,T]$ put $\lambda_t := -\langle X,Y\rangle_t\langle Y\rangle_t^{-1}$, provided $\langle Y\rangle_t > 0$. Then, by the above

$$0 \le \langle X\rangle_t - \frac{\langle X,Y\rangle_t^2}{\langle X\rangle_t} \quad \text{that is} \quad \langle X,Y\rangle_t^2 \le \langle X\rangle_t\langle Y\rangle_t$$

outside our fixed set of measure zero. If, on the other hand $\langle Y\rangle_t = 0$, picking $\lambda_t = -\frac{1}{2}\mathrm{sgn}\langle X,Y\rangle_t s$, for $s>0$, we find $|\langle X,Y\rangle_t| \le s^{-1}\langle X\rangle_t \to 0$ as $s\to\infty$, whence also in this case the claim holds. ☐

2.4. Exercises

(1) Fill in the gaps in the proofs of Lemma 2.1.4 and Theorem 2.2.5 that you find worthwhile to fill.

(2) Besides Brownian motion, there is another "basic" example of a martingale, namely the *compensated Poisson process*.

A Poisson process with intensity $\lambda > 0$ is an adapted, integer-valued process $(N(t))_{t\geq 0}$ which has paths which are right continuous with left limits such that $N(0) = 0$ almost surely and for $0 \leq s < t$, the random variable $N(t) - N(s)$ is independent of $\mathscr{F}_s$ and has Poisson distribution with mean $\lambda(t-s)$. Recall that the Poisson distribution with mean λ is the measure $e^{-\lambda}\sum_{k=0}^\infty \frac{\lambda^k}{k!}\delta_k$.

Show that the *compensated Poisson process* $N(t)-\lambda t$ is a martingale. Moreover, show that $(N(t)-\lambda t)^2 - \lambda t$ is also a martingale. Thus, in a way, λt is the quadratic variation of the compensated Poisson process.

(3) Let $W(t)$ be a Wiener process $\tau = \inf\{t > 0 : W(t) \leq -1\}$ and define

$$X(t) := \begin{cases} W(\frac{t}{1-t} \wedge \tau) & t < 1 \\ -1 & t \geq 1\,. \end{cases}$$

Then $(X(t))_{t\geq 0}$ is a continuous (this uses the fact that $\tau < \infty$ almost surely) process. Show that it is a local martingale which is not a martingale.

Hint: Use $\tau_n := \inf\{t : X(t) \geq n\} \wedge n$.

(4) A 2-dimensional Brownian motion with covariance matrix $Q \in \mathbb{R}^{2\times 2}$, which is assumed to by symmetric and positive semidefinite, is a stochastic process $W(t) = (W_1(t), W_2(t)) : \Omega \to \mathbb{R}^d$, defined on a filtered probability space $(\Omega, \Sigma, \mathbb{F}, \mathbb{P})$ such that $W(0) = 0$ almost surely and, for $t, s \geq 0$, the increment $W(t+s) - W(s)$ is independent of $\mathscr{F}_s$ and is normally distributed with mean $(0,0)$ and covariance sQ.

Given a symmetric, positive semidefinite matrix Q, construct a 2-dimensional Brownian motion with covariance matrix Q and determine $\langle W_1, W_2\rangle$.

Hint: For the construction, start with two independent Brownian motions and diagonalize Q.

CHAPTER 3

Stochastic Calculus

3.1. The Itô Integral

We now define Itô's integral with general continuous local martingales X as integrators. We begin with $X \in \mathsf{M}_2([0,T])$ as integrators and adapted step processes $\phi : \Omega \times [0,T] \to \mathbb{R}$ as integrands. Here an adapted step process ϕ is a process of the form

$$\phi(t,x) = \sum_{k=1}^{n} \eta_k(\omega)\mathbb{1}_{[a_k,b_k)}(t)$$

where $0 \le a_1 < b_1 \le a_2 < b_2 < \cdots \le a_n < b_n$ and $\eta_k \in L^\infty(\Omega, \mathscr{F}_{a_k}, \mathbb{P})$. We define

$$\int_0^t \phi(s)\, dX(s) := \sum_{k=1}^{n} \eta_k(X(b_k \wedge t) - X(a_k \wedge t)) \tag{3.1}$$

and write $\phi \diamond X$ for the integral process, i.e. $(\phi \diamond X)_t := \int_0^t \phi(s)\, dX(s)$.

THEOREM 3.1.1. *Let $X \in \mathsf{M}_2([0,T])$ and ϕ be an adapted step process. Then $\phi \diamond X \in M_2([0,T])$, its quadratic variation is given by*

$$\langle \phi \diamond X\rangle_t = \int_0^t |\phi(s)|^2\, d\langle X\rangle_s := \sum_{k=1}^{n} \eta_k^2\big(\langle X\rangle_{b_k\wedge t}) - \langle X\rangle_{a_k\wedge t}\big).$$

In particular, we have the Itô isometry

$$\mathbb{E}\Big|\int_0^t \phi(s)\, dX(s)\Big|^2 = \mathbb{E}\int_0^t |\phi(s)|^2\, d\langle X\rangle_s\,. \tag{3.2}$$

Finally, if $Y \in \mathsf{M}_2([0,T])$, then

$$\langle \phi \diamond X, Y\rangle_t = \int_0^t \phi d\langle X, Y\rangle_t$$

PROOF. It is clear that $\phi \diamond X$ is a continuous, adapted process with $(\phi \diamond X)_0 = 0$. To show that it is a martingale, we use Proposition 2.1.6 and let a stopping time τ taking at most two values be given. Then

$$\mathbb{E}(\phi\diamond X)_\tau = \mathbb{E}\sum_{k=1}^{n} \eta_k(X^\tau(b_k) - X^\tau(a_k)) = \mathbb{E}\sum_{k=1}^{n} \eta_k\mathbb{E}\big[X^\tau(b_k) - X^\tau(a_k)\big|\mathscr{F}_{a_k}\big] = 0 = \mathbb{E}(\phi\diamond X)_0$$

since X, hence X^τ is a martingale.

For the second part, since $t \mapsto \int_0^t |\phi(s)|^2\, d\langle X\rangle_s$ is an adapted, increasing, continuous process, it is enough to show that $(\phi \diamond X)_t^2 - \int_0^t |\phi(s)|^2\, d\langle X\rangle_s$ is a martingale.

To that end, we again consider a stopping time τ taking at most two values and note that

$$(\phi \diamond X)_\tau^2 = \sum_{k,l=1}^{n} \eta_k\eta_l(X^\tau(b_k) - X^\tau(a_k))(X^\tau(b_l) - X^\tau(a_l))$$

and

$$\int_0^\tau |\phi(s)|^2\,d\langle X\rangle_s = \sum_{k=1}^n \eta_k^2(\langle X\rangle_{b_k\wedge\tau} - \langle X\rangle_{a_k\wedge\tau}).$$

We now obtain

$$\begin{aligned}
\mathbb{E}(\phi\diamond X)_\tau^2 &= \mathbb{E}\sum_{k=1}^n \eta_k^2(X^\tau(b_k)-X^\tau(a_k))^2\\
&\quad\text{as the mixed terms vanish, due to the martingale property.}\\
&= \mathbb{E}\sum_{k=1}^n \eta_k^2\mathbb{E}\big[X^\tau(b_k)^2 - X^\tau(a_k)^2|\mathscr{F}_{a_k}\big]\\
&\quad\text{as } \mathbb{E}[X^\tau(b_k)X^\tau(a_k)|\mathscr{F}_{a_k}] = X^\tau(a_k)^2\\
&= \mathbb{E}\sum_{k=1}^n \eta_k^2(\langle X\rangle_{b_k\wedge\tau} - \langle X\rangle_{a_k\wedge\tau})\\
&\quad\text{since } X^2 - \langle X\rangle \text{ is a martingale}\\
&= \mathbb{E}\int_0^\tau |\phi(s)|^2\,d\langle X\rangle_s
\end{aligned}$$

Consequently, by Proposition 2.1.6 $(\phi\diamond X)_t^2 - \int_0^t |\phi(s)|^2\,d\langle X\rangle_s$ is a martingale.

The proof of the last assertion is similar. □

We now extend the stochastic integral to more integrands and to more integrators. First, let us stick with $X\in \mathsf{M}_2([0,T])$ and allow more integrands.

To that end, first note that for fixed ω, there is a unique Lebesgue-Stieltjes measure $\nu_{\langle X\rangle(\omega)}$ on $[0,T]$ with $\nu_{\langle X\rangle(\omega)}([a,b)) = \langle X\rangle_b(\omega) - \langle X\rangle_a(\omega)$. Moreover, using the measurability properties of the quadratic variation, a monotone class argument shows that $\omega\mapsto \nu_{\langle X\rangle(\omega)}(A)$ is measurable for all $A\in\mathscr{B}([0,T])$. We may thus define a measure $\mu_{\langle X\rangle}$ on $(\Omega\times[0,T],\Sigma\otimes\mathscr{B}([0,T]))$ by

$$(\mu_{\langle X\rangle})(A) := \int_\Omega\int_0^T \mathbb{1}_A(\omega,t)\,d\nu_{\langle X\rangle(\omega)}(t)\,d\mathbb{P}(\omega).$$

Note that $\mu_{\langle X\rangle}$ is *not* a probability measure! For an adapted step process ϕ, we have $\int_{\Omega\times[0,T]}|\phi|^2\,d\mu_{\langle X\rangle} = \mathbb{E}\int_0^T|\phi(t)|^2\,d\langle X\rangle_t$. We denote by $L^2_{\mathbb{F},X}(\Omega\times[0,T])$ the closure of the adapted step processes in $L^2(\Omega\times[0,T],\mu_{\langle X\rangle})$. We can now extend the Itô integral uniquely to $L^2_{\mathbb{F},X}(\Omega\times[0,T])$.

Given $\phi\in L^2_{\mathbb{F},X}(\Omega\times[0,T])$ there is a sequence ϕ_n of adapted step processes such that $\phi_n\to\phi$ in $L^2(\Omega\times[0,T],\mu_{\langle X\rangle})$. By Doob's maximal inequality and the Itô isometry 3.2,

$$\mathbb{E}\|\phi_n\diamond X - \phi_m\diamond X\|^2_{C([0,T])} \le 4\mathbb{E}|(\phi_n\diamond X)_T - (\phi_m\diamond X)_T|^2 \le 4\|\phi_n-\phi_m\|_{L^2(\Omega\times[0,T],\mu_{\langle X\rangle})}\to 0$$

as $n,m\to\infty$. Thus, $\phi_n\diamond X$ is a Cauchy sequence in $\mathsf{M}_2([0,T])$. Its limit does not depend on the approximating sequence ϕ_n. Indeed, if $\tilde\phi_n$ is another sequence of adapted step processes with $\tilde\phi_n\to\phi$ in $L^2(\Omega\times[0,T],\mu_{\langle X\rangle})$, then

$$\mathbb{E}\|\phi_n\diamond X - \tilde\phi_n\diamond X\|^2_{C([0,T])} \le 4\|\phi_n-\tilde\phi_n\|^2_{L^2(\Omega\times[0,T],\mu_{\langle X\rangle})}\to 0\,.$$

Hence, we may define

Definition 3.1.2. Let $X\in\mathsf{M}_2([0,T])$ and $\phi\in L^2_{\mathbb{F},X}(\Omega\times[0,T])$. The stochastic integral $\phi\diamond X$ is the unique process $Y\in\mathsf{M}_2([0,T])$ such that for every sequence ϕ_n of adapted step processes with $\phi_n\to\phi$ in $L^2_{\mathbb{F},X}(\Omega\times[0,T])$ we have $\phi_n\diamond X\to Y$ in $L^2(\Omega;C([0,T]))$.

REMARK 3.1.3. It follows from an approximation argument that

$$\langle \phi \diamond X \rangle_t = \int_0^t |\phi(s)|^2 \, d\langle X \rangle_s$$

where the integral $\int_0^t |\phi(s)|^2 \, d\langle X \rangle_s$ is defined as integral with respect to the Lebesgue-Stieltjes measure μ_ω. Note this integral is defined pathwise, hence we have to pick representatives of ϕ and $\langle X \rangle$. However, for different choices of representatives, the resulting process differ (as $C([0,T])$-valued random elements) only on a null set. Thus, the right-hand side defines a process in $L^0(\Omega; C([0,T]))$.

To extend the stochastic integral even further, it is useful to first establish a characterization of the integral in terms of covariation processes. We have

THEOREM 3.1.4. *Let $X \in \mathsf{M}_2([0,T])$ and $\phi \in L^2_{\mathbb{F},X}(\Omega \times [0,T])$. The stochastic integral $\phi \diamond X$ is the unique process $Y \in \mathsf{M}_2([0,T])$ such that*

$$\langle Y, Z \rangle_t = \int_0^t \phi(s) d\langle X, Z \rangle_s \tag{3.3}$$

for all $Z \in \mathsf{M}_2([0,T])$.

As a technical tool for the proof, we need the Kunita-Watanabe inequality.

THEOREM 3.1.5. *(Kunita-Watanabe)*
Let $X, Y \in \mathsf{M}_2([0,T])$, $\phi \in L^2_{\mathbb{F},X}(\Omega \times [0,T])$ and $\psi \in L^2_{\mathbb{F},Y}(\Omega \times [0,T])$. Then, almost surely,

$$\int_0^t |\phi(s)\psi(s)||d\langle X, Y \rangle_s| \leq \Big(\int_0^t |\phi(s)|^2 \, d\langle X \rangle_s \Big)^{\frac{1}{2}} \Big(\int_0^t |\psi(s)|^2 \, d\langle Y \rangle_s \Big)^{\frac{1}{2}}$$

for all $t \in [0,T]$. Here, $|d\langle X, Y \rangle|$ is the total variation of the measure $d\langle X, Y \rangle$.

PROOF. Let us first make some simplifying assumptions. First note that the integrals above are pathwise continuous as functions of t. Hence, it suffices to prove that the claimed inequality holds for every $t \in [0,T]$ almost surely (i.e. with exceptional set possibly depending on t). We fix $t \in [0,T]$.

It is enough to prove that

$$\Big| \int_0^t \phi(s)\psi(s) d\langle X, Y \rangle_s \Big| \leq \Big(\int_0^t |\phi(s)|^2 \, d\langle X \rangle_s \Big)^{\frac{1}{2}} \Big(\int_0^t |\psi(s)|^2 \, d\langle Y \rangle_s \Big)^{\frac{1}{2}},$$

for, if $\rho(s)$ is a density of $|d\langle X, Y \rangle|$ with respect to $d\langle X, Y \rangle$ with values in $\{-1, 1\}$ and we replace ϕ with $\tilde{\phi} = \phi\rho\,\mathrm{sgn}(\phi\psi)$, then the above with $\tilde{\phi}$ and ψ gives the claim for ϕ and ψ. We leave it to the reader to verify that $\tilde{\phi} \in L^2_{\mathbb{F},X}(\Omega \times [0,T])$.

Given ϕ and ψ as in the assumption, we find sequences of elementary step functions ϕ_n and ψ_n converging to ϕ, resp. ψ in $L^2_{\mathbb{F},X}(\Omega \times [0,T])$ resp. $L^2_{\mathbb{F},Y}(\Omega \times [0,T])$. In particular, $\int_0^t |\phi_n(s)|^2 \, d\langle X \rangle_s \to \int_0^t |\phi(s)|^2 \, d\langle X \rangle_s$ in $L^1(\Omega)$. And similarly for ψ. Passing to suitable subsequences, we may assume that we have convergence almost everywhere both for the processes and the integrals.

Thus, if the result holds for Step functions we can infer from Fatou's Lemma the general result.

For the rest of the proof, we assume that ϕ and ψ are elementary step functions. To further simplify, we may assume without loss of generality that ϕ and ψ are based on a common partition $0 = t_0 < t_1 < \cdots < t_n$. Finally, to simplify notation, we assume that $t = t_n$; this is no loss of generality as for $t < t_n$, we may simply replace t_n with t, if $t > t_n$, we add in $t_{n+1} = t$ and set ϕ, ψ zero on the last interval.

We thus assume that

$$\phi(t) = \sum_{k=1}^{n} \eta_k \mathbb{1}_{[t_{k-1},t_k)}(t) \quad \text{and} \quad \psi(t) = \sum_{k=1}^{n} \xi_k \mathbb{1}_{[t_{k-1},t_k)}(t)\,.$$

Next observe that, putting $\langle X, Y\rangle_s^r := \langle X, Y\rangle_r - \langle X, Y\rangle_s$ for $s \le t$, we have $|\langle X, Y\rangle_s^r| \le \big(\langle X, X\rangle_s^r\big)^{\frac{1}{2}}\big(\langle Y, Y\rangle_s^r\big)^{\frac{1}{2}}$. Indeed, for $\lambda \in \mathbb{R}$ the quantity

$$\langle X, X\rangle_s^r + 2\lambda\langle X, Y\rangle_s^r + \lambda^2\langle Y, Y\rangle_s^r = \langle X + \lambda Y, X + \lambda Y\rangle_s^r$$

is almost surely nonnegative for $s \le r$ and the proof can be finished as in that of Lemma 2.3.3.

Consequently,

$$\begin{aligned}
\Big|\int_0^t \phi(s)\psi(s)d\langle X, Y\rangle_s\Big| &\le \sum_{k=1}^{n} |\eta_k\xi_k||\langle X, Y\rangle_{t_{k-1}}^{t_k}| \\
&\le \sum_{k=1}^{n} |\eta_k\xi_k|\big(\langle X, X\rangle_{t_{k-1}}^{t_k}\big)^{\frac{1}{2}}\big(\langle Y, Y\rangle_{t_{k-1}}^{t_k}\big)^{\frac{1}{2}} \\
&\le \Big(\sum_{k=1}^{n} |\eta_k|^2\langle X, X\rangle_{t_{k-1}}^{t_k}\Big)^{\frac{1}{2}}\Big(\sum_{k=1}^{n} |\xi_k|^2\langle Y, Y\rangle_{t_{k-1}}^{t_k}\Big)^{\frac{1}{2}} \\
&= \Big(\int_0^t |\phi(s)|^2\, d\langle X\rangle_s\Big)^{\frac{1}{2}}\Big(\int_0^t |\psi(s)|^2\, d\langle Y\rangle_s\Big)^{\frac{1}{2}},
\end{aligned}$$

where we have used the Cauchy-Schwarz inequality. □

Proof of Theorem 3.1.4. Clearly, there can be at most one $Y \in \mathsf{M}_2([0, T])$ which satisfies (3.3). Indeed, if (3.3) holds with $Y = Y_1$ and $Y = Y_2$, then $\langle Y_1 - Y_2, Z\rangle = 0$ for all $Z \in \mathsf{M}_2([0, T])$. Picking $Z = Y_1 - Y_2$, we find $\langle Y_1 - Y_2\rangle = 0$, hence $Y_1 - Y_2 = 0$ as then $(Y_1 - Y_2)^2$ is a nonnegative martingale starting at 0.

It remains to verify that (3.3) holds for $Y = \phi \diamond X$. To that end, let ϕ_n be a sequence of elementary step processes converging to ϕ in $L^2_{\mathbb{F},X}(\Omega \times [0, T])$. Using Lemma 2.3.3, we obtain

$$|\langle \phi_n \diamond X - \phi \diamond X, Z\rangle_t|^2 \le \langle \phi_n \diamond X - \phi \diamond X\rangle_t\langle Z\rangle_t \le \langle \phi_n \diamond X - \phi \diamond X\rangle_T\langle Z\rangle_T$$

Thus taking the supremum over $t \in [0, T]$ and using Cauchy-Schwarz and the Itô-isometry, we find

$$\mathbb{E}\|\langle \phi_n \diamond X - \phi \diamond X, Z\rangle\|_\infty^2 \le \mathbb{E}\langle \phi_n \diamond X - \phi \diamond X\rangle_T\langle Z\rangle_T \le \big(\mathbb{E}\langle Z\rangle_T^2\big)^{\frac{1}{2}}\|\phi_n - \phi\|_{L^2_{\mathbb{F},X}(\Omega\times[0,T])} \to 0$$

as $n \to \infty$. Consequently, $\langle \phi_n \diamond X, Z\rangle \to \langle \phi \diamond X, Z\rangle$ in $L^2(\Omega; C([0, T]))$. Passing to a subsequence, we may assume that we have convergence pointwise almost everywhere. It follows from the Kunita-Watanabe inequality that

$$\Big|\int_0^t \phi_n\, d\langle X, Z\rangle - \int_0^t \phi\, d\langle X, Z\rangle\Big| \le \Big(\int_0^t |\phi_n - \phi|^2\, d\langle X\rangle\Big)^{\frac{1}{2}}\langle Z\rangle \to 0$$

Since the result is true for elementary simple functions, we have

$$\langle \phi \diamond X, Z\rangle = \lim \langle \phi_n \diamond X, Z\rangle = \lim \int_0^\cdot \phi_n\, d\langle X, Z\rangle = \int_0^\cdot \phi\, d\langle X, Z\rangle\,. \qquad \square$$

Corollary 3.1.6. *Let $X \in \mathsf{M}_2([0, T])$ and $\phi \in L^2_{\mathbb{F},X}(\Omega \times [0, T])$. Moreover, let τ be a stopping time. Then*

$$(\phi \diamond X)^\tau = \phi \diamond (X^\tau) = (\phi\mathbb{1}_{[0,\tau]}) \diamond X = (\phi\mathbb{1}_{[0,\tau]}) \diamond (X^\tau)$$

almost surely, where we have written Y^τ for the process $Y(\cdot \wedge \tau)$.

PROOF. Let $Y \in \mathsf{M}_2([0,T])$. Then we have

$$\int_0^t \phi(s)\mathbb{1}_{[0,\tau]}(s)\, d\langle X,Y\rangle = \int_0^{t\wedge\tau} \phi(s)\, d\langle X,Y\rangle_s = \int_0^t \phi(s)\, d\langle X,Y\rangle_{s\wedge\tau}$$

almost surely. Thus the claim follows from the characterization in Theorem 3.1.4, noting that $\langle X,Y\rangle^\tau = \langle X^\tau, Y^\tau\rangle = \langle X, Y^\tau\rangle = \langle X^\tau, Y\rangle$ by Lemma 2.3.2 □

We now obtain the following corollary, which will allow us to considerably extend the stochastic integral.

COROLLARY 3.1.7. *Let $X,Y \in \mathsf{M}_2([0,T])$, $\phi \in L^2_{\mathbb{F},X}(\Omega\times[0,T])$ and $\psi \in L^2_{\mathbb{F},Y}(\Omega\times[0,T])$. Moreover, let τ be a stopping time such that*

$$X^\tau = Y^\tau \quad \textit{and} \quad \phi\mathbb{1}_{[0,\tau]} = \psi\mathbb{1}_{[0,\tau]}$$

almost surely. Then $(\phi\diamond X)^\tau = (\psi\diamond Y)^\tau$ almost surely.

PROOF. For $Z \in \mathsf{M}_2([0,T])$, we have

$$\langle(\phi\diamond X)^\tau - (\psi\diamond Y)^\tau, Z\rangle = \big\langle \phi\mathbb{1}_{[0,\tau]}\diamond X^\tau - \psi\mathbb{1}_{[0,\tau]}\diamond Y^\tau, Z\big\rangle = 0$$

almost surely. Choosing $Z = (\phi\diamond X)^\tau - (\psi\diamond Y)^\tau$, the claim follows from Lemma 2.2.4. □

We can now extend the stochastic integral to its full generality.

DEFINITION 3.1.8. Let $X \in \mathsf{M}_{\mathrm{loc}}([0,T])$. By $L^0_{\mathbb{F},X}(\Omega\times[0,T])$ we denote the closure of the elementary step processes in $L^0(\Omega\times[0,T],\mu_{\langle X\rangle})$. By $\mathscr{I}(X)$, the *processes which are integrable with respect to X*, we denote the set of all $\phi \in L^0_{\mathbb{F},X}(\Omega\times[0,T])$ such that

$$\mathbb{P}\Big(\int_0^T |\phi(s)|^2\, d\langle X\rangle_s < \infty\Big) = 1\,.$$

Since $X \in \mathsf{M}_{\mathrm{loc}}([0,T])$, there exists a sequence of stopping times τ_n with $\tau_n \uparrow T$ almost surely such that $X^{\tau_n} \in \mathsf{M}_2([0,T])$. Now put

$$\sigma_n := n \wedge \inf\{t\in[0,T] \,:\, \int_0^t |\phi(s)|^2\, d\langle X\rangle(\omega) \geq n\}\,.$$

Then also σ_n is a stopping time. We set $\rho_n = \sigma_n\wedge\tau_n$. In this case, $(\phi\mathbb{1}_{[0,\rho_n]})\diamond X^{\rho_n}$ is well-defined. By Corollary 3.1.7, we also have

$$(\phi\mathbb{1}_{[0,\rho_n]})\diamond X^{\rho_n} = (\phi\mathbb{1}_{[0,\rho_m]})\diamond X^{\rho_m}$$

on $[0,\rho_n]$, for all $1\leq n\leq m$. We may thus define

$$\phi\diamond X := (\phi\mathbb{1}_{[0,\rho_n]})\diamond X^{\rho_n} \quad \text{on } [0,\rho_n]\,.$$

It also follows from Corollary 3.1.7 that this definition does not depend on the particular choice of the τ_n. Clearly, $\phi\diamond X$ is a local martingale and it is easy to see that $\langle\phi\diamond X, Y\rangle = \int_0^\cdot \phi(s)\, d\langle X,Y\rangle_s$ for all $Y \in \mathsf{M}_{\mathrm{loc}}([0,T])$. Moreover, localizing, it follows from Theorem 3.1.4 that this characterizes the stochastic integral.

DEFINITION 3.1.9. For $X \in \mathsf{M}_{\mathrm{loc}}([0,T])$ and $\phi \in \mathscr{I}(X)$, the stochastic integral $\phi\diamond X$ is the unique process $Y \in \mathsf{M}_{\mathrm{loc}}([0,T])$ such that for all $Z \in \mathsf{M}_{\mathrm{loc}}([0,T])$ we have

$$\langle Y,Z\rangle = \int_0^\cdot \phi(s)\, d\langle X,Z\rangle_s$$

almost surely

3.2. Itô's Formula

We now extend stochastic integration even further.

Definition 3.2.1. A *(continuous) semimartingale* is an adapted process $(X(t))_{t\in[0,T]}$ which has a decomposition

$$X(t) = X_0 + B(t) + M(t)$$

where $X_0 = X(0)$, $(B(t))_{t\in[0,T]}$ is a continuous, adapted process with pathwise bounded variation and $M(t) \in \mathsf{M}_{\mathrm{loc}}([0,T])$.

Remark 3.2.2. We note that the decomposition of a semimartingale into a bounded variation process and a local martingale is almost surely unique. Indeed, if $X(t) = X_0 + B(t) + M(t) = X_0 + \tilde{B}(t) + \tilde{M}(t)$, then $B - \tilde{B} = M - \tilde{M}$ is a continuous local martingale which has pathwise bounded variation. Consequently, by Lemma 2.2.4, $B - \tilde{B} = 0$ almost surely, hence also $M = \tilde{M}$ almost surely.

Definition 3.2.3. Let $X = X_0 + B + M$ be a continuous semimartingale. By $\mathscr{I}(B)$, we denote space of all adapted processes ϕ such that, almost surely $\phi(\omega) \in L^1([0,T], dB(\omega))$, where $dB(\omega)$ is the Lebesgue-Stieltjes measure induced by $B(\omega)$. We then write $\mathscr{I}(X) = \mathscr{I}(B) \cap \mathscr{I}(M)$ and define for $\phi \in \mathscr{I}(X)$

$$(\phi \diamond X)(t) := \int_0^t \phi(s)\, dB(s) + \int_0^t \phi(s)\, dM(s)$$

where the first integral is defined pathwise and the second is the stochastic integral as before.

Remark 3.2.4. With this definition, some results from the previous section can now be reformulated.

For example, for a local martingale X the stochastic integral $\phi \diamond X$ is the unique local martingale such that $\langle \phi \diamond X, Y\rangle = \phi \diamond \langle X, Y\rangle$ for all local martingales Y. Moreover, $\langle \phi \diamond X\rangle = \phi^2 \diamond \langle X\rangle$. In both equalities, the "$\diamond$" on the left hand side refers to the stochastic integral introduced in the previous section and the "$\diamond$" on the right hand side refers to a pathwise integral with respect to a bounded variation process.

Remark 3.2.5. Note that for a continuous semimartingale $X = X_0 + B + M$ and $\phi \in \mathscr{I}(X)$ also $\phi \diamond X$ is a continuous semimartingale. Indeed, $\phi \diamond M$ is again a local martingale and $\phi \diamond B$ is again of bounded variation.

Proposition 3.2.6. *(Chain rule)*

Let X be a continuous semimartingale and ϕ, ψ be adapted processes with $\psi \in \mathscr{I}(X)$. Then $\phi \in \mathscr{I}(\psi \diamond X)$ if and only if $\phi\psi \in \mathscr{I}(X)$. In this case, $\phi \diamond (\psi \diamond X) = (\phi\psi) \diamond X$.

Proof. Let $X = X_0 + B + M$ be the canonical decomposition of X. Then $\phi \in \mathscr{I}(\psi \diamond X)$ is equivalent to $\phi^2 \in \mathscr{I}(\langle \psi \diamond M\rangle)$ and $\phi \in \mathscr{I}(\psi \diamond B)$. On the other hand, $\phi\psi \in \mathscr{I}(X)$ if and only if $(\phi\psi)^2 \in \mathscr{I}(\langle M\rangle)$ and $\phi\psi \in \mathscr{I}(B)$. Since $\langle \psi \diamond M\rangle = \psi^2 \diamond \langle M\rangle$, the equivalence of the conditions follows from the property of the Stieltjes integral.

Also by the properties of the Stieltjes integral, $\phi \diamond (\psi \diamond B) = (\phi\psi) \diamond B$. To see the corresponding formula for the integrals involving M, let N be a continuous local martingale. Using the formula for the Stieltjes integral and the local version of Theorem 3.1.4, we find

$$\langle (\phi\psi) \diamond M, N\rangle = (\phi\psi) \diamond \langle M, N\rangle = \phi \diamond (\psi \diamond \langle M, N\rangle) = \phi \diamond \langle \psi \diamond M, N\rangle = \langle \phi \diamond (\psi \diamond M), N\rangle\,.$$

which proves that $\phi \diamond (\psi \diamond M) = (\phi\psi) \diamond M$ by the local version of Theorem 3.1.4. ☐

We extend the definitions of quadratic variation and covariation to arbitrary continuous semimartingales X and Y with decompositions $X = X_0 + B + M$ and $Y = Y_0 + A + N$ by setting $\langle X\rangle := \langle M\rangle$ and $\langle X, Y\rangle := \langle M, N\rangle$. This reflects that processes with bounded variation have quadratic variation 0, see Lemma 2.2.2.

Next we prove the fundamental "integration by parts formula". This will be the key tool to prove Itô's formula. It generalizes Lemma 1.2.7 and Example 1.3.2.

THEOREM 3.2.7. *Let X and Y be continuous semimartingales. Then, almost surely,*

$$XY = X_0Y_0 + X \diamond Y + Y \diamond X + \langle X, Y\rangle. \tag{3.4}$$

PROOF. We may assume that $X = Y$, as the general case will follow from polarization. We may also assume that $X_0 = 0$.

First, let $X = M \in \mathsf{M}_2([0,T])$. Then equation (3.4) reads $M^2 = 2M \diamond M + \langle M\rangle$. Note that for $M = W$, this is exactly what we have proved in Example 1.3.2. Repeating the computations there for general M and using the convergence in Theorem 2.2.5 instead of Lemma 1.3.3, the formula follows. The case where $X = M \in \mathsf{M}_{\mathrm{loc}}([0,T])$ follows from this by localization.

If, on the other hand, $X = B$ is of bounded variation, equation (3.4) reduces to $B^2 = 2B \diamond B$ which follows from the integration by parts formula for Lebesgue-Stieltjes integrals.

In the general case, we have to prove that

$$(B + M)^2 = B^2 + 2BM + M^2 \stackrel{!}{=} 2B \diamond B + 2B \diamond M + 2M \diamond B + 2M \diamond M + \langle M\rangle\,.$$

In view of the formulas for B^2 and M^2 proved so far, it remains to show that $BM = B \diamond M + M \diamond B$. Localizing again, we may assume that B and M are uniformly bounded. Fixing $t \in (0,T]$, we put $t_k := \frac{kt}{n}$ and

$$A_n(s) := \sum_{k=1}^{n} A(t_{k-1})\mathbb{1}_{[t_{k-1},t_k)}(s) \quad \text{and} \quad M_n(s) := \sum_{k=1}^{n} M(t_k)\mathbb{1}_{[t_{k-1},t_k)}(s)$$

and observe that

$$\begin{aligned}(A_n \diamond M)(t) + (M_n \diamond A)(t) &= \sum_{k=1}^{n} A(t_{k-1})\big[M(t_k - M(t_{k-1}))\big] + M(t_k)\big[A(t_k) - A(t_{k-1})\big]\\ &= A(t)M(t)\end{aligned}$$

by Abelian partial summation. Noting that $A_n \to A$ in $L^2_{\mathbb{F},M}(\Omega \times [0,t])$ by dominated convergence and $M_n \to M$ pointwise and dominated, it follows from continuity of the stochastic integral, resp. dominated convergence for Lebesgue-Stieltjes integrals, that $(A_n \diamond M)(t) \to (A \diamond M)(t)$ and $(M_n \diamond A)(t) \to (M \diamond A)(t)$ in measure. Thus, indeed, $BM = B \diamond M + M \diamond B$, finishing the proof. □

We are now ready to prove Itô's formula.

THEOREM 3.2.8. *Let $X_1, \ldots, X_d$ be continuous semimartingales, $X = (X_1, \ldots, X_d)$. Then, for every $f \in C^2(\mathbb{R}^d)$, $f(X)$ is a continuous semimartingale and*

$$f(X) = f(X_0) + \sum_{j=1}^{d} f_{x_j}(X) \diamond X_j + \frac{1}{2}\sum_{i,j=1}^{d} f_{x_ix_j}(X) \diamond \langle X_i, X_j\rangle\,.$$

PROOF. For notational convenience, we only treat the case $d = 1$, the multidimensional case follows similarly. Thus, fix a continuous semimartingale X. By $\mathscr{C}$ we denote the set of all $f \in C^2(\mathbb{R})$ such that

$$f(X) = f(X_0) + f'(X) \diamond X + \frac{1}{2}f''(X) \diamond \langle X\rangle\,.$$

Clearly, f is a vector space and contains the functions $f_1 \equiv 1$ and $f_2(x) = x$. We prove next that if $f, g \in \mathscr{C}$, then $fg \in \mathscr{C}$, whence $\mathscr{C}$ contains all polynomials.

To see this, let $f, g \in \mathscr{C}$. Then

$$\begin{aligned}(fg)(X) - (fg)(X_0) &= f(X)g(X) - f(X_0)g(X_0)\\ &= f(X) \diamond g(X) + g(X) \diamond f(X) + \langle f(X), g(X)\rangle \quad \text{by Theorem 3.2.7}\\ &= f(X) \diamond \big[g'(X) \diamond X + g''(X) \diamond \langle X\rangle\big]\end{aligned}$$

$$
\begin{aligned}
&+g(x)\diamond\big[f'(X)\diamond X+f''(X)\diamond\langle X\rangle\big]+\langle f'(X)\diamond X,g'(X)\diamond X\rangle\\
&\quad\text{since } f,g\in\mathscr{C}\\
=\;&(fg'+gf')(X)\diamond X+\frac{1}{2}(fg''+2f'g'+f''g)(X)\diamond\langle X\rangle\\
&\quad\text{by Proposition 3.2.6}\\
=\;&(fg)'(X)\diamond X+\frac{1}{2}(fg)''(X)\diamond\langle X\rangle\,.
\end{aligned}
$$

We now extend this to arbitrary $f\in C^2(\mathbb{R})$. By stopping, it suffices to consider semimartingales X which take values in a bounded interval $[a,b]$ of $\mathbb{R}$. As a consequence of the Weierstrass approximation theorem, given $f\in C^2(\mathbb{R})$, we find a sequence of polynomials p_n such that $p_n\to f$, $p_n'\to f'$ and $p_n''\to f''$, uniformly on $[a,b]$. Using the continuity of the stochastic integral and dominated convergence for the pathwise integral, we see that $f\in\mathscr{C}$. □

In its one-dimensional case, Itô's formula asserts that for every semimartingale X, we have

$$f(X(t))=f(X(0))+\int_0^t f'(X(s))\,dX(s)+\frac{1}{2}\int_0^t f''(X(s))\,d\langle X\rangle_s\,.$$

One should compare this formula with the fundamental theorem of calculus, where, for $X(t)=t$, we have $f(t)=f(0)+\int_0^t f'(s)\,ds$. Thus, replacing the deterministic t with a random semimartingale X, we need a correction term $\frac{1}{2}f''(X)\diamond\langle X\rangle$.

Often, and we have seen this already in the formulation of stochastic differential equations, one prefers to write the above in differential form:

$$df(X)=f'(X)dX+\frac{1}{2}f''(X)d\langle X,X\rangle\,.$$

Thinking of $d\langle X,X\rangle$ as $dXdX$, it is suggesting to think of Itô's formula as a second order Taylor expansion:

$$df(X)=f'(X)dX+\frac{1}{2}f''(X)dXdX\,.$$

3.3. First applications of Itô's Formula

3.3.1. Lévy's characterization of Brownian motion. In this section, we show that Brownian motion is the only continuous local martingale with quadratic variation t. This will have appliactions later on, for example in Tanaka's example (Example 5.1.5) or in the integral representation theorem (Theorem 6.1.6).

THEOREM 3.3.1. *Let $X\in\mathsf{M}_{\mathrm{loc}}([0,T])$. Then X is a Brownian motion if and only if $\langle X\rangle_t=t$.*

PROOF. Fix $r\in\mathbb{R}$ and define $f:\mathbb{R}^2\to\mathbb{C}$ by $f(t,x):=e^{irx+\frac{1}{2}r^2t}$. We will use Itô's formula for the semimartingales $X_1(t)=t$ and $X_2(t)=X(t)$. Note that X_1 has bounded variation, hence $\langle X_1\rangle\equiv0$ and $\langle X_1,X_2\rangle=0$. Moreover, we have established Itô's formula merely for real-valued functions. However, we may apply it to the real and imaginary part separately.

The derivatives of f are given by $f_t=\frac{1}{2}r^2f$, $f_x=irf$ and $f_xx=-r^2f$. Note that the derivatives f_{tt} and f_{tx} will not appear in Itô's formula as $\langle X_1\rangle\equiv0$ and $\langle X_1,X_2\rangle=\langle X_2,X_1\rangle\equiv0$.

Put $M(t)=f(X_1(t),X_2(t))=f(t,X(t))$. By Itô's formula

$$
\begin{aligned}
M(t)\;&=\;1+\int_0^t f_t(s,X(s))\,dX_1(s)+\int_0^t f_x(s,X(s))\,dX_2(t)+\frac{1}{2}\int_0^t f_{xx}(s,X(s))\,d\langle X_2\rangle_s\\
&=\;1+\frac{1}{2}r^2\int_0^t M(s)\,ds+ir\int_0^t M(s)\,dX(s)-\frac{1}{2}r^2\int_0^t M(s)\,ds
\end{aligned}
$$

$$= \quad 1 + ir\int_0^t M(s)\,dX(s)\,.$$

It follows that M is a continuous local martingale. In fact, we have $|M(t)| = e^{\frac{r^2}{2}t} \leq e^{\frac{r^2}{2}T} < \infty$ and hence M is a bounded local martingale and thus a martingale. Consequently, for $0 \leq s < t$ we have $\mathbb{E}[M(t)|\mathscr{F}_s] = M(s)$, i.e.

$$\mathbb{E}\big[e^{irX(t)-\frac{1}{2}r^2t}\big|\mathscr{F}_s\big] = e^{irX(s)-\frac{1}{2}r^2s}.$$

Or, equivalently,

$$\mathbb{E}\big[e^{ir(X(t)-X(s))}\big|\mathscr{F}_s\big] = e^{-\frac{r^2}{2}(t-s)}\,.$$

Taking expectations, it follows that the characteristic function of $X(t) - X(s)$ is given by $\varphi_{X(t)-X(s)}(r) = e^{-\frac{r^2}{2}(t-s)}$, i.e. $X(t) - X(s)$ has normal distribution with mean zero and variance $t - s$. Moreover, for every $A \in \mathscr{F}_s$, we have

$$\int_A e^{ir(X(t)-X(s))}\,d\mathbb{P} = e^{-\frac{r^2}{2}(t-s)}\mathbb{P}(A)\,.$$

which implies that $X(t)-X(s)$ is independent of $\mathscr{F}_s$. Consequently, X is indeed a Brownian motion. □

Remark 3.3.2. We should point out that it is important that X has *continuous paths* in Theorem 3.3.1. Indeed, the compensated Poisson process of intensity 1 is a martingale which has "quadratic variation" t but is different from a Brownian motion.

There is also a generalization to higher dimensions:

Corollary 3.3.3. *Let $X_1, \ldots, X_d$ be continuous local martingales with $\langle X_i, X_j\rangle = \delta_{ij}t$. Then $X_1, \ldots, X_d$ are independent Brownian motions.*

3.3.2. Burkholder-Davies-Gundy inequalities. In Lemma 2.1.15, we have seen that $\mathsf{M}_2([0,T])$ is a closed linear subspace of $L^2(\Omega, C([0,T]))$ and that $\|X\| := \big(\mathbb{E}|X(T)|^2\big)^{\frac{1}{2}}$ defines an equivalent norm on that space. Since $X^2 - \langle X\rangle$ is a martingale and $X(0) = 0 = \langle X\rangle_0$, we find $\mathbb{E}|X(T)|^2 = \mathbb{E}\langle X\rangle_T$. Thus, also $\big(\mathbb{E}\langle X\rangle_T\big)^{\frac{1}{2}}$ defines an equivalent norm on $M_2([0,T])$.

We now extend this also to the L^p-setting, even for $p \in (0,\infty)$.

Theorem 3.3.4. *For every $p \in (0,\infty)$ there exist constants $c_p > 0$ and C_p such that*

$$c_p\mathbb{E}\langle X\rangle_T^{\frac{p}{2}} \leq \mathbb{E}\|X\|^p_{C([0,T])} \leq C_p\mathbb{E}\langle X\rangle_T^{\frac{p}{2}} \tag{3.5}$$

for all $X \in \mathsf{M}_{\mathrm{loc}}([0,T])$.

Implicitly in (3.5) is the assertion that if one of the terms is finite, then the other term is also finit.

We first establish some special cases.

Lemma 3.3.5. *For $p \geq 2$ there exists a constant C_p such that $\mathbb{E}\|X\|^p_{C([0,T])} \leq C_p\mathbb{E}\langle X\rangle_T^{\frac{p}{2}}$ for all $X \in \mathsf{M}_{\mathrm{loc}}([0,T])$.*

Proof. By stopping, we may assume that X is bounded.

We apply Itô's formula for $f(x) = |x|^p$. and obtain

$$|X(t)|^p = \int_0^t p|X(s)|^{p-1}\mathrm{sgn}(X(s))\,dX(s) + \frac{1}{2}\int_0^t p(p-1)|X(s)|^{p-2}\,d\langle X\rangle_s\,.$$

As the middle term is a martingale, we find

$$\mathbb{E}|X(T)|^p \quad = \quad \frac{p(p-1)}{2}\mathbb{E}\int_0^T |X(s)|^{p-2}\,d\langle X\rangle_s$$

$$\begin{aligned} &\leq \frac{p(p-1)}{2}\mathbb{E}\|X\|_\infty^{p-2}\langle X\rangle_T \\ &\leq \frac{p(p-1)}{2}\big(\mathbb{E}\langle X\rangle_T^{\frac{p}{2}}\big)^{\frac{2}{p}}\big(\mathbb{E}\|X\|_\infty^p\big)^{\frac{p-2}{p}} \end{aligned}$$

where the last inequality follows from Hölder's inequality with $\frac{p}{2}$ and $\frac{p}{p-2}$. By Doob's maximal inequality (Corollary 2.1.9), $\mathbb{E}\|X\|_\infty^p \leq \frac{p}{p-1}\mathbb{E}|X(T)|^p$, whence

$$\mathbb{E}\|X\|_\infty^p \leq \frac{p^2}{2}\big(\mathbb{E}\langle X\rangle_T^{\frac{p}{2}}\big)^{\frac{2}{p}}\big(\mathbb{E}\|X\|_\infty^p\big)^{\frac{p-2}{p}}$$

which is equivalent with the assertion. □

Now we prove a lower estimate.

Lemma 3.3.6. *For $p \geq 4$ there exists a constant $c_p > 0$ such that $c_p\mathbb{E}\langle X\rangle_T^{\frac{p}{2}} \leq \mathbb{E}\|X\|_\infty^p$ for all $X \in \mathsf{M}_{\mathrm{loc}}([0,T])$.*

Proof. Again, we assume that X is bounded.

Since $X(t)^2 = 2\int_0^t X(s)\,dX(s) + \langle X\rangle_t$ (this is Itô's formula for $f(x) = x^2$ or the integration by parts formula in Theorem 3.2.7) we have

$$\mathbb{E}\langle X\rangle_T^{\frac{p}{2}} \leq a_p\Big(\mathbb{E}|X(T)|^p + 2^p\mathbb{E}\Big(\int_0^T X(s)\,dX(s)\Big|^{\frac{p}{2}}\Big) \leq 2^p a_p\Big(\mathbb{E}\|X\|_\infty^p + \mathbb{E}\Big|\int_0^T X(s)\,dX(s)\Big)^{\frac{p}{2}}\Big).$$

Applying Lemma 3.3.5 to the martingale $X \diamond X$, we find

$$\begin{aligned} \mathbb{E}\langle X\rangle_T^{\frac{p}{2}} &\leq 2^p a_p\Big(\mathbb{E}\|X\|_\infty^p + C_{\frac{p}{2}}\mathbb{E}\Big(\int_0^T X(s)^2\,d\langle X\rangle_s\Big)^{\frac{p}{4}}\Big) \\ &\leq 2^p a_p C_{\frac{p}{2}}\Big(\mathbb{E}\|X\|_\infty^p + \mathbb{E}\big(\|X\|_\infty^{\frac{p}{2}}\langle X\rangle_T^{\frac{p}{4}}\big)\Big) \\ &\leq 2^p a_p C_{\frac{p}{2}}\Big(\mathbb{E}\|X\|_\infty^p + \big(\mathbb{E}\|X\|_\infty^p\big)^{\frac{1}{2}}\big(\mathbb{E}\langle X\rangle_T^{\frac{p}{2}}\big)^{\frac{1}{2}}\Big). \end{aligned}$$

Setting $\alpha = 2^p a_p C_{\frac{p}{2}}$, $x = \big(\mathbb{E}\langle X\rangle_T^{\frac{p}{2}}\big)^{\frac{1}{2}}$ and $y = \big(\mathbb{E}\|X\|_\infty^p\big)^{\frac{1}{2}}$, this is equivalent with

$$0 \geq x^2 - \alpha xy - \alpha y^2 = \Big(x - \frac{\alpha y}{2} - \sqrt{\alpha + \frac{\alpha}{4}}y\Big)\Big(x - \frac{\alpha y}{2} + \sqrt{\alpha + \frac{\alpha}{4}}y\Big).$$

As x, y and α are nonnegative, this forces $x \leq (\frac{\alpha}{2} + \sqrt{\alpha + \frac{\alpha}{4}})y$. This is the assertion. □

To extend these results, we make use of the following

Lemma 3.3.7. *Let M be a continuous, adapted process and A be a continuous, adapted increasing process such that for all stopping times τ, we have $\mathbb{E}M_\tau \leq \mathbb{E}A_\tau$. Then for all $x, y > 0$, we have*

$$\mathbb{P}(\|M\|_\infty > x, A_T \leq y) \leq \frac{1}{x}\mathbb{E}(A_T \wedge y).$$

Proof. Define $\tau := \inf\{t \in [0,T] : A_t > y\}$ and $\sigma := \inf\{t \in [0,T] : M_t > y\}$. Note that $\{A_T \leq y\} = \{\tau = T\}$, as A is increasing. Moreover, $A_{\tau\wedge\sigma} \leq A_T \wedge y$ as A is increasing and continuous. Thus

$$\begin{aligned} \mathbb{P}\big(\|M\|_\infty > x, A_T \leq y\big) &= \mathbb{P}\big(\|M\|_\infty > x, \tau = T\big) \\ &\leq \mathbb{P}\big(M_\sigma \geq x, \sigma < T, \tau = T\big) \\ &\leq \frac{1}{x}\mathbb{E}M_{\tau\wedge\sigma} \leq \frac{1}{x}\mathbb{E}M_{\tau\wedge\sigma} \leq \mathbb{E}A_T \wedge y. \quad \square \end{aligned}$$

Corollary 3.3.8. *Under the Hypothesis of Lemma 3.3.7, for $r \in (0,1)$, we have*

$$\mathbb{E}\|M\|_\infty^r \leq \frac{2-r}{1-r}\mathbb{E}A_T^r.$$

Proof. Put $F(t) := t^r$. Then F is an increasing, continuous function with $F(0) = 0$. Thus

$$\begin{aligned}
\mathbb{E}\|M\|_\infty &= \mathbb{E}\int_0^\infty \mathbb{1}_{\{\|M\|_\infty > t\}}\, dF(t)\\
&= \int_0^\infty \mathbb{E}\mathbb{1}_{\{\|M\|_\infty > t\}}\, dF(t)\\
&\le \int_0^\infty \mathbb{P}(\|M\|_\infty > t, A_T \le t) + \mathbb{P}(A_T > t)\, dF(t)\\
&\le \int_0^\infty \frac{1}{t}\mathbb{E}(A_T \wedge t) + \mathbb{P}(A_T > t)\, dF(t)\\
&\le \int_0^\infty 2\mathbb{P}(A_T > t) + \frac{1}{t}\mathbb{E}A_T\mathbb{1}_{\{A_T \le t\}}\, dF(t)\\
&= 2\mathbb{E}F(A_T) + \mathbb{E}A_T\int_{A_T}^\infty \frac{1}{t}\, dF(t)\\
&= \frac{2-r}{1-r}\mathbb{E}A_T^r\,. \quad \square
\end{aligned}$$

We are now ready for the proof of Theorem 3.3.4.

Proof of Theorem 3.3.4. The Hypothesis of Lemma 3.3.7 is satisfied for $M(t) := (\sup_{s\le t}|X(s)|)^2$ and $A(t) := C_2\langle X\rangle_t$ as a consequence of Lemma 3.3.5 applied to stopped processes. Thus, Corollary 3.3.8 yileds the upper estimate in 3.5 also for $p \in (0, 2)$.

Similarly, applying Corollary 3.3.8 to $M(t) := \langle X\rangle^2$ and $A(t) := C_4(\sup_{s\le t}|X(s)|)^4$ gives the lower estimate in 3.5 also for $p \in (0, 4)$. $\square$

3.4. Exercises

(1) In this exercise, we take a closer look at the measurability assumptions which appear in connection with Itô's integral. Show that the following σ-algebras on $\Omega \times [0, T]$ coincide:
   (i) the σ-algebra generated by the adapted step processes.
   (ii) the σ-algebra generated by the adapted left-continuous processes.
   (iii) the σ-algebra generated by the adapted continuous processes.

   This σ-algebra is called the *predictable σ-algebra* $\mathscr{P}$.

   Also show that a map $X : \Omega \times [0, T]$ is $\mathscr{P}$-measurable if and only if it is the pointwise limit of a sequence of adapted step processes. It follows that

$$\mathscr{I}(W) = \left\{X \in L^0(\Omega \times [0, T], \mathscr{P}, \mathbb{P} \otimes \lambda) : X(\cdot, \omega) \in L^2([0, T]\ a.e.\right\}.$$

(2) We define the *Hermite Polynomials* $H_n(x, t)$ by

$$H_n(x,t) = \frac{\partial^n}{\partial \lambda^n} e^{\lambda x - \frac{1}{2}\lambda^2 t}\Big|_{\lambda=0}$$

   for $n \in \mathbb{N}_0$.
   (a) Prove that

$$\frac{1}{2}\frac{\partial^2}{\partial x^2}H_n + \frac{\partial}{\partial t}H_n = 0$$

   for all $n \in \mathbb{N}_0$ and that

$$\frac{\partial}{\partial x}H_n = nH_{n-1}\,.$$

   (b) Now, let $(W(t))_{t\ge 0}$ be a Brownian motion and $M^n(t) := H_n(W(t), t)$. Show that M^n is a martingale and determine $\langle M^n\rangle$.

(3) In this exercise, we give an introduction to *Brownian local time*. We have established Itô's formula for functions f which are twice continuously differentiable. Suppose, we want to apply it to $f(x) = |x| =: \mathrm{abs}(x)$ which is obviously *not* twice continuously differentiable. The distributional derivatives of abs are given by

$$\mathrm{abs}'(x) = \mathrm{sign}(x) = \begin{cases} 1 & , x > 0 \\ 0 & , x = 0 \\ -1 & , x < 0 \end{cases} \quad \text{and} \quad \mathrm{abs}''(x) = 2\delta_0 \, .$$

Being bold, we formally apply Itô's formula with this and obtain

$$|W(t)| = \int_0^t \mathrm{sign}(W(s))\, dW(s) + \int_0^t \delta_0(W(s))\, ds \, .$$

Thus the last term in this equation somehow measures how much time Brownian motion spends at zero.

Putting $Z := \{(t, \omega) : W(t, \omega) = 0\}$, we find with Fubini's theorem

$$(\mathbb{P} \otimes \lambda)(Z) = \mathbb{E} \int_0^T \mathbb{1}_{\{0\}}(W(s))\, ds = \int_0^T \mathbb{E}\mathbb{1}_{\{0\}}(W(s))\, ds = 0,$$

that is, almost surely, the Lebesgue measure of the time that Brownian motion spends at zero is 0.

Nevertheless, in this exercise we will prove that there exists a continuous, increasing process L such that

$$|W(t)| = \int_0^t \mathrm{sign}(W(s))\, dW(s) + L(t) \tag{3.6}$$

almost surely for all $t \in [0, T]$. This is *Tanaka's formula for Brownian local time.*

(a) Given a sequence $a_n \downarrow 0$, we pick functions $\varphi_n \in C_c^\infty((0, \infty))$ which are positive, supported in (a_{n+1}, a_n) and satisfy $\int_0^\infty \varphi_n(t)\, dt = 1$. We define $f_n : \mathbb{R} \to \mathbb{R}$ by

$$f_n(x) = \int_0^{|x|} \int_0^t \varphi_n(s)\, ds\, dt \, .$$

Show that $f_n \in C^\infty(\mathbb{R})$, that $f_n'(x) \to \mathrm{sign}(x)$ for all $x \in \mathbb{R}$, that $f_n(x) \to |x|$ uniformly on compact sets and that $f_n''(x) = \varphi_n(|x|)$.

(b) Applying Itô's formula to f_n, we obtain

$$\begin{aligned} X_n(t) := f_n(W(t)) &= \int_0^t f_n'(W(s))\, dW(s) + \frac{1}{2} \int_0^t f_n''(W(s))\, ds \\ &=: I_n(t) + L_n(t) \, . \end{aligned}$$

Moreover, we define $I(t) := \int_0^t \mathrm{sign}(W(s))\, dW(s)$ and $X(t) = |W(t)|$. Prove that $X_n \to X$ and $I_n \to I$ in $L^2(\Omega; C([0, T]))$. Conclude that L_n converges in $L^2(\Omega; C([0, T]))$ to a limit L which satisfies (3.6).

Hint: Use Doob's maximal inequality and dominated convergence

(c) Show that L is a continuous, increasing process which is adapted to the filtration generated by $|W(t)|$. Moreover, show that, $\mathbb{P}$-almost surely, the paths of L are λ-almost everywhere differentiable with $L' \equiv 0$. Hence, the paths of L look like the Cantor function: They are increasing but outside of a null set constant.

Hint: For almost every ω the set $Z(\omega) := \{t \in [0, T] : W(t, \omega) = 0\}$ is closed and has Lebesgue measure zero. It's complement is open. Prove that for t_0 in the complement, $L(\cdot, \omega)$ is differentiable in t_0 with derivative 0. This proves the last assertion.

(d) For later purposes, we introduce also the function $\operatorname{sgn}(W(s)) := \mathbb{1}_{(0,\infty)} - \mathbb{1}_{(-\infty,0]}$. Show that $\int_0^t \operatorname{sign}(W(s))\, dW(s) = \int_0^t \operatorname{sgn}(W(s))\, dW(s)$ almost surely for $t \in [0, T]$. Hence, (3.6) also holds when we replace sign with sgn.

CHAPTER 4

Stochastic Differential Equations with Locally Lipschitz Coefficients

4.1. Solutions via Banach's Fixed Point Theorem

In this section, we let $W = (W_1, \ldots, W_m)$ be an m-dimensional Brownian motion, i.e. a vector of m independent Brownian motions. Moreover, we are given a finite time horizon T and continuous functions $f : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times m}$ which are Lipschitz continuous and of linear growth in the second variable, uniformly in the first, i.e. there exists a constant $L \geq 0$ such that for all $t \in [0,T]$ and $x, y \in \mathbb{R}^d$

$$\|f(t,x) - f(t,y)\| \leq L\|x-y\| \quad \text{and} \quad \|\sigma(t,x) - \sigma(t,y)\| \leq L\|x-y\|$$

and there exist constants $a, b \geq 0$ such that

$$\|f(t,x)| \leq a + b\|x\| \quad \text{and} \quad |\sigma(t,x)\| \leq a + b\|x\|$$

for all $t \in [0,T]$ and $x \in \mathbb{R}^d$. Here, $\|\cdot\|$ denotes the euclidean norm on $\mathbb{R}^d$ resp. $\mathbb{R}^{d\times m}$. Note that the linear growth condition follows from the Lipschitz assumption if we additionally assume that $\sup_{t\in[0,T]} \|f(t,0)\| < \infty$.

We are concerned with the stochastic differential equation

$$\left\{ \begin{array}{lcl} dX(t) &=& f(t, X(t))dt + \sigma(t, X(t))dW(t) \\ X(0) &=& \eta \end{array} \right. \tag{4.1}$$

which we us as a shorthand for

$$dX_i(t) = f_i(t, X_1(t), \ldots, X_n(t))dt + \sum_{j=1}^m \sigma_{ij}(t, X_1(t), \ldots, X_n(t))dW_j(t) \quad X_i(0) = \eta_i$$

for $i = 1, \ldots, d$. This again is a shorthand notation for the integral equation

$$X_i(t) = \eta_i + \int_0^t f_i(s, X(s))\, ds + \sum_{j=1}^m \int_0^t \sigma_{ij}(s, X(s))\, dW_j(s)$$

for $i = 1, \ldots, d$.

We are given a filtered probability space $(\Omega, \Sigma, \mathbb{F}, \mathbb{P})$ and the Brownian motions W_j are $\mathbb{F}$-Brownian motions. Finally, $\eta \in L^2(\Omega, \mathscr{F}_0, \mathbb{P})$. A *solution* of equation (4.1) is a process $X \in L^2_{\mathbb{F}}(\Omega; C([0,T]; \mathbb{R}^d))$ such that, almost surely, (4.1) holds for all $t \in [0,T]$. Here, $L^2_{\mathbb{F}}(\Omega; C([0,T]; \mathbb{R}^d))$ denotes the subspace of $L^2(\Omega; C([0,T]; \mathbb{R}^d))$ consisting of all adapted processes.

Theorem 4.1.1. *Let $f : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times m}$ be continuous functions which are Lipschitz continuous and of linear growth in the second variable, uniformly in in the first. Then, for every $\eta \in L^2(\Omega; \mathscr{F}_0; \mathbb{P})$ there exists a unique solution $X \in L^2_{\mathbb{F}}(\Omega; C([0,T]; \mathbb{R}^d))$ which solves the stochastic differential equation* (4.1)

Proof. The main idea of the proof is to consider the map $\Phi : E \to E$, where E denotes the Banach space $L^2_{\mathbb{F}}(\Omega; C([0,T]; \mathbb{R}^d))$ and Φ is defined by

$$[\Phi(u)](t) := \eta + \int_0^t f(s, u(s))\, ds + \int_0^t \sigma(s, u(s))\, dW(s)\,,$$

show that it satisfies a suitable Lipschitz condition and apply Banach's fixed point theorem. We denote by L a common Lipschitz constant for both f and σ and a be a number such that $\sup_{0\le t\le T}\|f(t,0)\|$, $\sup_{0\le t\le T}\|f(t,0)\|\le a$.

Let us first verify that Φ is indeed well-defined.

It suffices to prove that the three terms in the definition of Φ again yield processes in E. Concerning η, there is nothing to prove. Concerning the deterministic integral, first note that for almost every ω the map $s\mapsto f(s,u(s))$ is measurable and bounded (by the linear growth assumption) hence integrable on $(0,t)$. Moreover, $\int_0^t f(u(s))\,ds$ is certainly $\mathscr{F}_t$-measurable since u, hence $f(\cdot,u(\cdot))$ is adapted. Thus, the deterministic integral is adapted. Finally, we have

$$\Big\|\int_0^t f(s,u(s))\,ds\Big\|\le\int_0^t a+L\|u(s)\|\,ds\le(a+L)T\|u\|_\infty$$

which is square integrable.

Let us now consider the stochastic integral. Obviously, the components of $\sigma(\cdot,u(\cdot))$ belong to $L^2_{\mathbb{F}}(\Omega;C([0,T]))$, hence they are stochastically integrable with respect to a Brownian motion, as $L^2_{\mathbb{F}}(\Omega;C([0,T]))\subset L^2_{\mathbb{F},W_j}(\Omega\times[0,T])$, cf. Exercise 1 from Chapter 3. It follows that $\int_0^\cdot\sigma(s,u(s))\,dW(s)$ is a continuous square integrable martingale, in particular, it belongs to $L^2_{\mathbb{F}}(\Omega;C([0,T]))$.

Next, we prove Lipschitz continuity of Φ. For $r\in[0,T]$ and $u,v\in E$, we have

$$\begin{aligned}\mathbb{E}\sup_{0\le t\le r}\Big\|[\Phi(u)](t)-[\Phi(v)](t)\Big\|^2 \;&\le\; 2\mathbb{E}\sup_{0\le t\le r}\Big\|\int_0^t f(s,u(s))-f(s,v(s))\,ds\Big\|^2\\ &\quad+2\mathbb{E}\sup_{0\le t\le r}\Big\|\int_0^t\sigma(s,u(s))-\sigma(s,v(s))\,dW(s)\Big\|^2.\end{aligned}$$

We now treat the two terms on the right hand side separately. Using Jensen's inequality,

$$\begin{aligned}\mathbb{E}\sup_{0\le t\le r}\Big\|\int_0^t f(s,u(s))-f(s,v(s))\,ds\Big\|^2 &\le \mathbb{E}\sup_{0\le r\le r} t\int_0^t\|f(s,u(s))-f(s,v(s))\|^2\,ds\\ &\le TL^2\int_0^r\mathbb{E}\|u(s)-v(s)\|^2\,ds\end{aligned}$$

Doob's maximal inequality and the Itô isometry yield for the stochastic integral

$$\begin{aligned}&\mathbb{E}\sup_{0\le t\le r}\Big\|\int_0^t\sigma(s,u(s))-\sigma(s,v(s))\,dW(s)\Big\|^2\le 2\mathbb{E}\Big\|\int_0^r\sigma(s,u(s))-\sigma(s,v(s))\,dW(s)\Big\|^2\\ =\;& 2\sum_{i=1}^d\sum_{j=1}^m\mathbb{E}\Big|\int_0^r\sigma_{ij}(s,u(s))-\sigma_{ij}(s,v(s))\,dW_j(s)\Big|^2\\ =\;& 2\sum_{i=1}^d\sum_{j=1}^m\mathbb{E}\int_0^r|\sigma_{ij}(s,u(s))-\sigma_{ij}(s,v(s))|^2\,ds\\ \le\;& 2\sum_{i=1}^d\sum_{j=1}^m\int_0^r L^2\mathbb{E}\|u(s)-v(s)\|^2\,ds\\ \le\;& 2mdL^2\int_0^r\mathbb{E}\|u(s)-v(s)\|^2\,ds\end{aligned}$$

Thus, altogether, we obtain that for a certain constant C, we have

$$\mathbb{E}\sup_{0\le t\le r}\big\|[\Phi(u)](t)-[\Phi(v)](t)\big\|^2\le C\int_0^r\mathbb{E}\|u(s)-v(s)\|^2\,ds$$

Iterating, we find

$$\mathbb{E}\sup_{0\leq t\leq r}\big\|[\Phi^n(u)](t)-[\Phi^n(v)](t)\big\|^2\leq\frac{(Cr)^n}{n!}\mathbb{E}\|u-v\|_\infty^2\,.$$

Consequently, $\|\Phi^n(u)-\Phi^n(v)\|_E^2\leq\frac{(CT)^n}{n!}\|u-v\|_E^2$. It now follows from a variant of Banach's fixed point theorem that Φ has a unique fixed point, i.e. our stochastic differential equation has a unique solution. □

4.2. Extension to locally Lipschitz Coefficients

It is a rather natural question, whether we can extend this result to stochastic differential equations with *locally* Lipschitz continuous coefficients and/or initial data which are still $\mathscr{F}_0$-measurable but not necessarily square integrable. Of course, we have to modify our solution concept slightly.

Definition 4.2.1. Given $f:[0,T]\times\mathbb{R}^d\to\mathbb{R}^d$ and $\sigma:[0,T]\times\mathbb{R}^d\to\mathbb{R}^{d\times m}$ Borel measurable and $\eta\in L^0(\Omega;\mathscr{F}_0,\mathbb{P})$, a *solution* of (4.1) is an element $X\in L^0_{\mathbb{F}}(\Omega;C([0,T];\mathbb{R}^d))$ such that

(1) Almost surely, $s\mapsto f_i(s,X(s))\in L^1([0,T])$ for all $1\leq i\leq d$.
(2) Almost surely $s\mapsto\sigma_{ij}(s,X(s))\in L^2([0,T])$ for all $1\leq i\leq d$ and $1\leq j\leq m$.
(3) Almost surely, we have

$$X(t)=\eta+\int_0^t f(s,X(s))\,ds+\int_0^t\sigma(s,X(s))\,dW(s)$$

for all $t\in[0,T]$.

In view of Exercise (1) of Chapter 3, requirement (2) is equivalent with $\sigma_{ij}(\cdot,X(\cdot))\in\mathscr{I}(W_j)$ for all $1\leq i\leq d$ and $1\leq j\leq m$.

We next prove that if f and σ are locally Lipschitz continuous, then we have uniqueness of solutions.

Proposition 4.2.2. *Assume that $f:[0,T]\times\mathbb{R}^d\to\mathbb{R}^d$ and $\sigma:[0,T]\times\mathbb{R}^d\to\mathbb{R}^{d\times m}$ are locally Lipschitz continuous in the second variable, uniformly in the first, i.e. for every $n\in\mathbb{N}$ there exists a constant L_n such that for x,y with $\|x\|,\|y\|\leq n$ we have*

$$\|f(t,x)-f(t,y)\|\leq L_n\|x-y\|\quad\text{and}\quad\|\sigma(t,x)-\sigma(t,y)\|\leq L_n\|x-y\|$$

for all $t\in[0,T]$. Then, if X and Y are solutions of (4.1), we have $X=Y$ almost surely.

Proof. We define the stopping time τ_n by

$$\tau_n:=\inf\{t>0\,:\,\|X(t)\|\geq m\}\wedge\inf\{t>0\,:\,\|Y(t)\|\geq m\}\,.$$

Then, since X and Y are solutions we find, with similar estimates as in the proof of Theorem 4.1.1,

$$\begin{aligned}
&\mathbb{E}\|X^{\tau_n}(t)-Y^{\tau_n}(t)\|^2\\
=\;&\mathbb{E}\Big\|\int_0^{t\wedge\tau_n}f(s,X(s))-f(s,Y(s))\,ds+\int_0^{t\wedge\tau_n}\sigma(s,X(s))-\sigma(s,Y(s))\,dW(s)\Big\|^2\\
\lesssim\;&T\mathbb{E}\int_0^{t\wedge\tau_n}\|f(s,X(s))-f(s,Y(s))\|^2\,ds+\mathbb{E}\int_0^{t\wedge\tau_n}\|\sigma(s,X(s))-\sigma(s,Y(s))\|^2\,ds\\
\leq\;&T\mathbb{E}\int_0^{t\wedge\tau_n}L_n\|X(s)-Y(s)\|^2\,ds+\mathbb{E}\int_0^{t\wedge\tau_n}L_n\|X(s)-Y(s)\|^2\,ds\\
\leq\;&\int_0^t\mathbb{E}\|X^{\tau_n}(s)-Y^{\tau_n}(s)\|^2\,ds\,.
\end{aligned}$$

Hence, putting $\varphi_m(t) := \mathbb{E}\|X^{\tau_n}(t) - Y^{\tau_n}(t)\|^2$, we have, for a suitable constant C_m, $\varphi_m(t) \leq C_m \int_0^t \varphi_m(s)\,ds$. By Gronwall's lemma, $\varphi_m(t) = 0$ for all $t \geq 0$. It hence follows that for every $t \in [0,T]$ we have $X^{\tau_n}(t) = Y^{\tau_n}(t)$ almost surely. Note that the exceptional set may depend on t. However, by continuity of the paths, taking the union of exceptional sets for $t \in [0,T] \cap \mathbb{Q}$, we find a null set outside of which $X^{\tau_n}(t) = Y^{\tau_n}(t)$ for all $t \in [0,T]$.

By continuity of the paths, we have $\Omega = \bigcup_{m\in\mathbb{N}}\{\tau_n = T\}$. Hence, we infer that $X = Y$ in $L^0(\Omega; C([0,T];\mathbb{R}^d))$ as claimed. □

In particular, we see that in the situation of Theorem 4.1.1, we do not find more solutions if we allow solutions to be defined in $L^0_{\mathbb{F}}(\Omega; C([0,T];\mathbb{R}^d))$ rather than $L^2_{\mathbb{F}}(\Omega; C([0,T];\mathbb{R}^d))$. From the proof of Proposition 4.2.2 we obtain the following Corollary, which, roughly speaking, states that solutions depend on the coefficients locally.

Corollary 4.2.3. *Assume that $f_1, f_2 : [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$ and $\sigma_1, \sigma_2 : [0,T]\times\mathbb{R}^d \to \mathbb{R}^{d\times m}$ are locally Lipschitz continuous in the second variable, uniformly in the first, and satisfy $f_1(t,x) = f_2(t,x)$ and $\sigma_1(t,x) = \sigma_2(t,x)$ for all $t \in [0,T]$ and x with $\|x\| \leq m$. Moreover, let $\eta_1, \eta_2 \in L^0(\Omega, \mathscr{F}_0, \mathbb{P})$ be such that $\eta_1 \mathbb{1}_{\{\|\eta_1\|\leq m\}} = \eta_2 \mathbb{1}_{\{\|\eta_2\|\leq m\}}$.*

Finally, let X_i be a solution of the stochastic differential equation with coefficients f_i and σ_i and initial datum η_i and $\sigma_i = \inf\{t \in [0,T] : \|X_i\| \geq m\}$.

Then, almost surely, $\sigma_1 = \sigma_2$ and $X_1^{\sigma_1} = X_2^{\sigma_2}$.

Proof. It suffices to copy the proof of Proposition 4.2.2 with τ_n replaced with $\tau_n \wedge \sigma_1$. Note that in this case, we can have $f_1(s, X_1(s)) = f_2(s, X_1(s))$ for all $s \leq \tau_n \wedge \sigma_1$ and similarly for the σ's. Hence we obtain $X_1^{\sigma_1} = X_2^{\sigma_1}$. Interchanging the roles of X_1 and X_2, we obtain $X_1^{\sigma_2} = X_2^{\sigma_2}$. Which yields the claim. □

Given locally Lipschitz continuous coefficients, we can "freeze" the coefficients outside a Ball of radius n thus obtaining globally Lipschitz continuous coefficients. The resulting equation we can solve using Theorem 4.1.1, obtaining a solution X_n. Putting $\tau_n := \inf\{t \in [0,T] : \|X_n(t)\| \geq n\}$ we obtain from Corollary 4.2.3, that $X_n^{\tau_n} = X_m^{\tau_n}$ for all $m \geq n$ in particular, $\tau_n \geq \tau_n$.

We may thus "glue together" these solutions to a *maximal solution* $(X(t))_{t\in[0,\tau)}$ by setting $\tau := \sup_n \tau_n$ and $X(t) := X_n(t)$ for $t \leq \tau_n$. The stopping time τ is called the *life time* of the process. By definition, it is clear that if $\tau < T$ then $\|X(t)\| \to \infty$ as $t \uparrow \tau$. Hence τ is also called *explosion time*. Already deterministic equations exhibit this behavior. For example, if $n = d = 1$, $f(t,x) = -x^2$ and $\sigma \equiv 0$, our equation becomes $dX(t) = -X(t)^2dt$ or, as an ODE, $u' = -u^2$. This ODE with initial condition $u(0) = 1$ has the unique solution $u(t) = (1-t)^{-1}$ which explodes at time $\tau = 1$.

In these lecture notes, we are more interested in situations where $\tau \equiv T$ and no explosion occurs, i.e. we have a solution in the sense of Definition 4.2.1. Therefore, we do not develop the concept of "maximal solution" for stochastic differential equations.

Let us now give a first example where we obtain solutions in the sense of Definition 4.2.1.

Proposition 4.2.4. *Let $f : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times m}$ be Borel measurable and locally Lipschitz continuous in the second variable, uniformly in the first, i.e. for all $n \in \mathbb{N}$ there exists a constant L_n such that*

$$\|f(t,x) - f(t,y)\|, \|\sigma(t,x) - \sigma(t,y)\| \leq L_n\|x - y\|$$

for all $t \in [0,T]$ and x, y with $\|x\|, \|y\| \leq n$. Moreover, assume that there exist a, b such that

$$\|f(t,x)\|, \|\sigma(t,x)\| \leq a + b\|x\|$$

for all $(t,x) \in [0,T] \times \mathbb{R}^d$.

Then, for every $\eta \in L^2(\Omega, \mathscr{F}_0; \mathbb{R}^d)$, there exists a unique solution of equation (4.1).

PROOF. Uniqueness is immediate from Proposition 4.2.2. To prove existence, we put

$$f_n(t,x) := \begin{cases} f(t,x) & \|x\| \leq n \\ f\big(t, \frac{nx}{\|x\|}\big) & \|x\| > n \end{cases} \quad \text{and} \quad \sigma_n(t,x) := \begin{cases} \sigma(t,x) & \|x\| \leq n \\ \sigma\big(t, \frac{nx}{\|x\|}\big) & \|x\| > n \end{cases}.$$

Then f_n and g_n satisfy the hypothesis of Theorem 4.1.1. We let X_n be the unique solution of equation (4.1) with f and σ replaced with f_n and σ_n and initial datum η. To prove global existence of a solution to equation (4.1), it suffices to prove that $\mathbb{E}\|X_n\|_\infty^2 \leq C$ for a constant C independent of n. Indeed, In this case, by Chebyshev's inequality,

$$\mathbb{P}(\|X_n\|_\infty \geq n) \leq Cn^{-2}.$$

As n^{-2} is summable, the Borel-Cantelli lemma yields

$$\mathbb{P}\Big(\bigcap_{k\geq 1}\bigcup_{n\geq k}\{\|X_n\|_\infty \geq n\}\Big) = 0$$

Hence, almost surely, $\tau_n := \inf\{t \in [0,T] : \|X_n(t)\| \geq n\} = T$ for all n large enough. Consequently, gluing together the X_n as described above, we obtain a solution of our initial equation defined for all $t \in [0,T]$.

It remains to prove the boundedness of $\mathbb{E}\|X_n\|_\infty^2$. Using the linear growth assumption on the coefficients, similar estimates as in the proof of Theorem 4.1.1 yield the estimated $\mathbb{E}\|X_n\|_\infty^2 \leq C(T)\mathbb{E}\|X_n\|_\infty^2 + C_1 + C_2\mathbb{E}\|\eta\|^2$, where C_1, C_2 are coefficients which only depend on the coefficients in the linear growth assumption and $C(T) \to 0$ for $T \to 0$. Hence, for T_0 small enough,

$$\mathbb{E}\|X_n\|_{C([0,T_0])}^2 \leq \frac{1}{1-C(T_0)}(C_1 + C_2\mathbb{E}\|\eta\|^2).$$

Next note that on the filtered probability space $(\Omega, \Sigma, (\mathscr{F}_t)_{t\in[T_0,T]}, \mathbb{P})$ the process $\tilde{W}(t) := W(T_0 + t) - W(t)$ is a Brownian motion. Moreover, the processes $(X_n(t))_{t\in[T_0,T]}$ solve the differential equation (4.1) with initial datum $X_n(T_0)$ at time T_0 and coefficients f_n and σ_n. Thus,

$$\begin{aligned}\mathbb{E}\|X_n\|_{C([T_0,2T_0])}^2 &\leq \frac{1}{1-C(T_0)}(C_1 + C_2\mathbb{E}\|X_n(T_0)\|^2) \\ &\leq \frac{1}{1-C(T_0)}\Big(C_1 + C_2\Big(\frac{1}{1-C(T_0)}(C_1 + C_2\mathbb{E}\|\eta\|^2)\Big)\Big).\end{aligned}$$

Inductively, we obtain constants C_k such that $\mathbb{E}\|X_n\|_{C([0,kT_0])}^2 \leq C_k(1+\mathbb{E}\|\eta\|^2)$ for all $n \in \mathbb{N}$. Eventually, we have proved the boundedness of $\mathbb{E}\|X_n\|_{C([0,T])}^2$ independently of n. □

4.3. Examples

4.3.1. Geometric Brownian motion. Let $X(t)$ denote the price of an "asset" at time t. Suppose first that the asset is risk free and interest is paid on the asset continuously with an interest rate r. Then the change of X is proportional to the value of X and the proportionality constant is exactly r. Thus, $X(t)$ is a solution of the (deterministic) equation

$$X'(t) = rX(t).$$

Now suppose that the interest rate r is itself a random variable. A simple model could be that r consists of a fixed number μ (the *expected* interest rate) plus a multiple σ of a "noise"-term Ξ which has zero expectation. Here σ measures the size of the noise. Thus, our ODE becomes

$$X'(t) = \mu X(t) + \sigma X(t)\Xi(t).$$

The question is, what is a good choice for Ξ. One of the simplest choices is to take "white noise", i.e. $\Xi = \frac{dW(t)}{dt}$. Inserting this and multiplying with dt, we obtain the stochastic differential equation

$$dX(t) = \mu X(t)dt + \sigma X(t)dW(t). \tag{4.2}$$

Note that this equation has Lipschitz continuous coefficients $f(t,x) = \mu x$ and $\sigma(t,x) = \sigma x$. Hence, by Theorem 4.1.1 it has a unique solution for every initial datum η. For simplicity, let us take $\eta = 1$, i.e. we normalize the price of our asset at time 0. How can we *compute* a solution to this equation?

We try to separate the variables. Dividing by X and integrating, we find that formally

$$\int_0^t \frac{dX(s)}{X(s)} = \int_0^t \mu ds + \int_0^t \sigma dW(s) = \mu t + \sigma W(t)$$

as $W(0) = 0$. It remains to compute the integral on the left. We "guess" that the integral should somehow involve $\log X(t)$. By Itô's formula

$$\begin{aligned} d\log(X(t)) &= +\frac{1}{X(t)}dX(t) + \frac{1}{2}(-\frac{1}{X(t)^2})d\langle X\rangle_t \\ &= \frac{dX(t)}{X(t)} - \frac{1}{2X(t)^2}\sigma^2 X(t)^2 dt = \frac{dX(t)}{X(t)} - \frac{1}{2}\sigma^2 t \end{aligned}$$

since $\langle X\rangle_t = \langle(\sigma X)\diamond W\rangle_t = \int_0^t \sigma^2 X(s)^2 dt$, hence $d\langle X\rangle_t = \sigma^2 X(t)^2 dt$. Combining, we find

$$\log X(t) = \log\eta + \mu t + \sigma W(t) - \frac{1}{2}\sigma^2 t$$

and thus

$$X(t) = e^{(\mu - \frac{\sigma^2}{2})t + \sigma W(t))}.$$

Note that we cannot apply Itô's formula with log, as this function cannot be defined on the whole real line. However, applying Itô's formula with $f(t,x) = \exp((\mu - \frac{\sigma^2}{2})t + \sigma x)$ to the semimartingale $(t, W(t))$ we find that indeed the above process solves our stochastic differential equation.

We should note that the above equation is nothing but a model for the price of an asset. The solution depends on our interpretation of the stochastic differential equation. Indeed, if instead of the Itô integral we would have used the Stratonovich integral (which we merely glanced at in Exercise (3) of Chapter 1), the solution would have been $X(t) = e^{\mu t + \sigma W(t)}$, which might be closer to the what one would have guessed as a solution. We note that in our (Itô-)solution above, we again have an Itô correction term $e^{-\frac{\sigma^2}{2}t}$.

In applications, it is also worth discussing whether the noise should be given by a Brownian motion, hence whether it is appropriate to use the stochastic differential $dW(t)$. More generally, we could use $dZ(t)$ where $Z(t)$ is some other stochastic process. We do not discuss such modeling issues here.

4.3.2. Linear equations. Similar to the Ornstein-Uhlenbeck equation in Example 1.2.6, we can consider general linear equations. Indeed, if $A \in \mathbb{R}^{d\times d}$ and $B \in \mathbb{R}^{d\times m}$ we may consider the general linear equation

$$\begin{cases} dX(t) &= AX(t)dt + BdW(t) \\ X(0) &= x_0 \end{cases}$$

where W is an m-dimensional Brownian motion. By Theorem 4.1.1, for every $x_0 \in \mathbb{R}^d$, there exists a unique solution $X(t)$ of this equation. Note that being a solution means that

$$X(t) = x_0 + \int_0^t AX(s)\,ds + \int_0^t BdW(s) = x_0 + \int_0^t AX(s)\,ds + B\begin{pmatrix} W_1(t) \\ \vdots \\ W_m(t) \end{pmatrix}.$$

Writing $S(t) := e^{tA}$, the matrix exponential function generated by A, we expect that the variation of constants formula holds, so that the solution is in fact given by

$$X(t) = S(t)x_0 + \int_0^t S(t-s)B\,dW(s)\,.$$

We leave the verification of this as an exercise to the reader

As for ordinary differential equations, we can rewrite higher order equations as systems of equations of order one. As is well-known, the second order equation $y'' = f(y)$ describes the movement of a point in a force field which only depends on the location. This equation could be disturbed by noise, i.e. $y'' = f(y) + \text{noise}$. Let us for simplicity consider the function $f(y) = -y$ which corresponds to the harmonic oscillator. We model the noise as before as $\frac{dW(t)}{dt}$, where W is a one-dimensional Brownian motion. Setting $X_1(t) = y(t)$ and $X_2(t) = y'(t)$, the resulting equation can be rewritten as the system

$$dX(t) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} X(t)dt + \begin{pmatrix} 0 \\ 1 \end{pmatrix} dW(t)$$

It is well-known that

$$\exp t \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}.$$

Consequently, by the variation of constants formula,

$$X(t) = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} x_0 + \int_0^t \begin{pmatrix} \sin(t-s) \\ \cos(t-s) \end{pmatrix} dW(s)$$

Using the integration by parts formula, we see that

$$\int_0^t \begin{pmatrix} \sin(t-s) \\ \cos(t-s) \end{pmatrix} dW(s) = \begin{pmatrix} \int_0^t \cos(t-s)W(s)\,ds \\ W(t) - \int_0^t \sin(t-s)W(s)\,ds \end{pmatrix}$$

4.3.3. Integrating Factors. An important technique in solving ordinary differential equations is to multiply with an "integrating factor" so that the differential equation becomes exact. In this section, we see that for equations of the form

$$dX(t) = f(X(t))dt + \sigma X(t)dW(t)$$

we can find an integrating factor which transforms the stochastic differential equation into an ordinary differential equation with random coefficients.

To be more precisely, let $F(t) := \exp(-\sigma W(t) + \frac{\sigma^2}{2}t)$. By Itô's formula (or Section 4.3.1), we find

$$dF(t) = \sigma^2 F(t)dt - \sigma F(t)dW(t)\,.$$

Thus, with integration by parts, we have

$$\begin{aligned} d(F(t)X(t)) &= F(t)dX(t) + X(t)dF(t) + \langle F(t), X(t)\rangle \\ &= F(t)f(X(t))dt + \sigma F(t)X(t)dW(t) \\ &\quad +\sigma^2 F(t)X(t)dt - \sigma F(t)W(t)dW(t) \\ &\quad -\sigma^2 F(t)X(t)dt \\ &= F(t)f(X(t))dt\,. \end{aligned}$$

Here used the chain rule and the characterization of the stochastic integral in Theorem 3.1.4 which yields $\langle X, F\rangle_t = -\sigma^2 \int_0^t F(t)X(t)\,dt$. Consequently, if we write $Y(t) := F(t)X(t)$, then

$$dY(t) = F(t)f(F(t)^{-1}Y(t))dt$$

that is, Y solves an *ordinary* differential equation with random coefficients. Solving this equation pathwise, we can compute the solution X of our given equation via $X(t) = F(t)Y(t)$. Some applications of this technique will be presented in the exercises.

4.4. Exercises

(1) Let $\phi : [0,T] \to \mathbb{R}^{d\times m}$. For an m-dimensional Brownian motion W, we write, as before, $\int_0^T \phi(s)\,dW(s)$ for the vector with components

$$\Big(\int_0^T \phi(s)\,dW(s)\Big)_i = \sum_{j=1}^m \int_0^T \phi_{ij}(s)\,dW_j(s)\,.$$

Use Itô's formula to prove the integration by parts formula

$$\int_0^T \phi(s)\,dW(s) = \phi(T)W(T) - \int_0^T \phi'(s)W(s)\,ds.$$

Here, differentiation is component wise and the last integral is also understood component wise.

Hint: Try $f(t,x) = \phi(t)x$ for $t \in [0,T]$ and $x \in \mathbb{R}^m$.

Then use the integration by parts formula to prove the variation of constants formula for the linear stochastic differential equation.

Hint: First apply Itô's formula to determine $d(e^{-At}X(t))$ where X is the solution.

(2) Use an integrating factor to determine a solution of the following stochastic differential equations

(a) $dX(t) = adt + \sigma X(t)dW(t)$.

(b) $dX(t) = X^\gamma(t)dt + \sigma X(t)dW(t)$. Here, $X(0) = x_0 > 0$. For which γ do the solutions explode?

(3) It is interesting to extend the existence result from Proposition 4.2.4 to more general coefficients f and σ which are locally Lipschitz continuous but not necessarily of linear growth. In this exercise we take a look at an important technique using *Lyapunov functions*. For simplicity, we restrict ourselves to the case $d = m = 1$ and coefficients f, σ which do not depend on the time, i.e. $f, \sigma : \mathbb{R} \to \mathbb{R}$. We also restrict ourselves to deterministic initial data $\eta \in \mathbb{R}$.

As in the proof of Proposition 4.2.4, we define f_n and σ_n by freezing the functions f resp. σ outside the ball of radius n and denote by X_n the unique solutions of the equation with coefficients f_n and σ_n and initial datum η.

(a) Let $\tau_n := \inf\{t \in [0,T] : |X_n(t)| > n\}$. Change the proof of Proposition 4.2.4 to show that if

$$\mathbb{P}(\tau_n \le t) \to 0 \quad \text{as } n \to \infty \quad \forall\, t \in [0,T]\,, \tag{4.3}$$

then there exists a solution to the equation with coefficients f and σ and initial datum η.

To verify (4.3), we make use of *Lyapunov functions*. Given f and σ, a Lyapunov function for our equation is a function $V \in C^2(\mathbb{R})$ such that

(i) $V(x) \ge 0$ for all $x \in \mathbb{R}$.

(ii) $\lim_{|x|\to\infty} V(x) = \infty$.

(iii) $\mathscr{A}_{f,\sigma}V(x) := \frac{1}{2}\sigma^2(x)V''(x) + V'(x)f(x) \le \kappa V(x)$ for a certain $\kappa > 0$.

We now assume that there exists a Lyapunov function for our equation.

(b) Let $V_n := V\phi_n$, where $0 \le \phi_n \in C_c^\infty(\mathbb{R})$ is such that $\phi_n(x) = 1$ whenever $|x| \le n$ and $\phi_n(x) = 0$ whenever $|x| \ge 2n$. Moreover $\phi_n \le 1$. Apply Itô's formula for the function V_n and the stopped process X^{τ_n} and use the Gronwall lemma to infer that

$$\mathbb{E}V_n(X(t\wedge\tau_n)) \le e^{\kappa t}V_n(\eta)$$

whenever $n \ge |\eta|$.

(c) Derive from (b) that

$$\inf_{|x|=n} V(x)\mathbb{P}(\tau_n \le t) \le e^{\kappa T} V(\eta)$$

and infer (4.3) from this.

(4) Find Lyapunov functions for the stochastic differential equations with coefficients f and σ, where
   (a) $f(t) = t - t^3$ and $\sigma(t) = t$.
   (b) $f(t) = t$ and $\sigma(t) = t^2$.

CHAPTER 5

Yamada-Watanabe Theory

5.1. Different notions of existence and uniqueness

In the last chapter, we have discussed existence and uniqueness of stochastic differential equations with (locally) Lipschitz continuous coefficients. The solutions were defined on a given probability space and with respect to a given Brownian motion; uniqueness meant that if X and Y are two solutions, then $\mathbb{P}(X(t) = Y(t) : t \in [0,T]) = 1$.

However, for some applications these notions of existence and uniqueness are too narrow. We define here more general concepts. Throughout, we consider the stochastic differential equation

$$\left\{ \begin{array}{lcl} dX(t) & = & f(t,X(t))dt + \sigma(t,X(t))dW(t) \\ X(0) & = & \eta \end{array} \right. \tag{5.1}$$

We will assume that $f : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : [0,T] \times \mathbb{R}^{d\times m}$ are measurable; W is interpreted as an m-dimensional Brownian motion. In contrast to the last section, η is assumed to be a vector in $\mathbb{R}^d$, it is *not* random.

Definition 5.1.1. We will say that (5.1) admits a *strong solution* if whenever $(\Omega, \Sigma, \mathbb{F}, \mathbb{P})$ is a stochastic basis on which an m-dimensional $\mathbb{F}$-Brownian motion W is defined, we find an $\mathbb{F}$-adapted, continuous process $X : \Omega \to \mathbb{R}^d$ such that

(1) $\mathbb{P}(X(0) = \eta) = 1$.
(2) $\mathbb{P}(\int_0^T |f_i(s,X(s))| + |\sigma_{ij}(s,X(s))|^2\, ds < \infty) = 1$ for all $1 \le i \le d$ and $1 \le j \le m$.
(3) The integrated version of (5.1) holds true, i.e. almost surely

$$X(t) = \eta + \int_0^t f(s,X(s))\, ds + \int_0^t \sigma(s,X(s))\, dW(s) \quad \forall\, t \in [0,T].$$

Here, the deterministic integral is defined pathwise, the stochastic integral is Itô's integral. Note that both are well-defined in view of condition (2).

This is exactly Definition 4.2.1. Note that here we require that we can find such a solution on *every* stochastic basis carrying a Wiener process and that it is a solution with respect to that given Wiener process.

For a *weak solution*, we make the stochastic basis a part of the solution thus, we cannot allow random initial data, as we are not given a priori a probability space. It is, however, possible to generalize this definition by prescribing a certain initial *distribution*. We will not go into details at this moment.

Definition 5.1.2. A *weak solution of* (5.1) is a tupel $(\Omega, \Sigma, \mathbb{F}, \mathbb{P}, W, X)$, consisting of a filtered probability space $(\Omega, \Sigma, \mathbb{F}, \mathbb{P})$, an m-dimensional $\mathbb{F}$-Brownian motion W on that probability space and a continuous, $\mathbb{F}$-adapted process Omega, X defined on $(\Omega, \Sigma, \mathbb{F})$ such that (1), (2) and (3) of Definition 5.1.1 hold.

Obviously, every strong solution is also a weak solution and a weak solution is a strong solution on the stochastic basis and relative to the Brownian motion which is part of the solution.

We will see below that a stochastic differential equation may have weak solutions but not strong solutions. Next, we also introduce two uniqueness concepts. The first one is that of *pathwise uniqueness* which we have used so far.

DEFINITION 5.1.3. We say that *pathwise uniqueness* (or *strong uniqueness*) holds for equation (5.1) if whenever X and Y are (strong) solutions defined on the same stochastic basis and with respect to the same Brownian motion[1] with $X(0) = Y(0)$, then $\mathbb{P}(X(t) = Y(t) : \forall\, t \in [0,T]) = 1$.

As for general weak solutions, how could they be unique? Certainly, the stochastic basis could differ from one solution to another. This of course implies that (being maps defined on that basis) the Brownian motion W and the process X will differ from one solution to another. However, often the sole interest in a solution lies in its *distribution*.

DEFINITION 5.1.4. We say that *uniqueness in law* (or *weak uniqueness*) holds for equation (5.1) if whenever X and Y are two weak solutions[2] starting at the same $\eta \in \mathbb{R}^d$, the laws of X and Y as $C([0,T])$-valued random variables are the same.

EXAMPLE 5.1.5. (Tanaka)

We let $\operatorname{sgn} x = \mathbb{1}_{(0,\infty)} - \mathbb{1}_{(-\infty,0]}$ and consider the one-dimensional stochastic differential equation

$$dX(t) = \operatorname{sgn} X(t) dW(t)\,.$$

Let us first discuss uniqueness of equations. If this equation has a solution with initial datum $\eta \in \mathbb{R}$, then $X(t) = \eta + \int_0^t \operatorname{sgn} X(s)\, dW(s)$. Consequently, $X - \eta$ is a local martingale with quadratic variation

$$\langle X - \eta\rangle_t = \int_0^t \big[\operatorname{sgn} X(s)\big]^2\, ds = \int_0^t s\, ds = t\,.$$

Thus, by Theorem 3.3.1, $X - \eta$ is a Brownian motion. This implies that uniqueness in law holds for Tanaka's equation.

On the other hand, if X is a solution starting at 0 on the stochastic basis $(\Omega, \Sigma, \mathbb{F}, \mathbb{P})$ with respect to the Brownian motion W, then also $-X$ is a solution on the same stochastic basis and with respect to the same Brownian motion. Consequently, *if* there exists solutions, they cannot be pathwise unique.

Let us now turn to existence. We start with a stochastic basis $(\Omega, \Sigma, \mathbb{F}, \mathbb{P})$ on which an $\mathbb{F}$-Brownian motion X (yes, this Brownian motion will be our solution!) is defined. To be definite, take the one constructed in Chapter 1. Now, we define W by $\int_0^t \operatorname{sgn} X(s)\, dX(s)$. By the above, W is a Brownian motion defined on our stochastic basis. With the chain rule from Proposition 3.2.6, we see that

$$\int_0^t \operatorname{sgn}(X(s))\, dW(s) = \int_0^t \operatorname{sgn}(X(s))\operatorname{sgn}(X(s))\, dX(s) = \int_0^t 1\, dX(s) = X(t).$$

Hence, $(\Omega, \Sigma, \mathbb{F}, \mathbb{P}, W, X)$ is a weak solution of Tanaka's equation starting at 0. To obtain solutions starting at a general point $\eta \in \mathbb{R}$, we repeat this construction with X replaced with $X + \eta$. This shows that there exist weak solutions of Tanaka's equation.

However, strong solutions do not exist. Indeed, given a Brownian motion W on a probability space $(\Omega, \Sigma, \mathbb{P})$, we endow the probability space with the filtration $\mathbb{F}^W$ generated by W. Now suppose we had a solution X starting at 0 on the stochastic basis $(\Omega, \Sigma, \mathbb{F}^W, \mathbb{P})$

[1] Equivalently, if $(\Omega, \Sigma, \mathbb{F}, \mathbb{P})$ is a stochastic basis on which an m-dimensional $\mathbb{F}$-Brownian motion W is defined and $(\Omega, \Sigma, \mathbb{F}, \mathbb{P}, W, X)$ and $(\Omega, \Sigma, \mathbb{F}, \mathbb{P}, W, Y)$ are weak solutions of (5.1)

[2] Here, as is customary, we do not mention explicitly the stochastic basis on which the processes are defined and the Brownian motion with respect to which we have a solution. It would be more accurate to say "whenever $(\Omega, \Sigma, \mathbb{F}, \mathbb{P}, W, X)$ and $(\tilde\Omega, \tilde\Sigma, \tilde{\mathbb{F}}, \tilde{\mathbb{P}}, \tilde W, Y)$ are solutions". However, we want to avoid such sentences.

with respect to the Brownian motion W. Then X is a Brownian motion. Using again the chain rule and Tanaka's formula (3.6), we find that

$$W(t) = \int_0^t \operatorname{sgn}(X(s))\, dX(s) = |X(t)| + L(t)\,.$$

As the right-hand side is $\sigma(|X(s)| : s \leq t) =: \mathscr{F}_t^{|X|}$-measurable, we find that $\mathbb{F}^W \subset \mathbb{F}^{|X|}$ and consequently, since X is $\mathbb{F}^W$-adapted, $\mathbb{F}^X \subset \mathbb{F}^{|X|}$, which is absurd. Hence, a process X as above cannot exist and we see that indeed, we cannot expect strong existence of solutions.

Tanaka's example shows that we can have weak existence without strong existence and weak uniqueness without strong uniqueness. Several questions are natural and immediate:

- Does strong uniqueness imply weak uniqueness?
- Does weak existence and strong uniqueness imply strong existence?

Such, and related question are the topic of this chapter. We call this Yamada-Watanabe theory, as such questions were studied by Yamada and Watanabe in their seminal papers [**9, 10**].

5.2. On strong and weak uniqueness

In this section, we address the first question raised above, whether strong uniqueness implies weak uniqueness. We shall prove that this is indeed the case.

The main difficulty to overcome is that we have to prove something for solutions defined on possibly *different* probability spaces, while we only know something about solutions defined on *the same* probability space. Thus, if we want to use strong uniqueness to prove weak uniqueness, we have to transfer solutions onto a common probability space. This will be done by means of *regular conditional probability*

Definition 5.2.1. Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space and $\mathscr{F}$ be a sub-σ-algebra of Σ. A *regular conditional probability for* Σ *given* $\mathscr{F}$ is a map $q : \Omega \times \mathscr{G} \to [0,1]$ such that

(1) for every $\omega \in \Omega$, $q(\omega, \cdot)$ is a probability measure on (Ω, Σ).
(2) for every $A \in \Sigma$, the map $\omega \mapsto q(\omega, A)$ is $\mathscr{F}$-measurable.
(3) for every $A \in \Sigma$, we have $q(\omega, A) = \mathbb{E}[\mathbb{1}_A|\mathscr{F}]$ almost surely.

Thus, in a way, regular conditional probability is a version of the conditional probability $\mathbb{P}(A|\mathscr{F}) := \mathbb{E}[\mathbb{1}_A|\mathscr{F}]$. Here, we have to make a choice for every $A \in \Sigma$ (that's quite a lot) in such a way, that whenever A_k is a sequence of disjoint sets in σ, we have $q(\omega, \bigcup_k A_k) = \sum_k q(\omega, A_k)$ for all $\omega \in \Omega$. It is not clear at all whether this is possible.

Theorem 5.2.2. *Let (E, d) be a complete, separable metric space, $\mathbb{P}$ be a probability measure on the Borel σ-algebra $\mathscr{B}(E)$ and $\mathscr{F}$ be a sub-σ-algebra of $\mathscr{B}(E)$. Then there exists a regular conditional probability for $\mathscr{B}(E)$ given $\mathscr{F}$.*

Proof. We proceed in several steps.

Step 1: Preparation.

Let $\{x_n : n \in \mathbb{N}\}$ be a countable, dense subset of E and let $\mathscr{B}$ be the collection of all open balls, centered at some x_n with rational radius. Then $\mathscr{B}$ is a countable set which generates the Borel σ-algebra. We denote by $\mathscr{R}$ the ring generated by $\mathscr{B}$, i.e.

$$\mathscr{R} = \Big\{ \cup_{k=1}^n \cap_{j=1}^m A_{kj} \,:\, n, m \in \mathbb{N}, A_{kj} \in \mathscr{B} \text{ or } A_{kj}^c \in \mathscr{B}\}$$

Then $\mathscr{R}$ is a countable set generating the Borel σ-algebra. Adding $\emptyset$ to $\mathscr{B}$, we may, and shall, assume that $E \in \mathscr{R}$.

Since $\mathbb{P}$ is a regular measure, see Lemma D.2, for every $A \in \mathscr{R}$, there exists a sequence $(K_j^A)_{j\in\mathbb{N}}$ of compact sets contained in A such that $\mathbb{P}(A) = \sup_j \mathbb{P}(K_j^A)$. We let $\mathscr{R}^*$ be the ring generated by $\mathscr{R}$ and all the K_j^A's. Then $\mathscr{R}^*$ is a countable ring, generating the Borel σ-algebra.

Step 2 : Preliminary selection.

For every $R \in \mathscr{R}^*$, we choose a version $q_0(x, R)$ of $\mathbb{E}[\mathbb{1}_R|\mathscr{F}]$. Note that $q_0(\cdot, R)$ is $\mathscr{F}$-measurable. By linearity of conditional expectation, if $R_1, \ldots, R_n$ are pairwise disjoint sets of $\mathscr{R}^*$, we have

$$q_0(x, R_1 \cup \cdots \cup R_n) = \sum_{j=1}^{n} q_0(x, R_j)$$

for $\mathbb{P}$-almost every x. Since $\mathscr{R}^*$ is countable, there are only countably many such finite families in $\mathscr{R}^*$. Consequently, we find a set $N_0 \in \mathscr{F}$ with $\mathbb{P}(N_0) = 0$ such that $q_0(x, \cdot)$ is a finitely additive set function on $E \setminus N_0$.

We now enlarge N_0 somewhat. By construction, for $R \in \mathscr{R}$, we have $\mathbb{1}_{K_j^R} \uparrow \mathbb{1}_R$. By the properties of conditional expectation, $q_0(x, K_j^R) \uparrow q_0(x, R)$ almost surely. Thus, we find $N_R \in \mathscr{F}$ with $\mathbb{P}(N_R) = 0$ such that $q_0(x, K_j^R) \uparrow q_0(x, R)$ for all $x \in E \setminus N_R$. Finally, let $N_\infty \in \mathscr{F}$ be a null set such that $q_0(x, E) = 1$ on $\Omega \setminus N_\infty$.

Define $N := N_0 \cup N_\infty \cup \bigcup_{R \in \mathscr{R}} N_R$. Then $\mathbb{P}(N) = 0$ and for $x \in E \setminus N$ we have that $q_0(x, \cdot)$ is a finitely additive set function with total mass 1 such that $q_0(x, R) = \sup_j q_0(x, K_j^R)$ for all $R \in \mathscr{R}$.

Step 3: We extend $q_0(x, \cdot)$ to a measure $q(x, \cdot)$ on $\mathscr{B}(E)$ for $x \in E \setminus N$.

Fix $x \in E \setminus N$ and let A_n be a sequence in $\mathscr{R}$ with $A_n \downarrow \emptyset$. Given $\varepsilon > 0$ for each n we find a compact set C_n (a suitable $K_j^{A_n}$) in $\mathscr{R}^*$ with $C_n \subset A_n$ and $q_0(x, A_n \setminus C_n) \leq 2^{-n}\varepsilon$. As $A_n \downarrow \emptyset$, we have that $\bigcap C_n = \emptyset$. As the sets C_n are compact, there exists some n_0 such that $C_1 \cap \cdots \cap C_{n_0} = \emptyset$. Consequently, $\bigcup_{n=1}^{n_0} A_n \setminus C_n = \bigcup_{n=1}^{n_0} A_n \cap \big(\bigcup_{n=1}^{n_0} C_n\big)^c = \bigcup_{n=1}^{n_0} A_n$, so that $A_{n_0} \subset \bigcup_{n=1}^{n_0} A_n \setminus C_n$. Thus, since $q_0(x, \cdot)$ is finitely additive on $\mathscr{R}^*$,

$$q_0(x, A_{n_0}) \leq \sum_{n=1}^{n_0} q_0(x, A_n \setminus C_n) \leq \sum_{n=1}^{\infty} 2^{-n}\varepsilon = \varepsilon.$$

It follows that $q_0(x, A_n) \downarrow 0$, hence $q_0(x, \cdot)$ is a pre-measure. By the Caratheodory extension theorem, there exists a unique measure $q(x, \cdot)$ on $\mathscr{B}(E)$ such that $q(x, R) = q_0(x, R)$ for all $R \in \mathscr{R}$. We finally put $q(x, \cdot) = \delta_x$ for $x \in N$.

Step 4: We show that q is the sought-after regular conditional probability.

Condition (1) in the definition is true by construction. For (2), let us show that $x \mapsto q(x, A)$ is $\mathscr{F}$-measurable for all $A \in \Sigma$. Indeed, the sets of all A for which this is true is easily seen to be a Dynkin system. Moreover, it contains $\mathscr{R}$ (observe: $N \in \mathscr{F}$). Thus, by Dynkin's π - λ theorem, this is indeed true for all $A \in \Sigma$.

Finally, for $R \in \mathscr{R}$ we have $q(x, R) = \mathbb{E}[\mathbb{1}_R|\mathscr{F}]$ almost surely by construction. It is straightforward to prove that the sets R for which this is true is a Dynkin system. Invoking Dynkin's π - λ theorem again, it is true for all $R \in \mathscr{B}(E)$ hence (3). □

In the case where $\mathscr{F}$ is generated by a random element, we get a related result:

COROLLARY 5.2.3. *Let $(\Omega, \Sigma, \mathbb{P})$ be as in Theorem 5.2.2, and X be a measurable map from (Ω, Σ) to a measurable space $(S, \mathscr{S})$. The distribution of X is denoted by μ_X. Then there exists a function $q : S \times \Sigma \to [0, 1]$ called* regular conditional probability of Σ given X, *such that*

(1) *for every $x \in S$, $q(x, \cdot)$ is a probability measure on (Ω, Σ).*
(2) *for every $A \in \Sigma$, the map $x \mapsto q(x, A)$ is $\mathscr{S}$-measurable.*
(3) *for each $A \in \Sigma$, we have $q(x, A) = \mathbb{P}(A|X = x)$, μ_X-a.e. This should be interpreted as*

$$q(X, A) = \mathbb{E}[\mathbb{1}_A|X] \quad \mathbb{P}\, a.e.$$

Hence, the $\sigma(X)$-measurable random variable $q(X, A)$ is a version of conditional expectation of $\mathbb{1}_A$ given X.

We are now ready to prove that pathwise uniqueness implies uniqueness in law. To bring two solutions, defined on possibly different probability spaces, together, we use regular conditional probability.

Now let $(\Omega_j, \Sigma_j, \mathbb{P}_j, W_j, X_j)$ be two weak solutions of equation (5.1) such that $X_1(0)$ has the same distribution μ as $X_2(0)$. We set $Y_j := X_j - X_j(0)$ and regard solution j as consisting of the parts $X_j(0), W_j$ and X_j. For $j = 1, 2$, this triple induces a measure $\mathbf{P}_j$ on

$$E_* := \mathbb{R}^d \times C([0,T];\mathbb{R}^m) \times C([0,T];\mathbb{R}^d),$$

endowed with the σ-algebra

$$\Sigma_* := \mathscr{B}(\mathbb{R}^d) \otimes \mathscr{B}(C([0,T];\mathbb{R}^m)) \otimes \mathscr{B}(C([0,T];\mathbb{R}^d))$$

via

$$\mathbf{P}_j(A) := \mathbb{P}_j((X_j(0), W_j, Y_j) \in A) \quad \forall\, A \in \Sigma.$$

Note that as a product of three complete, separable metric spaces, E_* is itself a complete separable metric space and Σ_* is in fact the Borel σ-algebra of E_*. We write $\theta_* = (x, \mathbf{w}, \mathbf{y})$ for a generic element of E_*.

Now let q_j be a regular conditional probability of Σ_* given $(x, \mathbf{w})$. In fact, we will only be interested in the regular conditional probability of sets of the form $\mathbb{R}^d \times C([0,T];\mathbb{R}^m) \times A$. By slight abuse of notation we will thus write $q_j(x, \mathbf{w}, A)$ instead of $q_j(x, \mathbf{w}, \mathbb{R}^d \times C([0,T];\mathbb{R}^m) \times A$ and consider q_j as a map from $\mathbb{R}^d \times C([0,T];\mathbb{R}^m) \times \mathscr{B}(C([0,T];\mathbb{R}^d)) \to [0,1]$.

Note that the push-forward of $\mathbb{P}_j$ under $(X_j(0), W_j)$, i.e. the distribution of $(x, \mathbf{w})$, is the measure $\mu \otimes \mathbb{W}$, where $\mathbb{W}$ is m-dimensional Wiener measure, on $\mathscr{B}(\mathbb{R}^d) \otimes \mathscr{B}(E([0,T];\mathbb{R}^m))$. In particular, it is independent of j. We write $\mathbf{P}_0$ for this measure. We obtain for $A \in \mathscr{B}(\mathbb{R}^d) \otimes \mathscr{B}(C([0,T];\mathbb{R}^m))$ and $B \in \mathscr{B}(C([0,T];\mathbb{R}^d))$ that

$$\mathbf{P}_j(A \times B) = \int_A q_j(x, \mathbf{w}, B)\, d\mathbf{P}_0(x, \mathbf{w})\,.$$

We can now glue things together. We add an additional copy of $C([0,T];\mathbb{R}^d)$ to E to make room for *two* $\mathbf{y}$'s. Thus, we put $E := E \times C([0,T];\mathbb{R}^d)$ and $\Sigma := \Sigma_* \otimes \mathscr{B}(C([0,T];\mathbb{R}^d)$. We write $\theta = (x, \mathbf{w}, \mathbf{y}_1, \mathbf{y}_2)$ for a generic element of E. On (E, Σ), we define a probability measure $\mathbf{P}$ by setting for $A \in \mathscr{B}(\mathbb{R}^d) \otimes \mathscr{B}(C([0,T];\mathbb{R}^m))$ and $B, C \in \mathscr{B}(C([0,T];\mathbb{R}^d))$

$$\mathbf{P}(A \times B \times C) = \int_A q_1(x, \mathbf{w}, B) q_2(x, \mathbf{w}, C)\, d\mathbf{P}_0(x, \mathbf{w})\,.$$

We note that for the marginals, we have $\mathbf{P}(A \times B \times C([0,T];\mathbb{R}^d)) = \mathbf{P}_1(A \times B)$ and $\mathbf{P}(A \times C([0,T];\mathbb{R}^d) \times C) = \mathbf{P}_2(A \times C)$. In particular, $(x, \mathbf{w}, \mathbf{y}_1)$ is a "distributional copy" of $(X_1(0), Y_1, W_1)$ and $(x, \mathbf{w}, \mathbf{y}_2)$ is a distributional copy of $(X_2(0), Y_2, W_2)$. We also define $\mathscr{F}_t := \sigma(x, \mathbf{w}(s), \mathbf{y}_(s), \mathbf{y}_2(s) \,:\, s \leq t)$ and $\mathbb{F} := (\mathscr{F}_t))_{t \in [0,T]}$.

Lemma 5.2.4. *With the definitions above,* $(E, \Sigma, \mathbb{F}, \mathbf{P}, \mathbf{w}, x + \mathbf{y}_1)$ *and* $(E, \Sigma, \mathbb{F}, \mathbb{P}, \mathbf{w}, x + \mathbf{y}_2)$ *are weak solutions of equation* (5.1)*.*

Taking Lemma 5.2.4 for granted (we will indicate a proof in the exercises), we can now prove the following result.

Theorem 5.2.5. *(Yamada and Watanabe)*
Pathwise uniqueness implies uniqueness in law.

Proof. Given two weak solutions $(\Omega_j, \Sigma_j, \mathbb{P}_j, W_j, X_j)$ for $j = 1, 2$, we have constructed two weak solutions $x + \mathbf{y}_j$ on a single stochastic basis $(E, \Sigma, \mathbb{F}, \mathbf{P})$ and with respect to the same Brownian motion $\mathbf{w}$. Now pathwise uniqueness implies

$$\mathbf{P}(x + \mathbf{y}_1 = x + \mathbf{y}_2 : \forall\, t \in [0,T]) = 1$$

or, equivalently,

$$\mathbf{P}(\{(x,\mathbf{w},\mathbf{y}_1,\mathbf{y}_2)\,:\,\mathbf{y}_1=\mathbf{y}_2\})=1. \tag{5.2}$$

Consequently, for $A\in\mathscr{B}(\mathbb{R})\otimes\mathscr{B}(C([0,T];\mathbb{R}^m))$ and $B\in\mathscr{B}(C([0,T];\mathbb{R}^d))$, we have

$$\begin{aligned}
\mathbb{P}_1((X_1(0),W,Y_1)\in A\times B) &= \mathbf{P}_1((x,\mathbf{w},\mathbf{y}_1)\in A\times B)\\
&= \mathbf{P}((x,\mathbf{w},\mathbf{y}_1,\mathbf{y}_2)\in A\times B\times C([0,T];\mathbb{R}^d))\\
&\overset{(5.2)}{=} \mathbf{P}((x,\mathbf{w},\mathbf{y}_1,\mathbf{y}_2)\in A\times B\times B)\\
&= \mathbf{P}((x,\mathbf{w},\mathbf{y}_1,\mathbf{y}_2)\in A\times C([0,T];\mathbb{R}^d)\times B)\\
&= \mathbf{P}_1((x,\mathbf{w},\mathbf{y}_2)\in A\times B)\\
&= \mathbb{P}_2((X_2(0),W_2,Y_2)\in A\times B)\,.
\end{aligned}$$

This is uniqueness in law. □

It is a remarkable Corollary of the Yamada-Watanabe result that pathwise uniqueness implies strong existence. That is, if pathwise uniqueness holds for a certain equation and if there exist solutions (otherwise, there is nothing to prove anyway), then on every given stochastic basis with a given Brownian motion, there exists a solution with respect to that Brownian motion.

The basic strategy is to show that there exists a measurable function $k(x,\mathbf{w})$ such that

$$\mathbf{P}((x,\mathbf{w},\mathbf{y}_1,\mathbf{y}_2):y_1=y_2=k(x,\mathbf{w}))=1\,.$$

That is, $\mathbf{y}_j$ is a measurable function of x and $\mathbf{w}$. Then we set $h(x,\mathbf{w})=x+k(x,\mathbf{w})$. It follows that our initially given solutions $X_j=h(X_j(0),W_j)$ almost surely. Conversely, given a stochastic basis $(\Omega,\Sigma,\mathbb{F},\mathbb{P})$ on which an $\mathbb{F}$-Brownian motion W and an independent η with distribution μ is defined, we may define $X:=h(\eta,W)$. Then $(X_j(0),W_j)$ and (η,W) induce the same measure $\mathbf{P}_0=\mu\otimes\mathbb{W}$ on $\mathscr{B}(\mathbb{R}^d)\otimes\mathscr{B}(C([0,T]))$. However, on that space the process $X:=h(x,\mathbf{w})$ solves (5.1) with respect to $\mathbf{w}$. By a variant of Lemma 5.2.4, $X=h(\eta,W)$ solves that equation.

The details of the proof (in particular the construction of h and the verification of measurability conditions) are rather technical, so we do not go into details here and do not prove this result.

5.3. Pathwise uniqueness for some one-dimensional equations

Our approach to solve stochastic differential equations with (locally) Lipschitz continuous coefficients yields pathwise uniqueness, cf. Corollary 4.2.3. Hence, by the Yamada-Watanabe result, we also have uniqueness in law. In particular, the distribution of the solution does not depend on the stochastic basis on which we solve the equation. A natural question is, whether there are examples of stochastic differential equations with coefficients which are *not* (locally) Lipschitz continuous, yet still pathwise uniqueness holds for such equations.

In this section, we give a one-dimensional example, due to Yamada and Watanabe. At this moment, we merely prove that pathwise uniqueness holds. Existence of solutions will follow from the results of the next chapter.

THEOREM 5.3.1. *We consider the stochastic differential equation*

$$dX(t)=f(t,X(t))dt+\sigma(t,X(t))dW(t) \tag{5.3}$$

in dimension one, more precisely, $m=d=1$. *Moreover, assume that there exists a constant* L *and a strictly increasing function* $h:[0,\infty)\to[0,\infty)$ *with* $h(0)=0$ *and*

$$\int_{(0,\varepsilon)}h^{-2}(s)\,ds=\infty\quad\forall\,\varepsilon>0\,.$$

Finally, we assume that the coefficients f and σ satisfy

$$|f(t,x)-f(t,y)|\leq L|x-y| \quad \textit{and} \quad |\sigma(t,x)-\sigma(t,y)|\leq h(|x-y|) \qquad \forall\, x,y\in\mathbb{R}, t\in[0,T]\,.$$

Then pathwise uniqueness holds for equation (5.3).

Proof. By the assumption on h, there exists a strictly decreasing sequence a_n with $a_0 = 1$ and $a_n \downarrow 0$ such that $\int_{a_n}^{a_{n-1}} h^{-2}(s)\,ds = n$. We now pick for $n\in\mathbb{N}$ a continuous function φ_n supported in (a_n, a_{n-1}) with $0\leq\varphi_n(x)\leq\frac{2}{nh^2(x)}$ and $\int_0^\infty\varphi_n(x)\,dx = 1$. Then

$$\psi_n(x) := \int_0^{|x|}\int_0^t \varphi_n(s)\,ds\,dt$$

is an even, twice continuously differentiable function with $|\psi'(x)|\leq 1$ and $\psi_n(x)\uparrow|x|$, cf. Exercise 3 in Chapter 3.

Now let $(\Omega,\Sigma,\mathbb{F},\mathbb{P},W,X_j)$ be a solution of (5.3) for $j=1,2$. By stopping, it suffices to prove that $X_1=X_2$ almost surely additionally assuming that $\mathbb{E}\int_0^T|\sigma(t,X_j(t))|^2\,dt<\infty$.

We have

$$\Delta X(t) := X_1(t)-X_2(t) = \int_0^t f(s,X_1(s))-f(s,X_2(s))\,ds+\int_0^t \sigma(s,X_1(s))-\sigma(s,X_2(s))\,dW(s)\,.$$

By Itô's formula,

$$\begin{aligned}\psi_n(\Delta X(t)) &= \int_0^t \psi_n'(\Delta X(s))\big[f(s,X_1(s))-f(s,X_2(s))\big]\,ds\\ &\quad+\frac12\int_0^t\psi_n''(\Delta X(s))\big[\sigma(s,X_1(s))-\sigma(s,X_2(s))\big]^2\,ds\\ &\quad+\int_0^t\psi_n'(\Delta X(s))\big[\sigma(s,X_1(s))-\sigma(s,X_2(s))\big]\,dW(s)\,.\end{aligned}$$

By our additional assumption, the stochastic integral is a martingale, hence it has expectation zero. Taking expectations,

$$\begin{aligned}\mathbb{E}\psi_n(\Delta X(t)) &\leq \mathbb{E}\Big(\int_0^t|\psi_n'(\Delta X(s))||\Delta X(s)|\,ds+\frac12\int_0^t\phi_n(|\Delta X(s)|)h^2(|\Delta X(s)|)\,ds\Big)\\ &\leq \mathbb{E}\Big(\int_0^t 1\cdot L|\Delta X(s)|\,ds+\int_0^t\frac{1}{nh^2(|\Delta X(s)|)}h^2(|\Delta X(s)|)\,ds\Big)\\ &\leq \int_0^t L\mathbb{E}|\Delta X(s)|\,ds+\frac{t}{n}\,.\end{aligned}$$

Upon $n\to\infty$, we obtain $\mathbb{E}|\Delta X(t)|\leq L\int_0^t\mathbb{E}|\Delta X(s)|\,ds$ and, by Gronwall's lemma, it follows that $\mathbb{E}|\Delta X(t)|=0$ for all $t\in[0,T]$. Taking continuity of the paths into account, pathwise uniqueness follows. □

Typical examples for the function h are $h(t)=|t|^\alpha$ for $\alpha\geq\frac12$. This shows that we obtain pathwise uniqueness for functions σ which are Hölder continuous of order $\alpha\geq\frac12$.

Example 5.3.2. Consider the stochastic equation

$$\begin{cases} dX(t) &= |X(t)|^\alpha dW(t)\\ X(0) &= 0\end{cases}$$

Clearly, $X\equiv 0$ is a solution. By Theorem 5.3.1, this is the unique solution for $\alpha\geq\frac12$.

It can be shown, that for $\alpha\in(0,\frac12)$ we do not have pathwise uniqueness, not even uniqueness in distribution.

This should be compared with the deterministic equation $dX(t)=|X(t)|^\alpha dt$, or as a differential equation, $u'(t)=|u(t)|^\alpha$. With initial datum $u(0)=0$, the solution $u(0)=0$ is the only solution for $\alpha\geq 1$ (!). For $\alpha\in(0,1)$ also $u(t)=(1-\alpha)^{\frac{1}{1-\alpha}}t^{\frac{1}{1-\alpha}}$ is a solution.

EXAMPLE 5.3.3. A real-world example which fits into the situation of this section is the *Cox-Ingersoll-Ross* model for interest rates. According to it, the short rate r (= instantaneous interest rate) solves the stochastic differential equation

$$dr(t) = a(b - r(t))dt + \sigma\sqrt{|r(t)|}dW(t)$$

where a, b and σ are model parameter, $a, b, \sigma \geq 0$.

In our abstract framework, $f(t, x) = a(b-x)$ which is certainly Lipschitz continuous and $\sigma(t, x) = \sigma\sqrt{|x|}$ which is Hölder continuous of order $\frac{1}{2}$. Thus, pathwise uniqueness holds for this equation.

5.4. Exercises

(1) Find as many different strong solutions of the stochastic equation

$$dX(t) = 3X(t)^{\frac{1}{3}}dt + 3X(t)^{\frac{2}{3}}dW(t)$$

as possible. Here $x^{\frac{1}{3}} = -(-x)^{\frac{1}{3}}$ for $x < 0$ and $X^{\frac{2}{3}} = \big(X^{\frac{1}{3}}\big)^2$.

(2) The *Heston model* is a so-called stochastic volatility model. Let us again consider the model for an asset price S from Section 4.3.1:

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t)\,.$$

Here, μ and σ are constant model parameter. In stochastic volatility models, the parameter σ (which is a measure for the amount of noise in the process) is assumed to be itself a stochastic process. In the Heston model, we have the following equations:

$$\begin{cases} dS(t) &= \mu S(t)dt + \sqrt{\sigma(t)}S(t)\,dW_1(t) \\ d\sigma(t) &= \kappa(\theta - \sigma(t))dt + \rho\sqrt{\sigma(t)}\,dW_2(t). \end{cases}$$

Note that σ solves the Cox-Ingersoll-Ross equation.

Assuming that W_1 and W_2 are independent, this fits into our framework. Indeed, setting $d = m = 2$ and

$$f(t, x, y) := \begin{pmatrix} \mu x \\ \kappa(\theta - y) \end{pmatrix} \quad \text{and} \quad \sigma(t, x, y) := \begin{pmatrix} \sqrt{y}x & 0 \\ 0 & \rho\sqrt{y} \end{pmatrix}$$

this equation is of the form (5.1). More appropriately, we should replace y with $|y|$ in the square roots. However, as we will see in the next exercise, if the initial datum for σ is positive (and this is the case which is interesting for applications) the process is positive almost surely.

Show that we have pathwise uniqueness for the above equations.

Hint: The results of this section show that pathwise uniqueness holds for σ. Now prove pathwise uniqueness for the first equation as in Proposition 4.2.2.

In applications, often W_1 and W_2 are assumed to be correlated (thus not independent). What happens in this case?

(3) In this exercise, we prove a comparison result. Let f_1, f_2 and σ be given, such that they satisfy the assumptions of Theorem 5.3.1 and $f_1(t, x) \leq f_2(t, x)$ for all $(t, x) \in [0, T] \times \mathbb{R}$. Moreover, let X_j be a solution of the equation

$$\begin{cases} dX(t) &= f_j(t, X(t))dt + \sigma(t, X(t))dW(t) \\ X(0) &= \eta_j \end{cases}$$

for $j = 1, 2$. Prove that if $\eta_1 \leq \eta_2$ almost surely, then $\mathbb{P}(X_1(t) \leq X_2(t)\,\forall\, t \in [0, T]) = 1$.

Hint: If ψ_n is as in the proof of Theorem 5.3.1, put $\tilde{\psi}_n := \psi_n \mathbb{1}_{(0,\infty)}$ and repeat the proof of Theorem 5.3.1.

Use this result to prove that in the Cox-Ingersoll-Ross model respects positivity, i.e. we have $r(t) \geq 0$ whenever $r(0) \geq 0$.

(4) Prove Lemma 5.2.4. For simplicity, we only consider the case $m = d = 1$, the general case being similar.

We are given two filtered probability spaces $(\Omega_j, \Sigma_j, \mathbb{F}_j, \mathbb{P}_j)$ for $j = 1, 2$ and on the j-th space, an $\mathbb{F}_j$-Brownian motion W_j, a continuous, adapted process X_j and an $\mathscr{F}_0^{(j)}$-measurable random variable η_j. Moreover, we are given measurable functions $f, \sigma : \mathbb{R} \to \mathbb{R}$. We need to prove that if the distribution of (η_1, X_1, W_1) is the same as that of (η_2, X_2, W_2) (as $\mathbb{R} \times C([0,T]; \mathbb{R}^m) \times C([0,T]; \mathbb{R}^d)$-valued random variables), then one is a solution if and only if the other is a solution. That is, we need to prove that

$$\begin{aligned} &\mathbb{P}_1\Big[X_1(t) = \eta_1 + \int_0^t f(X_1(s))\,ds + \int_0^t \sigma(X_1(s))\,dW(s)\Big] \\ = \;&\mathbb{P}_2\Big[X_2(t) = \eta_2 + \int_0^t f(X_2(s))\,ds + \int_0^t \sigma(X_2(s))\,dW(s)\Big] \end{aligned}$$

along with similar statements concerning integrability of $f(X_j(\cdot))$ and stochastic integrability of $\sigma(X_j(\cdot))$. Proceed along the following steps:

(a) (Reduction)
Show that it suffices to prove that for every $t \in [0,T]$ the vectors

$$\begin{aligned} &\Big(\eta_1, \int_0^t f(X_1(s))\,ds, \int_0^t f(X_1(s))\,dW_j(s)\Big) \quad \text{and} \\ &\Big(\eta_2, \int_0^t f(X_2(s))\,ds, \int_0^t f(X_2(s))\,dW_2(s)\Big) \end{aligned} \tag{5.4}$$

have the same distribution. We may moreover assume that f and σ are bounded and that $\mathbb{P}_1(\|X_1\|_\infty \le c) = \mathbb{P}_2(\|X_2\|_\infty \le c) = 1$ for some $c > 0$.

(b) (Proof of the reduced statement for continuous f and σ)
Show that if f and σ are bounded and continuous (and the additional assumptions from (a) holds) then (5.4) holds. To that end, take partitions π_n of $(0,t)$ with $|\pi_n| \to 0$ and consider for both integrals Riemann sums for the partition (evaluate at the left end points!). Show that (5.4) holds if we replace the integrals with their Riemann sum approximation and finish by letting $n \to \infty$.

(c) (Proof of the reduced statement for measurable f and σ) Use the point above, approximation, a Dynkin system argument and linearity to extend to general measurable coefficients.

CHAPTER 6

Martingale Problems

6.1. The Martingale Problem associated to an SDE

Let us consider the general time homogeneous stochastic differential equation

$$\begin{cases} dX(t) &= f(X(t))dt + \sigma(X(t))dW(t) \\ X(0) &= \eta \end{cases} \tag{6.1}$$

with measurable coefficients $f : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$. Here, W is an m-dimensional Brownian motion. Again, η is an element of $\mathbb{R}^d$. In this chapter we focus on coefficients which are time independent for notational convenience. Note that time can always be added as an additional (spatial) variable.

In Chapter 4, we have proved (strong) existence and uniqueness of solutions in the case where f and σ are (locally) Lipschitz continuous using Banach's fixed point theorem. Our approach was very similar to that in the Picard-Lindelöf theorem for ordinary differential equations.

There is a second Theorem for ordinary differential equations, namely Peano's theorem, which asserts existence (not uniqueness, however) of solutions for ODE with merely continuous coefficients. The proof is based on *compactness*.

In this section, we want to establish similar results for stochastic differential equations, i.e. we want to prove existence of (in general: weak) solutions of (6.1) for coefficients f and σ which are merely continuous.

As we are expecting weak solutions, the main problem to overcome is where to look for solutions. We resolve this problem by not constructing weak solutions directly, but rather to construct *the distribution* of such a solution. Thus, we look for an element of $\mathscr{P}(C([0,T];\mathbb{R}^d))$, the set of all probability measures on $C([0,T];\mathbb{R}^d)$. Our first task is to characterize those measures, which are distributions of solutions.

This is done via the *martingale problem*.

We denote by $C_c^2(\mathbb{R}^d)$ the set of all functions $\varphi : \mathbb{R}^d \to \mathbb{R}$ which are twice continuously differentiable and have compact support, i.e. $\overline{\{x : \varphi(x) \neq 0\}}$ is compact.

Now assume that we have a (weak) solution X of (6.1), defined on some stochastic basis. Thus, for the components $X_1, \ldots, X_d$, we have

$$X_i(t) = \eta_i + \int_0^t f_i(X(s))\, ds + \sum_{j=1}^m \int_0^t \sigma_{ij}(X(s))\, dW_j(s)\,.$$

Writing $\rho_{ij}(s) = \sigma_{ij}(X(s))$, we obtain

$$\begin{aligned} \langle X_k, X_l\rangle &= \Big\langle \textstyle\sum_{\nu=1}^m \rho_{k\nu}\diamond W_\nu, \sum_{\mu=1}^m \rho_{l\mu}\diamond W_\mu\Big\rangle = \displaystyle\sum_{\nu=1}^m \rho_{k\nu}\diamond\Big\langle W_\nu, \textstyle\sum_{\mu=1}^m \rho_{l\mu}\diamond W_\mu\Big\rangle \\ &= \sum_{\nu=1}^m \rho_{k\nu}\diamond\Big(\sum_{\mu=1}^m \rho_{l\mu}\diamond\langle W_\mu, W_\nu\rangle\Big) = \sum_{\nu=1}^m \rho_{k\nu}\diamond(\rho_{l\nu}\langle W_\nu\rangle) = \sum_{\nu=1}^m (\rho_{k\nu}\rho_{l\nu})\diamond\langle W_\nu\rangle \end{aligned}$$

by using the biliniearity of the covariation and the chain rule.

Thus, applying Itô's formula to $\varphi(X)$ for $\varphi \in C_c^2(\mathbb{R}^d)$, we obtain

$$\begin{aligned}
\varphi(X(t)) &= \varphi(\eta) + \sum_{k=1}^{d} \frac{\partial \varphi}{\partial x_k}(X(s)) \diamond X_k + \sum_{k,l=1}^{d} \frac{\partial^2 \varphi}{\partial x_k \partial x_l}(X(s)) \diamond \langle X_k, X_l \rangle \\
&= \varphi(\eta) + \sum_{k=1}^{d} \int_0^t \big[\frac{\partial \varphi}{\partial x_k}(X(s))\big] f_k(X(s))\, ds \qquad (6.2) \\
&\quad + \sum_{k=1}^{d} \sum_{j=1}^{m} \int_0^t \big[\frac{\partial \varphi}{\partial x_k}(X(s))\big] \sigma_{kj}(X(s))\, dW_j(s) \\
&\quad + \frac{1}{2} \sum_{k=1}^{d} \sum_{l=1}^{d} \sum_{\nu=1}^{m} \int_0^t \big[\frac{\partial^2 \varphi}{\partial x_i \partial x_j}(X(s))\big] \sigma_{i\nu}(X(s)) \sigma_{j\nu}(X(s))\, ds
\end{aligned}$$

If we write $\sigma = (\sigma_{ij})$ and $A = \sigma\sigma^T$, then $a_{ij} = \sum_{\nu=1}^{m} \sigma_{i\nu}\sigma_{j\nu}$. We can now introduce the associated operator:

Definition 6.1.1. Given $f : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$, we write $A = \sigma\sigma^T$. The *associated operator* to (6.1) is the differential operator $\mathscr{A}_{f,\sigma} : C_c^2(\mathbb{R}^d) \to B_b(\mathbb{R}^d)$, the bounded, measurable functions on $\mathbb{R}^d$, given by

$$\begin{aligned}
\mathscr{A}_{f,\sigma} u(x) &= \sum_{i=1}^{d} f_i(x) u_{x_i} + \frac{1}{2} \sum_{k,l=1}^{d} a_{kl} u_{x_k x_l} \\
&= \nabla u(x) \cdot f(x) + \frac{1}{2} \mathrm{tr}(A(x) H_u(x))
\end{aligned}$$

where ∇u and H_u refer to the gradient and the Hessian of u, respectively.

We can now rewrite equation (6.2) as

$$\varphi(X(t)) - \varphi(\eta) - \int_0^t \big[\mathscr{A}_{f,\sigma}\varphi\big](X(s))\, ds = \sum_{k=1}^{d} \sum_{j=1}^{m} \int_0^t \frac{\partial \varphi}{\partial x_k}(X(s)) \sigma_{kj}(X(s))\, dW_j(s)\,.$$

Note the integrands in the stochastic integrals are bounded since φ is compactly supported and σ is bounded on bounded subsets of $\mathbb{R}^d$. Consequently, the stochastic integrals are martingales and thus, the right-hand side in the above equation is a martingale.

Hence, we have proved that if X is a solution of (6.1), then for every $\varphi \in C_c^\infty(\mathbb{R}^d)$, the process

$$f(X(t)) - f(X(0)) - \int_0^t \mathscr{A}_{f,\sigma}\varphi(X(s))\, ds$$

is a martingale. We are now very close to the martingale problem. However, we are still working with martingales on our given probability space. We change this now.

Definition 6.1.2. We consider the measurable space $C([0,T];\mathbb{R}^d)$ endowed with its Borel σ-algebra. The *natural filtration* is $\mathbb{F}^{\mathbf{x}} := (\mathscr{F}_t^{\mathbf{x}})_{t\in[0,T]}$, where $\mathscr{F}_t^{\mathbf{x}}(\mathbf{x}(s)\, s \le t)$. Here, we have identified $\mathbf{x}(t)$ with the measurable map $\pi_t : C([0,T];\mathbb{R}^d) \to \mathbb{R}^d$ given by $\pi_t(\mathbf{x}) = \mathbf{x}(t)$. We shall do so in what follows without further notice.

A *solution to the martingale problem associated with* (6.1) is a probability measure $\mathbf{P}$ on $(C([0,T];\mathbb{R}^d), \mathscr{B}(C([0,T],\mathbb{R}^d)))$ such that

(1) $\mathbf{P}(\mathbf{x}(0) = \eta) = 1$.
(2) For every $\varphi \in C_c^2(\mathbb{R}^d)$ the process

$$M_\varphi(t) := \varphi(\mathbf{x}(t)) - \varphi(\mathbf{x}(0)) - \int_0^t \big[\mathscr{A}_{f,\sigma}\varphi\big](\mathbf{x}(s))\, ds$$

is a martingale with respect to $\mathbb{F}^{\mathbf{x}}$ under $\mathbf{P}$.

We will briefly say that $\mathbf{P}$ solves the martingale problem $[f, \sigma, \eta]$. If instead we are given a probability measure μ on $\mathbb{R}^d$ (an *initial distribution*) we say that $\mathbf{P}$ solves the martingale problem $[f, \sigma, \mu]$ if (2) above holds and (1′) $\mathbf{P}(\mathbf{x}(0) \in A) = \mu(A)$ for all $A \in \mathscr{B}(\mathbb{R}^d)$, i.e. under $\mathbf{P}$ the random variable $\mathbf{x}(0)$ has distribution μ.

PROPOSITION 6.1.3. *Let $f : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ be measurable and bounded on bounded subsets of $\mathbb{R}^d$. If $(\Omega, \Sigma, \mathbb{F}, \mathbb{P}, W, X)$ is a weak solution to* (6.1)*, then the distribution $\mathbf{P}$ of X (considered as an $C([0, T]; \mathbb{R}^d)$-valued random variable) is a solution to the martingale problem associated with* (6.1)*.*

PROOF. First, note that the maps $X : (\Omega, \Sigma, \mathbb{P}) \to C([0, T]; \mathbb{R}^d)$ and

$$\mathbf{x} : (C([0, T]; \mathbb{R}^d), \mathscr{B}(C([0, T]; \mathbb{R}^d)), \mathbf{P}) \to C([0, T]; \mathbb{R}^d)$$

have the same distribution. In particular, $\mathbf{P}(\mathbf{x}(0) = \eta) = 1$.

Now let $0 \le r_1 < r_2 < \cdots < r_n \le s < t$, function $h_j \in B_b(\mathbb{R}^d)$, for $j = 1, \ldots, n$ and a function $\varphi \in C_c^2(\mathbb{R}^d)$ be given. We define $\Phi : C([0, T]; \mathbb{R}^d) \to \mathbb{R}$ by

$$\Phi(\mathbf{x}) := \Big[\varphi(\mathbf{x}(t)) - \varphi(\mathbf{x}(s)) - \int_s^t \big[\mathscr{A}_{f,\sigma}\varphi\big](\mathbf{x}(s))\, ds\Big] \cdot \prod_{j=1}^n h_j(\mathbf{x}(r_j))\,.$$

Then Φ is a bounded, measurable map from $C([0, T]; \mathbb{R}^d)$ to $\mathbb{R}$. Thus, $\Phi(X)$ and $\Phi(\mathbf{x})$, as random variables on $(\Omega, \Sigma, \mathbb{P})$ and $(C([0, T]; \mathbb{R}^d), \mathscr{B}(C([0, T]; \mathbb{R}^d)), \mathbf{P})$ respectively, have the same distribution, in particular, they have the same expectation.

Now note that $\prod_{j=1}^n h_j(X(r_j))$ is $\mathscr{F}_s$-measurable, as X is adapted. By the above computations, the process

$$M(t) := \varphi(X(t)) - \varphi(X(0)) - \int_0^t \big[\mathscr{A}_{f,\sigma}\varphi\big](X(s))ds$$

is an $\mathbb{F}$ martingale. Consequently,

$$\mathbb{E}\Phi(X) = \mathbb{E}\big[(M(t) - M(s)) \prod_{j=1}^n h_j(X(r_j))\big] = \mathbb{E}\Big[\mathbb{E}\big[(M(t) - M(s)) \prod_{j=1}^n h_j(X(r_j))|\mathscr{F}_s\big]\Big] = 0$$

Thus also $\Phi(\mathbf{x})$ has expectation 0 under $\mathbf{P}$. Taking $h_j = \mathbb{1}_{A_j}$ for some $A_j \in \mathscr{B}(\mathbb{R}^d)$, we see that

$$\begin{aligned} &\int_{\{\mathbf{x}(r_1)\in A_1,\ldots,\mathbf{x}(r_n)\in A_n\}} \varphi(\mathbf{x}(t)) - \varphi(\mathbf{x}(0)) - \int_0^t \big[\mathscr{A}_{f,\sigma}\varphi\big](\mathbf{x}(r))\, dr\, d\mathbf{P} \\ = &\int_{\{\mathbf{x}(r_1)\in A_1,\ldots,\mathbf{x}(r_n)\in A_n\}} \varphi(\mathbf{x}(s)) - \varphi(\mathbf{x}(0)) - \int_0^s \big[\mathscr{A}_{f,\sigma}\varphi\big](\mathbf{x}(r))\, dr\, d\mathbf{P}\,. \end{aligned}$$

By a monotone class argument, we can replace the cylinder set $\{\mathbf{x}(r_1) \in A_1, \ldots, \mathbf{x}(r_n) \in A_n\}$ with an arbitrary element of $\mathscr{F}_s^{\mathbf{x}}$. This proves that

$$\varphi(\mathbf{x}(t)) - \varphi(\mathbf{x}(0)) - \int_0^t \big[\mathscr{A}_{f,\sigma}\varphi\big](\mathbf{x}(r))\, dr$$

is an $\mathbb{F}^{\mathbf{x}}$-martingale under $\mathbf{P}$. □

Let us pause at this moment and look at the operators associated to some stochastic differential equations.

EXAMPLE 6.1.4. (1) For $f \equiv 0$ and $\sigma = I_{d\times d}$, the solution of (6.1) is an d-dimensional Wiener process. In this case, we have $\sigma\sigma^T = I_{d\times d}$ and we find for the associated operator

$$\mathscr{A}_{0,I} = \frac{1}{2}\sum_{k,l=1}^{d}\delta_{kl}\frac{\partial^2}{\partial x_k \partial x_l} = \frac{1}{2}\sum_{k=1}^{d}\frac{\partial^2}{\partial x_k^2} = \frac{1}{2}\Delta\,.$$

Thus, the operator associated to Browinian motion is half the Laplace operator.

(2) For the Ornstein-Uhlenbeck equation, we have $m = d = 1$ and $f(x) = ax$ and $\sigma(x) = \sigma$. The associated operator is

$$\mathscr{A}_{f,\sigma} = ax\frac{d}{dx} + \frac{1}{2}\sigma^2\frac{d^2}{dx^2}\,.$$

(3) For the equation in Section 4.3.2 with $d = 2$ and $m = 1$ and coefficients

$$f(x,y) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ -x \end{pmatrix} \quad \text{and} \quad \sigma(x,y) \equiv \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

we find

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

Thus, for the associated operator, we have

$$\mathscr{A}_{f,\sigma} = y\frac{\partial}{\partial x} - x\frac{\partial}{\partial y} + \frac{1}{2}\frac{\partial^2}{\partial y^2}$$

(4) For the Cox-Ingersoll-Ross modell with $m = d = 1$ and $f(x) = a(b-x)$ and $\sigma(x) = \sigma\sqrt{x}$, we have

$$\mathscr{A}_{f,\sigma} = a(b-x)\frac{d}{dx} + \frac{1}{2}\sigma^2|x|\frac{d^2}{dx^2}\,.$$

So far, we have proved that if the stochastic differential equation (6.1) has a solution, then the associated martingale problem has a solution. The more interesting question is, whether given a solution of the martingale problem, we can find a solution of the stochastic differential equation (which has the given measure as a distribution).

DEFINITION 6.1.5. Let $(\Omega_j, \Sigma_j, \mathbb{F}_j, \mathbb{P}_j)$ be stochastic Bases for $j = 1,2$ with $\Omega_1 \cap \Omega_2 = \emptyset$. The *canonical extension* of $(\Omega_1, \Sigma_1, \mathbb{F}_1, \mathbb{P}_1)$ by $(\Omega_2, \Sigma_2, \mathbb{F}_j, \mathbb{P}_2)$ is the stochastic basis

$$(\Omega_1 \times \Omega_2, \Sigma_1 \otimes \Sigma_2, (\mathscr{F}_t^1 \otimes \mathscr{F}_t^2)_{t\in[0,T]}, \mathbb{P}_1 \otimes \mathbb{P}_2)\,.$$

Any probability space which is a canonical extension of $(\Omega_1, \Sigma_1, \mathbb{F}_1, \mathbb{P}_1)$ by some other stochastic basis will be called a *canonical extension* of $(\Omega_1, \Sigma_1, \mathbb{F}_1, \mathbb{P}_1)$.

We will use canonical extensions to glue to a given probability space another probability space on which a suitable Brownian motion lives. For example, if $\mathbf{P}$ is a probability measure on $(C([0,T];\mathbb{R}^d), \mathscr{B}(C([0,T];\mathbb{R}^d))$ and $\mathbb{W}$ is m-dimensional Wiener measure on $(C([0,T];\mathbb{R}^m), \mathscr{B}(C([0,T];\mathbb{R}^m))$ and we denote by $\mathbb{F}^{\mathbf{x}}$ and $\mathbb{F}^{\mathbf{w}}$ the natural filtrations, then with $\mathbb{F} = (\mathscr{F}_t^{\mathbf{x}} \otimes \mathscr{F}_t^{\mathbf{w}})_{t\in[0,T]}$ the extension

$$(C([0,T];\mathbb{R}^d) \times C([0,T];\mathbb{R}^m), \mathscr{B}(C([0,T];\mathbb{R}^d) \otimes \mathscr{B}(C([0,T];\mathbb{R}^m), \mathbb{F}, \mathbf{P} \otimes \mathbb{W})$$

is a stochastic extension of the first stochastic basis. Note that on this extension $(\mathbf{x}, \mathbf{w}) \mapsto \mathbf{x}$ has the same law as $\mathbf{x}$ on the old basis is adatped to $\mathbb{F}$ and that $\mathbf{w}$ is an m-dimensional Wiener process even with respect to $\mathbb{F}$ (as Σ_1 and Σ_2 are independent on the enlarged space).

We now prove the following integral representation theorem due to Doob:

THEOREM 6.1.6. *Let $(\Omega, \Sigma, \mathbb{F}, \mathbb{P})$ be a stochastic basis and, on this basis, Let X be a vector in $\mathbb{R}^d$ consisting of continuous local martingales with $X(0) = 0$ and*

$$\langle X_k, X_l \rangle_t = \int_0^t \sum_{j=1}^m \sigma_{kj}(s)\sigma_{lj}(s)ds$$

for a certain $\sigma : \Omega \times [0,T] \to \mathbb{R}^{d\times m}$, every component of which belongs to $L^0(\Omega\times[0,T], \mathscr{P}, \mathbb{P})$, cf. Exercise 1 of Chapter 3.

Then there exists a canonical extension of the basis, on which an m-dimensional Brownian motion W is satisfied such that

$$X_k(t) = \sum_{j=1}^m \int_0^t \sigma_{kj}(s)\, dW_j(s)$$

for all $1 \le k \le d$.

PROOF. For $t \in [0,T]$, we denote by $K(t)$ and $R(t)$ the kernel, resp. the range of $\sigma(t)$ and by $K(t)^\perp$ and $R(t)^\perp$ their orthogonal complements. By $P_{K(t)} \in \mathbb{R}^{m\times m}, P_{R(t)} \in \mathbb{R}^{d\times d}, P_{K(t)^\perp} \in \mathbb{R}^{m\times m}$ and $P_{R(t)^\perp} \in \mathbb{R}^{d\times d}$ we denote the orthogonal projection onto the corresponding space. We note that $\sigma(t)$ is a bijection from $K(t)^\perp$ to $R(t)$, we write $\sigma(t)^{-1}$ for its inverse.

All these matrices are obtained via standard algorithms from linear algebra. Inspecting these algorithms, we see that they are Borel-measurable functions of $\sigma(t)$. In particular, they are again progressive.

We now consider a standard extension of $(\Omega, \Sigma, \mathbb{F}, \mathbb{P})$, on which an m-dimensional Brownian motion $B = (B_1, \dots, B_m)$ independent of Σ is defined. We define:

$$W(t) := \int_0^t \sigma(s)^{-1} P_{R(t)}\, dX(s) + \int_0^t P_{K(t)}\, dB(t)$$

where we have used the matrix-vector notation. Then $W(t)$ is a continuous local martingale. Let us compute the quadratic variation. With computations as at the beginning of this chapter, we see that the matrix of quadratic variations has density with respect to Lebesgue measure as

$$(\sigma^{-1}P_R)\sigma\sigma^T(\sigma^{-1}P_R)^T + P_K P_K^T = P_{K^\perp}P_{K^\perp}^T + P_K = P_{K^\perp} + P_K = I\,.$$

Hence, by Corollary 3.3.3, W is an m-dimensional Wiener process. Using the chain rule from Proposition 3.2.6, we find

$$\sigma \diamond W = (\sigma\sigma^{-1}P_R) \diamond X + \sigma P_K \diamond B = P_R \diamond X + 0 = (P_R + P_{R^\perp}) \diamond X = X\,.$$

Here, we have used that $P_{R^\perp} \diamond X = 0$ almost surely, as its matrix of quadratic variations is $P_{R^\perp}\sigma\sigma^T P_{R^\perp} = 0$ □

We can now prove the main result of this section.

THEOREM 6.1.7. *Let $f : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ be measurable and bounded on bounded subsets. Moreover, let $\eta \in \mathbb{R}^d$. Then there exists a weak solution of equation* (6.1) *with distribution $\mathbf{P}$ if and only if $\mathbf{P}$ solves the associated martingale problem.*

PROOF. It was seen in Proposition 6.1.3 that the distribution of any weak solution of (6.1) solves the associated martingale problem. For the converse, let $\mathbf{P}$ solve the associated martingale problem.

Pick $\varphi_{i,n} \in C_c^2(\mathbb{R}^d)$ such that $\varphi_{i,n} = x_i$ for all $x \in \mathbb{R}^d$ with $\|x\| \le n$. Since $\mathbf{P}$ solves the martingale problem, $M_{\varphi_{i,n}}$ is a martingale for all i, n. Using the stopping time $\tau_n := \inf\{t \in$

$[0,T] : \|\mathbf{x}(t)\| \le n\}$, we see that

$$M^{\tau_n}_{\varphi_{i,n}}(t) = \mathbf{x}_i(t \wedge \tau_n) - \mathbf{x}_i(0) - \int_0^{t\wedge\tau_n} f_i(\mathbf{x}(s))\, ds$$

hence $M_i(t) := \mathbf{x}(t) - \mathbf{x}(0) - \int_0^t f_i(\mathbf{x}(s))\, ds$ is a continuous local martingale.

Let us now compute the covariation processes. Similarly as above, using $\varphi_{k,l,n} \in C_c^2(\mathbb{R}^d)$ with $\varphi_{k,l,n}(x) = x_k x_l$ for $\|x\| \le n$, it follows that

$$M_{k,l}(t) = \mathbf{x}_k(t)\mathbf{x}_l(t) - \mathbf{x}_k(0)\mathbf{x}_l(0) - \int_0^t \mathbf{x}_k(s) f_l(\mathbf{x}(s)) + \mathbf{x}_l(s) f_k(\mathbf{x}(s)) + \frac{1}{2} a_{kl}(\mathbf{x}(s))\, ds$$

is a local martingale. Here, $A = (a_{kl}) = \sigma\sigma^T$ as before.

A direct computation shows

$$\begin{aligned}
& M_k(t)M_l(t) - \int_0^t a_{kl}(\mathbf{x}(s))\, ds \\
= \;& M_{kl}(t) - \mathbf{x}_k(0)M_l(t) - \mathbf{x}_l(0)M_k(t) + \int_0^t (\mathbf{x}_k(s) - \mathbf{x}_k(t)) f_l(\mathbf{x}(s))\, ds \\
& + \int_0^t (\mathbf{x}_l(s) - \mathbf{x}_l(t)) f_k(\mathbf{x}(s))\, ds + \int_0^t f_k(\mathbf{x}(s))\, ds \int_0^t f_l(\mathbf{x}(s))\, ds \\
= \;& M_{kl}(t) - \mathbf{x}_k(0)M_l(t) - \mathbf{x}_l(0)M_k(t) + \int_0^t \big[M_k(s) - M_k(t)\big] f_l(\mathbf{x}(s))\, ds \\
& + \int_0^t \big[M_l(s) - M_l(t)\big] f_k(\mathbf{x}(s))\, ds \\
= \;& M_{kl}(t) - \mathbf{x}_k(0)M_l(t) - \mathbf{x}_l(0)M_k(t) + \int_0^t M_k(s)\, dF_l(s) - M_k(t)F_l(t) \\
& + \int_0^t M_l(s)\, dF_k(s) - M_l(t)F_k(t)
\end{aligned}$$

where we have set $F_i(t) = \int_0^t f_i(\mathbf{x}(s))\, ds$. Noting that F_i is a bounded variation process and integrating by parts (Theorem 3.2.7), it follows that

$$\begin{aligned}
M_k(t)M_l(t) - \int_0^t a_{kl}(\mathbf{x}(s))\, ds =& M_{kl}(t) - \mathbf{x}_k(0)M_l(t) - \mathbf{x}_l(0)M_k(t) - \int_0^t F_k(s)\, dM_l(s) \\
& - \int_0^t F_l(s)\, dM_k(s)\,,
\end{aligned}$$

a local martingale. Thus,

$$\langle M_k, M_l \rangle_t = \int_0^t a_{kl}(\mathbf{x}(s))\, ds = \int_0^t \sum_{j=1}^m \sigma_{kj}(\mathbf{x}(s))\sigma_{lj}(\mathbf{x}(s))\, ds\,.$$

By Theorem 6.1.6, there exists a canonical extension $(\Omega, \Sigma, \mathbb{F}, \mathbb{P})$ of the stochastic basis $(C([0,T];\mathbb{R}^d), \mathscr{B}(C([0,T];\mathbb{R}^d)), \mathbb{F}^{\mathbf{x}}, \mathbf{P})$ on which an m-dimensional Brownian motion W is defined such that

$$M_k(t) = \mathbf{x}_k(t) - \mathbf{x}_k(0) - \int_0^t f_k(\mathbf{x}(s))\, ds = \sum_{j=1}^m \int_0^t \sigma_{kj}(\mathbf{x}(s))\, dW(s)$$

for all $1 \le k \le d$, i.e.

$$\mathbf{x}(t) = \mathbf{x}(0) + \int_0^t f(\mathbf{x}(s))\, ds + \int_0^t \sigma(\mathbf{x}(s))\, dW(s)\,.$$

Since $\mathbb{P}(\mathbf{x}(0) = \eta) = \mathbf{P}(\mathbf{x}(0) = \eta) = 1$, it follows that $(\Omega, \Sigma, \mathbb{F}, \mathbb{P}, W, \mathbf{x})$ is a weak solution of (6.1). $\square$

The theory developed so far has an extension to *time* dependent functions. We write $C_b([0,T] \times \mathbb{R}^d)$ for the bounded, continuous functions on $[0,T] \times \mathbb{R}^d$ and $C^{1,2}((0,T) \times \mathbb{R}^d)$ for the functions on $(0,T) \times \mathbb{R}^d$ which are once continuously differentiable with respect to t and twice continuously differentiable with respect to $x \in \mathbb{R}^d$.

PROPOSITION 6.1.8. *A measure $\mathbf{P}$ solves the martingale problem $[f, \sigma, \eta]$ if and only if $\mathbf{P}(\mathbf{x}(0) = \eta) = 1$ and for all $u \in C_b([0,T] \times \mathbb{R}^d) \cap C^{1,2}((0,T) \times \mathbb{R}^d)$ the process*

$$t \mapsto u(t, \mathbf{x}(t)) - u(0, \mathbf{x}(0)) - \int_0^t \big[\frac{\partial u}{\partial t} - \mathscr{A}_{f,\sigma} u\big](t, \mathbf{x}(s))\, ds$$

is a local martingale under $\mathbf{P}$.

The proof of this proposition is left as an exercise.

6.2. Existence of weak solutions

We can now construct weak solutions of (6.1) via compactness arguments. However, we will not assume that certain *processes* are relatively compact, but we will assume that certain *measures* (namely, solutions of certain martingale problems) are relatively compact, i.e. we assume that they are tight, see Appendix D

Given $f : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ and $\eta \in \mathbb{R}^d$, we will refer to the martingale problem associated with (6.1) as *the martingale problem* $[f, \sigma, \eta]$.

THEOREM 6.2.1. *Let sequences f_n, σ_n and η_n be given which converge to f, σ resp. η. For the functions, we assume that the convergence is uniformly on compact subsets of $\mathbb{R}^d$. Moreover, we assume that the functions are measurable and uniformly bounded on bounded sets.*

Let $\mathbf{P}_n$ be a solution of the martingale problem $[f_n, \sigma_n, \eta_n]$. If the sequence $(\mathbf{P}_n)$ is tight, then every accumulation point of the sequence is a solution of the martingale problem $[f, \sigma, \eta]$.

PROOF. Passing to a subsequence, we may assume that $\mathbf{P}_n$ converges weakly to the probability measure $\mathbf{P}$. Testing against $\psi \circ \pi_0$ for $\psi \in C_b(\mathbb{R}^d)$, we infer that the distribution of $\mathbf{x}(0)$ under $\mathbf{P}_n$ converges weakly to that under $\mathbf{P}$. Thus, since $\eta_n \to \eta$, $\mathbf{P}(\mathbf{x}(0) = \eta) = 1$.

Fix $0 \leq r_1 \leq r_2 \leq \cdots \leq r_n \leq s < t \leq T$, functions $h_j \in C_b(\mathbb{R}^d)$ and a function $\varphi \in C_c^2(\mathbb{R}^d)$. We define $\Phi_n : C([0,T]; \mathbb{R}^d) \to \mathbb{R}$ by

$$\Phi_n(\mathbf{x}) = \Big[\varphi(\mathbf{x}(t)) - \varphi(\mathbf{x}(s)) - \int_s^t \big[\mathscr{A}_{f_n,\sigma_n}\varphi\big](\mathbf{x}(s))\, ds\Big] \prod_{j=1}^n h_j(\mathbf{x}(r_j))\,.$$

We defined Φ similarly, replacing f_n, σ_n with f, σ.

Using our assumptions, is is easy to see that $\mathscr{A}_{f_n,\sigma_n}\varphi$ is uniformly bounded, say by M, and converges to $\mathscr{A}_{f_n,\sigma_n}\varphi$ uniformly on compact subsets of $\mathbb{R}^d$.

It follows that Φ_n is uniformly bounded, namely by $(2\|\varphi\|_\infty + (t-s)M) \prod_{j=1}^d \|h_j\|_\infty$ and converges to Φ, uniformly on compact subsets of $C([0,T]; \mathbb{R}^d)$. Indeed, if $\mathscr{C}$ is a compact subset of $C([0,T]; \mathbb{R}^d)$, then there is a compact subset K of $\mathbb{R}^d$ such that $\mathbf{x}(t) \in K$ for all $\mathbf{x} \in \mathscr{C}$ and all $t \in [0,T]$. Thus, the claimed convergence follows easily from the uniform convergence of $\mathscr{A}_{f_n,\sigma_n}\varphi$ to $\mathscr{A}_{f,\sigma}\varphi$.

Now, let $\varepsilon > 0$ be given. By tightness, we find a compact set $\mathscr{C}$ such that $\mathbf{P}_n(\mathscr{C}^c) \leq \varepsilon$ for all $n \in \mathbb{N}$. Moreover, we find n_0 such that $|\Phi_n(\mathbf{x}) - \Phi(\mathbf{x})| \leq \varepsilon$ and $|\int \Phi\, d\mathbb{P}_n - \int \Phi\, d\mathbb{P}| \leq \varepsilon$ for all $n \geq n_0$. Hence, for $n \geq n_0$, we find

$$\Big|\int \Phi_n\, d\mathbb{P}_n - \int \Phi\, d\mathbb{P}\Big| \quad \leq \quad \int |\Phi_n - \Phi|\, d\mathbb{P}_n + \Big|\int \Phi\, d\mathbb{P}_n - \int \Phi\, d\mathbb{P}\Big|$$

$$\begin{aligned} &\le \int_{\mathscr{C}} \varepsilon \, d\mathbb{P}_n + \int_{\mathscr{C}^c} 2M \, d\mathbb{P}_n + \varepsilon \\ &\le \varepsilon + 2M\varepsilon + \varepsilon = (2+2M)\varepsilon \, . \end{aligned}$$

As $\varepsilon > 0$ was arbitrary, $\int \Phi_n \, d\mathbb{P}_n \to \int \Phi \, d\mathbb{P}$. Since $\mathbb{P}_n$ solves the martingale problem $[f_n, \sigma_n, \eta_n]$, we have $\int \Phi_n \, d\mathbb{P}_n \equiv 0$, cf. the proof of Proposition 6.1.3. Consequently, $\int \Phi \, d\mathbb{P} = 0$. As in the proof of Proposition 6.1.3 (with an additional Dynkin system argument to get from continuous h_j's to indicator functions), we infer that under $\mathbb{P}$, the process

$$\varphi(\mathbf{x}(t)) - \varphi(\mathbf{x}(0)) - \int_0^t \big[\mathscr{A}_{f,\sigma}\big](\mathbf{x}(s)) \, ds$$

is a martingale. As φ was arbitrary, $\mathbf{P}$ solves the martingale problem $[f, \sigma, \eta]$ as claimed. □

Next, we have to find a suitable criterion to prove tightness of measures on $C([0,T];\mathbb{R}^d)$. Recall that the Arzelá Ascoli theorem asserts that a subset of $C([0,T];\mathbb{R}^d)$ is relatively compact if and only if it pointwise bounded and equicontinuous. In particular, if $\|\mathbf{x}(t) - \mathbf{x}(s)\| \le C|t-s|^\alpha$ all $\mathbf{x} \in \mathscr{C}$ and certain constants C, α independent of $\mathbf{x} \in \mathscr{C}$, i.e. the elements of $\mathscr{C}$ are Hölder continuous with constant independent of $\mathbf{x}$, then the set $\mathscr{C}$ is equicontinuous.

Lemma 6.2.2. *Let for $n \in \mathbb{N}$ $(\Omega_n, \Sigma_n, \mathbb{P}_n)$ be probability space and $X_n : \Omega_n \to C([0,T],\mathbb{R}^d)$ be such that for certain constants $M, C, \alpha, \gamma \ge 1$, and $\beta > 0$ we have*

(1) $\mathbb{E}_n\|X_n(0)\|^\gamma \le M$ *for all* $n \in \mathbb{N}$
(2) $\mathbb{E}_n\|X_n(t) - X_n(s)\|^\alpha \le C|t-s|^{1+\beta}$ *for all* $n \in \mathbb{N}$.

Then the distributions $\mathbf{P}_n$ of X_n are tight.

Proof. By the Chebyshev inequality and Corollary A.3, for $\gamma < \frac{\beta}{\alpha}$ we can pick C_1, C_2 such that

$$\mathbb{P}_n(\|X_n\|_\gamma > C_1) \le \frac{\varepsilon}{2} \quad \text{and} \quad \mathbb{P}_n(\|X_n(0)\| > C_2) \le \frac{\varepsilon}{2} \, .$$

Then, for $\mathscr{C} := \{\mathbf{x} : \|\mathbf{x}\|_\gamma \le C_1 \, , \, \|\mathbf{x}(0)\| \le C_2\}$, we find

$$\mathbf{P}_n(\mathscr{C}^c) = \mathbb{P}_n(\|X_n\|_\gamma > C_1 \text{ or } \|X_n(0)\| > C_2) \le \varepsilon \, .$$

By the Arzelá Ascoli theorem, $\mathscr{C}$ is compact in $C([0,T];\mathbb{R}^d)$. Hence, we have proved that the measures $\mathbb{P}_n$ are tight. □

This criterion can be applied in the following situation.

Lemma 6.2.3. *Let $f : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ be measurable and satisfy the growth condition*

$$\|f(x)\| \le a + b\|x\| \quad \textit{and} \quad |\sigma(x)\| \le a + b\|x\| \, .$$

If X is a weak solution of of equation (6.1)*, then*

$$\mathbb{E}\|X\|_{C([0,T])}^{2m} \le C(1 + \|\eta\|^{2m})$$

and

$$\mathbb{E}\|X(t) - X(s)\|^{2m} \le C(1 + \|\eta\|^{2m})|t-s|^m \quad (t, s \in [0,T])$$

for all $m \ge 1$ and a constant C which only depends on m, a and b.

Proof. Since X is a weak solution, we have that

$$\|X(t)\|^{2m} \le C\Big(\|\eta\|^{2m} + \Big\|\int_0^t f(X(s)\, ds\Big\|^{2m} + \Big\|\int_0^t \sigma(X(s))\, dW(s)\Big\|^{2m}\Big)$$

almost surely where C is a constant only depending on m. In what follows, C denotes a generic constant and may change from ocurrence to ocurrence.

By Jensen's inequality,

$$\begin{aligned}\Big\|\int_0^t f(X(s))\,ds\Big\|^{2m} &= \Big[\sum_{i=1}^d\Big(\int_0^t f_i(X(s))\,ds\Big)^2\Big]^m \leq t^m\Big[\int_0^t \|f(X(s))\|^2\,ds\Big]^m\\ &\leq t^{2m-1}\int_0^t \|f(X(s))\|^{2m}\,ds\,.\end{aligned}$$

As for the stochastic integral, we have

$$\mathbb{E}\sup_{0\leq r\leq t}\Big\|\int_0^r \sigma(X(s))\,dW(s)\Big\|^{2m} \leq C\mathbb{E}\Big(\int_0^t |\sigma(X(s))|^2\,ds\Big)^m \leq Ct^{m-1}\int_0^t |\sigma(X(s))|^{2m}\,ds$$

by the Burkholder-Davies-Gundy inequality. It follows that

$$\mathbb{E}\Big(\sup_{0\leq r\leq t}\|X(r)\|^{2m}\Big) \leq C\Big(\|\eta\|^{2m} + \int_0^t \|f(X(s))\|^{2m} + \|\sigma(X(s))\|^{2m}\,ds\Big)\,.$$

Setting $\varphi(t) := \mathbb{E}\Big(\sup_{0\leq r\leq t}\|X(r)\|^{2m}\Big)$ and using the linear growth assumptions, we see that

$$\varphi(s) \leq C\Big(1 + \|\eta\| + \int_0^t \varphi(s)\,ds\Big)\,.$$

Now the first assertion follows from the Gronwall lemma.

Similarly, we can prove that

$$\mathbb{E}\|X(t)-X(s)\|^{2m} \leq C(t-s)^{m-1}\int_0^t 1 + \sup_{0\leq r\leq s}\|X(r)\|^{2m}\,ds \leq C(1+\|\eta\|^{2m})(t-s)^m$$

by the first estimate. □

We can now prove existence of weak solutions for arbitrary continuous coefficients f and σ of linear growth.

Theorem 6.2.4. *Let $f:\mathbb{R}^d\to\mathbb{R}^d$ and $\sigma:\mathbb{R}^d\to\mathbb{R}^{d\times m}$ be continuous and of linear growth. Then for every $\eta\in\mathbb{R}^d$, there exists a weak solution of equation* (6.1).

Proof. By Theorem 6.1.7, it suffices to construct a solution to the martingale problem $[f,\sigma,\eta]$. If f and σ are additionally Lipschitz continuous, then a solution exists by Theorem 4.1.1 and Proposition 6.1.3.

Let

$$\mathscr{H} := \{(f,\sigma) : \text{problem } [f,\sigma,\eta] \text{ has a solution}\}\,.$$

Then $\mathscr{H}$ contains $\mathsf{Lip}(\mathbb{R}^d,\mathbb{R}^d)\times\mathsf{Lip}(\mathbb{R}^d;\mathbb{R}^{d\times m})$ by the above. Moreover, if (f_n,σ_n) is a sequence in $\mathscr{H}$ which is uniformly of linear growth and converges to (f,σ) uniformly on compact subsets of $\mathbb{R}^d$, then $(f,\sigma)\in\mathscr{H}$. Indeed, if $\mathbf{P}_n$ solves the problem $[f_n,\sigma_n,\eta]$, then the sequence $\mathbf{P}_n$ is tight as a consequence of Lemmas 6.2.2 and 6.2.3. Thus, this sequence has an accumulation point which, by Theorem 6.2.1 solves the problem $[f,\sigma,\eta]$.

To finish the proof, it suffices to note that for every continuous f and σ of linear growth, there exist sequences f_n and σ_n of Lipschitz continuous functions which are uniformly of linear growth and converge uniformly on compact subsets to f resp. σ. Thus any such pair of functions belongs to $\mathscr{H}$. □

We can now establish existence of solutions for the Cox-Ingersoll-Ross modell.

Corollary 6.2.5. *Let $a,b,\sigma\geq 0$. Then for every $\eta\geq 0$ there exists a pathwise unique strong solution of the stochastic differential equation*

$$\begin{cases} dX(t) &= a(b-X(t))dt + \sqrt{|X(t)|}dW(t)\\ X(0) &= \eta\,. \end{cases}$$

Proof. Existence of weak solutions follows immediately from Theorem 6.2.4 as the coefficients are continuous and of linear growth. Pathwise uniqueness was proved in Theorem 5.3.1 and that pathwise uniqueness implies strong existence was discussed after Theorem 5.2.5. $\square$

6.3. Uniqueness of solutions

Having established existence of solutions for continuous coefficients f and σ with the help of the martingale problem, we now turn to the question of uniqueness (in law) of solutions. As it turnes out, this question is intimately connected with existence of solutions to some partial differential equation involving the associated operator $\mathscr{A}_{f,\sigma}$.

In this section, we will also consider martigale problems where we prescribe an initial distribution μ rather than an initial value η. This is due to the fact that even if we just want to prove uniqueness for deterministic initial values, we have to consider solutions with random initial distributions on the way.

We start with the following

Lemma 6.3.1. *Let f and σ be continuous. Suppose that for every $\varphi \in C_c^\infty(\mathbb{R}^d)$ there exists a solution $u_\varphi \in C_b([0,T]\times\mathbb{R}^d)\cap C^{1,2}((0,T)\times\mathbb{R}^d))$ of the Cauchy problem*

$$\begin{cases} \frac{du}{dt} &= \mathscr{A}_{f,\sigma}u \\ u(0) &= \varphi\,. \end{cases}$$

Then, if $\mathbf{P}$ and $\mathbf{Q}$ are two solutions of the martingale problem $[f,\sigma,\mu]$, then the one-dimensional marginals agree, i.e. for all $t\in[0,T]$ we have

$$\mathbf{P}(\mathbf{x}(t)\in A) = \mathbf{Q}(\mathbf{x}(t)\in A))$$

for all $A\in\mathscr{B}(\mathbb{R}^d)$.

Proof. Fix $\varphi\in C_c^\infty(\mathbb{R}^d)$, $t\in[0,T]$ and put $g(s,x) := u_\varphi(t-s,x)$ for $0\le s\le t$ and $x\in\mathbb{R}^d$. Then $g\in C([0,t]\times\mathbb{R}^d)\cap C^{1,2}((0,t)\times\mathbb{R}^d)$ satisfies

$$\frac{du}{dt} - \mathscr{A}_{f,\sigma}u = 0 \quad\text{and}\quad g(t) = \varphi\,.$$

By Proposition 6.1.8, the process $g(s,\mathbf{x}(s))$ is a local martingale under both $\mathbf{P}$ and $\mathbf{Q}$. In fact, as $g(s,\mathbf{x}(s))$ is bounded, it is actually a martingale. Consequently,

$$\mathbb{E}_{\mathbf{P}}(\varphi(\mathbf{x}(t))) = \mathbb{E}_{\mathbf{P}}(g(t,\mathbf{x}(t))) = \mathbb{E}_{\mathbf{P}}(g(0,\mathbf{x}(0))) = \int_{\mathbb{R}^d} g(0,x)\,d\mu(x) = \mathbb{E}_{\mathbf{Q}}(\varphi(\mathbf{x}(t)))\,.$$

As $\varphi\in C_c^\infty(\mathbb{R}^d)$ was arbitrary, it follows from a monotone class argument that the push-forward of $\mathbf{P}$ and $\mathbf{Q}$ under $\mathbf{x}(t)$ are equal. That is the assertion. $\square$

Lemma 6.3.1 asserts that any two solutions of the martingale problem (with identical initial distribution) have the same one-dimensional marginals. Note that since the Borel σ-algebra on $C([0,T];\mathbb{R}^d)$ is generated by the point evaluations, see Proposition B.2, a measure on $C([0,T];\mathbb{R}^d)$ is uniquely determined by its finite-dimensional marginals.

We now extend Lemma 6.3.1 to a uniqueness result, by showing that certain conditional distributions also solve the martingale problem.

Theorem 6.3.2. *Let f and σ be continuous. Suppose that for every $f\in C_c^\infty(\mathbb{R}^d)$ there exists a solution $u_f\in C_b([0,T]\times\mathbb{R}^d)\cap C^{1,2}((0,T)\times\mathbb{R}^d))$ of the Cauchy problem*

$$\begin{cases} \frac{du}{dt} &= \mathscr{A}_{f,\sigma}u \\ u(0) &= f\,. \end{cases}$$

Then, if $\mathbf{P}$ and $\mathbf{Q}$ are two solutions of the martingale problem $[f,\sigma,\mu]$, we have $\mathbf{P}=\mathbf{Q}$.

PROOF. As a consequence of Proposition B.2, it suffices to prove that for every $0 \le t_1 < t_2 < \cdots < t_n \le T$ and all $A_1, \ldots, A_n \in \mathscr{B}(\mathbb{R}^d)$, we have that

$$\mathbf{P}(\mathbf{x}(t_1) \in A_1, \ldots, \mathbf{x}(t_n) \in A_n) = \mathbf{Q}(\mathbf{x}(t_1) \in A_1, \ldots, \mathbf{x}(t_n) \in A_n)\,.$$

By a Dynkin system argument, it suffices to prove that for positive measurable functions $f_1, \ldots, f_n$ with $f_j \ge \varepsilon$ for a certain $\varepsilon > 0$, we have

$$\mathbb{E}_{\mathbf{P}}\Big[\prod_{j=1}^n f_k(\mathbf{x}(t_k))\Big] = \mathbb{E}_{\mathbf{Q}}\Big[\prod_{j=1}^n f_k(\mathbf{x}(t_k))\Big] \tag{6.3}$$

We prove this by induction on n. The case $n = 1$ follows from Lemma 6.3.1. Now assume that (6.3) holds for all $n \le m$. We prove that it also holds for $n = m + 1$. We thus assume that $t_m < T$. We set $E = C([t_m, T]; \mathbb{R}^d)$ and define the measures $\mathbf{P}_1, \mathbf{Q}_1$ on E by setting, for $A \in \mathscr{B}(E)$,

$$\mathbf{P}_1(A) := \frac{\mathbb{E}_{\mathbf{P}}\Big[\mathbb{1}_A \prod_{j=1}^m f_k(\mathbf{x}(t_k))\Big]}{\mathbb{E}_{\mathbf{P}}\Big[\prod_{j=1}^m f_k(\mathbf{x}(t_k))\Big]} \quad \text{and} \quad \mathbf{Q}_1(A) := \frac{\mathbb{E}_{\mathbf{Q}}\Big[\mathbb{1}_A \prod_{j=1}^m f_k(\mathbf{x}(t_k))\Big]}{\mathbb{E}_{\mathbf{Q}}\Big[\prod_{j=1}^m f_k(\mathbf{x}(t_k))\Big]}\,.$$

Then $\mathbf{P}_1$ and $\mathbf{Q}_1$ are probablity measures on E. Moreover, by using (6.3) with $n = m$, it follows that

$$\begin{aligned}
\mathbb{E}_{\mathbf{P}_1}(f(\mathbf{x}(t_m))) &= \frac{\mathbb{E}_{\mathbf{P}}\Big[f(\mathbf{x}(t_m)) \prod_{j=1}^m f_k(\mathbf{x}(t_k))\Big]}{\mathbb{E}_{\mathbf{P}}\Big[\prod_{j=1}^m f_k(\mathbf{x}(t_k))\Big]} \\
&= \frac{\mathbb{E}_{\mathbf{Q}}\Big[f(\mathbf{x}(t_m)) \prod_{j=1}^m f_k(\mathbf{x}(t_k))\Big]}{\mathbb{E}_{\mathbf{Q}}\Big[\prod_{j=1}^m f_k(\mathbf{x}(t_k))\Big]} = \mathbb{E}_{\mathbf{Q}_1} f(\mathbf{x}(t_m))\,,
\end{aligned}$$

for all strictly positive, measurable f. It follows that $\mathbf{P}_1$ and $\mathbf{Q}_1$ have the same initial distributions, say ν. Moreover, $\mathbf{P}_1$ and $\mathbf{Q}_1$ solve the martingale problem for $[f, \sigma, \nu]$. Indeed, if $t_m \le r_1 < \ldots, r_l \le s < t$ and $h_1, \ldots, h_l$ are bounded, measurable functions, and $\varphi \in C_c^2(\mathbb{R}^d)$ we may put

$$\Phi(\mathbf{x}) := \Big[\varphi(\mathbf{x}(t)) - \varphi(\mathbf{x}(s)) - \int_s^t [\mathscr{A}_{f,\sigma}]\varphi(\mathbf{x}(r))\, dr\Big] \prod_{i=1}^l h_i(\mathbf{x}(r_i))\,.$$

Then $\mathbb{E}_{\mathbf{P}_1}\Phi(\mathbf{x})$ is a multiple of

$$\mathbb{E}_{\mathbf{P}}\Big(\Big[\varphi(\mathbf{x}(t)) - \varphi(\mathbf{x}(s)) - \int_s^t [\mathscr{A}_{f,\sigma}]\varphi(\mathbf{x}(r))\, dr\Big] \prod_{i=1}^l h_i(\mathbf{x}(r_i)) \prod_{j=1}^m f_j(\mathbf{x}(t_j))\Big) = 0$$

as $\mathbf{P}$ solves the martingale problem. Now a standard argument shows that $\mathbf{P}_1$ solves the martingale problem, cf. the proof of Proposition 6.1.3. Similarly one sees that $\mathbf{Q}_1$ solves the martingale problem.

Now Lemma 6.3.1 yields that $\mathbf{P}_1$ and $\mathbf{Q}_1$ have the same one-dimensional marginals. In particular,

$$\frac{\mathbb{E}_{\mathbf{P}}\Big[f_{m+1}(\mathbf{x}(t_{m+1}) \prod_{j=1}^m f_k(\mathbf{x}(t_k))\Big]}{\mathbb{E}_{\mathbf{P}}\Big[\prod_{j=1}^m f_k(\mathbf{x}(t_k))\Big]} = \frac{\mathbb{E}_{\mathbf{Q}}\Big[f_{m+1}(\mathbf{x}(t_{m+1}) \prod_{j=1}^m f_k(\mathbf{x}(t_k))\Big]}{\mathbb{E}_{\mathbf{Q}}\Big[\prod_{j=1}^m f_k(\mathbf{x}(t_k))\Big]}$$

for all bounded, measurable functions f_{m+1}. As the nominators above are equal by induction hypothesis, we have proved (6.3) for $n = m + 1$. □

We have thus reduced the question of uniqueness of solutions to the martingale problem, i.e. of uniqueness in law for our stochastic differential equation to a queasion of existence to some partial differential equation. If the matrix $A(x) = \sigma(x)\sigma(x)^T$ is positive definite, i.e. there exists a $\kappa > 0$ such that

$$\sum_{ij=1}^{d} a_{ij}(x)\xi_i\xi_j \geq \kappa|\xi|^2$$

and the coefficients A and f are bounded and Hölder continuous, then the Cauchy problem $u_t = \mathscr{A}_{f,\sigma}$ has a solution for every initial value $f \in C_c^\infty$, see Theorem 12 in [**3**, Section 1.7].

6.4. Exercises

(1) Prove Proposition 6.1.8.

(2) Let us consider the stochastic differential equation with coefficients f and σ, which we assume to be locally Lipschitz continuous and of linear growth. The results of Chapter 4 yield that for every $\eta \in L^0(\Omega, \mathscr{F}_0, \mathbb{P})$ there exists a (pathwise) unique, strong solution of the equation with initial datum η. It now follows from Theorem 5.2.5 and the results of this chapter, that for every probability measure μ, there exists a unique solution $\mathbf{P}_\mu$ of the associated martingale problem with initial distribution μ.

Given $f \in B_b(\mathbb{R}^d)$, the bounded, borel-measurable functions on $\mathbb{R}^d$, we define for $t \geq 0$ the function $T(t)f$ by

$$\big[T(t)f\big](x) = \mathbb{E}_x f(\mathbf{x}(t)) = \mathbb{E} f(X(t,x))$$

where $X(t,x)$ denotes a strong solution with initial datum x and $\mathbb{E}$ denotes expectation with respect to the probability measure on that space and $\mathbb{E}_x$ denotes expectation on $C([0,T];\mathbb{R}^d)$ with respect to the probability measure $\mathbf{P}_{\delta_x}$.

Show the following properties:

(a) If $f \in C_b(\mathbb{R}^d)$, then $T(t)f \in C_b(\mathbb{R}^d)$. If $f \in B_b(\mathbb{R}^d)$, then $T(t)f \in B_b(\mathbb{R}^d)$. Moreover, show that if f_n is a bounded sequence converging pointwise to f, then $T(t)f_n \to T(t)f$ pointwise.

Hint: For the first part, use Theorem 6.2.1. Then show the last assertion and use this and a Dynkin class argument to prove the second assertion

(b) Show that for $s, t \geq 0$ we have $T(t+s)f = T(t)T(s)f$.

Hint: Inspect the proof of Theorem 6.3.2.

Part (b) asserts that $T(t)$ is a *semigroup* on either $B_b(\mathbb{R}^d)$ or $C_b(\mathbb{R}^d)$. It is called the *transition semigroup* of the stochastic differential equation.

APPENDIX A

Continuity of Paths for Brownian Motion

Definition A.0.1. Let $(X(t))_{t\in I}$ and $(Y(t))_{t\in I}$ be stochastic processes defined on the same probability space $(\Omega, \Sigma, \mathbb{P})$. Then $(X(t))_{t\in I}$ and $(Y(t))_{t\in I}$ are called *versions of each other* if $\mathbb{P}(X(t) = Y(t)) = 1$ for all $t \in I$.

Theorem A.1. *(Kolmogorov)*
Let $(X(t))_{t\in[0,T]}$ be a stochastic process such that there exist $C, \alpha, \beta > 0$ such that

$$\mathbb{E}|X(t) - X(s)|^\alpha \leq C|t-s|^{1+\beta} \quad \forall\, t, s \in [0,T]\ . \tag{A.1}$$

Then, for all $0 \leq \gamma < \frac{\beta}{\alpha}$, the process X has a version Y whose paths are Hölder continuous of exponent γ, i.e. for all $\omega \in \Omega$ there exists a constant $C(\omega)$ such that

$$|Y(t,\omega) - Y(s,\omega)| \leq C(\omega)|t-s|^\gamma \quad \forall\, t, s \in [0,T]\ .$$

Proof. We may assume that $T = 1$ for notational simplicity.
For $k \in \mathbb{N}$, define the real valued random variables Y_k by

$$Y_k := \sup_{0\leq j\leq 2^k-1} \big|X\big(\frac{j-1}{2^k}\big) - X\big(\frac{j}{2^k}\big)\big|\ .$$

Hence, by assumption,

$$\mathbb{E}Y_k^\alpha \leq \sum_{j=0}^{2^k-1} \mathbb{E}\big|X\big(\frac{j-1}{2^k}\big) - X\big(\frac{j}{2^k}\big)\big| \leq 2^k \cdot C(2^{-k})^{1+\beta} = C2^{-\beta k}\ .$$

Note that for $c < \frac{\beta}{\alpha}$ we have

$$\mathbb{E}\sum_{n=1}^\infty (2^{cn}Y_n)^\alpha = \sum_{n=1}^\infty 2^{c\alpha n}\mathbb{E}Y_n^\alpha \leq C\sum_{n=1}^\infty 2^{c\alpha n}2^{-\beta n} < \infty\ .$$

It follows that almost surely, there exists a constant $C(\omega)$ such that $|Y_n| \leq C(\omega)2^{-cn}$. Indeed, the above proves that the series $\sum_{n=1}^\infty (2^{cn}Y_n)^\alpha$ converges in $L^1(\Omega)$. However, by positivity of the summands, it converges almost surely. Thus, if for some $\omega \in \Omega$ there was no such constant, then $Y_n \geq 2^{-cn}$ infinitely often but then $(2^{cn}Y_n)^\alpha \geq 1$ for infinitely many n. But then for such an ω the series $\sum_{n=1}^\infty (2^{cn}Y_n)^\alpha$ diverges.

Now put $D_k := \{j2^{-k} : j = 0, \dots, 2^k - 1\}$ and $D := \bigcup_{k\in\mathbb{N}} D_k$. Define the real valued random variable Z by

$$Z := \sup\{|t-s|^{-\gamma}|X(t) - X(s)| \,:\, s,t \in D\,, s \neq t\}\ .$$

Then Z is finite almost everywhere.
To see this, note that

$$\begin{aligned} Z &\leq \sup_{k\in\mathbb{N}} \Big\{2^{(k+1)\gamma} \sup_{2^{-(k+1)}<|t-s|\leq 2^{-k}} |X(t) - X(s)|\,,\, s,t \in D\Big\} \\ &\overset{(*)}{\leq} \sup_{k\in\mathbb{N}} 2^{(k+1)\gamma} \cdot 2\sum_{n\geq k} Y_k \leq 2^{\gamma+1}\sum_{n=0}^\infty 2^{\gamma n}Y_n < \infty \text{ a.e.}\ , \end{aligned}$$

since, almost surely, $Y_n \leq 2^{-(\gamma+\varepsilon)n}$, where $\varepsilon > 0$ is such that $\gamma + \varepsilon < \frac{\beta}{\alpha}$.

It remains to verify $(*)$. To that end, let $s,t \in D$ with $|t-s| \leq 2^k$, be given. For $n \geq k$ put $s_n := \sup\{r \in D_n : r \leq s\}$ and $t_n := \sup\{r \in D_n : r \leq t\}$ and note that $s_n = s$ and $t_n = t$ eventually. Hence

$$X(t) = X(t_k) + \sum_{n\geq k} \big(X(t_{n+1}) - X(t_n) \big)$$
$$\text{and} \quad X(s) = X(s_k) + \sum_{n\geq k} \big(X(s_{n+1}) - X(s_n) \big) \ .$$

Furthermore, t_{n+1} is either equal to t_n or to $t_n + 2^{-(n+1)}$. It follows that $|X(t_{n+1}) - X(t_n)| \leq Y_{n+1}$, and, similarly, $|X(s_{n+1}) - X(s_n)| \leq Y_{n+1}$. Thus, by the triangle inequality,

$$|X(t) - X(s)| \leq |X(t_k) - X(s_k)| + 2\sum_{n\geq k} Y_{k+1} \ .$$

Since $|t-s| \leq 2^{-k}$, the difference $|t_k - s_k|$ is either 0 or 2^{-k}, hence $|X(t_k) - X(s_k)| \leq Y_k$. Putting this together, $(*)$ is proved.

We have proved that, almost surely, the map $D \ni t \mapsto X(t)$ is Hölder continuous of exponent γ, more precisely,

$$\sup_{t,s\in D\,,\, t\neq s} |X(t) - X(s)| \leq Z|t-s|^\gamma \ .$$

Define $Y(t) := \lim_{s\to t\, s\in D} X(s)$ on the set $\{Z < \infty\}$ and $Y(t) := 0$ on the remaining null set. Then Y has almost surely Hölder continuous paths of exponent γ. Indeed, for $t_,1, t_2 \in [0,1]$ and $s_1, s_2 \in D$, we have

$$\begin{aligned} |X(t_1) - X(t_2)| &\leq |X(t) - X(s_1)| + Z(s_1 - s_2)^\gamma + |X(t_2) - X(s_2)| \\ &\to Z(t_1 - t_2)^\gamma \end{aligned}$$

as $s_1 \to t_1$ and $s_2 \to t_2$, whenever $Z < \infty$.

Finally, we note that Y is a version of X. Indeed, if (t_n) is a sequence in D converging to $t \in [0,1]$, then $Y(t_n) \to Y(t)$ almost surely. However, by the Kolmogorov condition A.1, $|X(t_n) - X(t)|^\alpha \to 0$ as $n \to \infty$. Passing to a subsequence, $X(t_n) \to X(t)$ almost surely. Thus $Y(t) = X(t)$ almost surely. ☐

Corollary A.2. *Let $(W(t))_{t\geq 0}$ be a Brownian motion and $\alpha \in (0, \frac{1}{2})$. Then $(W(t))_{t\geq 0}$ has a version with α-Hölder continuous paths.*

Proof. For $n \in \mathbb{N}$, we have

$$\mathbb{E}|(W(t) - W(s)|^{2n} = (2n-1)!!|t-s|^n$$

as for a Gaussian random variable γ with mean 0 and variance r, we have $\mathbb{E}\gamma^{2n} = (2n - a)!!r^n$.[1] By Theorem A.1 with $\alpha = 2n$ and $\beta = n-1$, it follows that Brownian motion has a version with $\frac{n-1}{2n}$-Hölder continuous paths. As $n \to \infty$, this coefficient tends to $\frac{1}{2}$. ☐

We also note

Corollary A.3. *In the situation of TheoremA.1, additionally assume that $\alpha \geq 1$. In this case, for $\gamma < \frac{\beta}{\alpha}$, we have*

$$\mathbb{E}\|Y\|_\gamma^\alpha \leq \tilde{C}$$

where $\|Y\|_\gamma := \sup_{t\neq s} |t-s|^{-\gamma}|Y(t) - Y(s)|$ and $\tilde{C}$ is a constant which only depends on C, α, β and γ.

[1] This follows easily by induction.

PROOF. Clearly, $|Y(t) - Y(s)| = |X(t) - X(s)| \leq Z|t - s|^\gamma$ for all $t, s \in D$. By continuity, it follows that $\|Y\|_\gamma \leq Z$. We have seen in the proof of Theorem A.1 that

$$Z \leq 2^{\gamma+1} \sum_{n=0}^{\infty} 2^{\gamma n} Y_n$$

almost surely. Taking norms in $L^\alpha(\Omega)$, we obtain

$$(\mathbb{E}Z^\alpha)^{\frac{1}{\alpha}} \leq 2^{\gamma+1} \sum_{n=0}^{\infty} 2^{\gamma n} (\mathbb{E}Y_n^\alpha)^{\frac{1}{\alpha}} \leq 2^{\gamma+1} \sum_{n=0}^{\infty} 2^{\gamma n} (C 2^{-\beta n})^{\frac{1}{\alpha}} < \infty \,.$$

$\square$

APPENDIX B

Stochastic Processes as Random Elements

Throughout these lecture notes, we are concerned with ($\mathbb{R}^d$-valued) stochastic processes $(X(t))_{t\in[0,T]}$. According to our definition, we are given a probability space $(\Omega, \Sigma, \mathbb{P})$ and, for each $t \in [0,T]$ a random variable $X(t) : \Omega \to \mathbb{R}^d$. Writing $X(t,\omega)$ for $X(t)(\omega)$, we are thus requiring that every $t \in [0,T]$ the map $\omega \mapsto X(t,\omega)$ is measurable.

However, we also keep ω fixed and vary t thus obtaining a function $t \mapsto X(t,\omega)$ which depends on ω. In our applications, the functions $t \mapsto X(t,\omega)$ are usually continuous, i.e. elements of $C([0,T];\mathbb{R}^d)$. In this appendix, we discuss the measurability of maps from Ω to $C([0,T];\mathbb{R}^d)$.

We endow $C([0,T];\mathbb{R}^d)$ with the topology induced by $\|\cdot\|_\infty$, defined by

$$\|f\|_\infty = \max_{j=1,\dots,d} \sup_{t\in[0,T]} |f_j(t)|$$

and denote the Borel σ-algebra by $\mathscr{B}(C([0,T];\mathbb{R}^d))$.

We note that $(C([0,T]), \|\cdot\|_\infty)$ is a complete metric space which is separable by the Stone-Weierstrass theorem. We thus obtain

LEMMA B.1. *$\mathscr{B}(C([0,T];\mathbb{R}^d))$ is generated by either the open balls $B(x,r)$ or the closed balls $\bar{B}(x,r)$ for $x \in C([0,T])$ and $r > 0$. In fact, it suffices in both cases to consider a countable union of such balls.*

PROOF. Let $(f_n)_{n\in\mathbb{N}}$ be a sequence which is dense in $C([0,T];\mathbb{R}^d)$. And define $S := \{B(f_n, k^{-1}) : n, k \in \mathbb{N}\}$. Then every set U which is open in $C([0,T];\mathbb{R}^d))$ is the union of all balls in S which are contained in U. Indeed, if $g \in U$, then $B(g, k^{-1}) \subset U$ for large enough k. Since the f_n's are dense, there is some $m \in \mathbb{N}$ with $f_m \in B(g, \frac{1}{2}k^{-1})$, thus $g \in B(f_m, \frac{1}{2}k^{-1}) \subset B(g, k^{-1}) \subset U$. This shows that every element of U is contained in a ball in S which is entirely contained in U. Hence, U is contained in all balls of S contained in U. The converse is trivial.

This proves that the Borel σ-algebra is generated by all open balls. As every open ball is a countable union of closed balls, it is also generated by the closed balls. ☐

Of particular importance are the point evaluations $\pi_t : C([0,T];\mathbb{R}^d) \to \mathbb{R}^d$, given by $\pi_t(f) = f(t)$. These maps are continuous, hence measurable. Thus, if $X : \Omega \to C([0,T];\mathbb{R}^d)$ is measurable, then $\pi_t(X) : \Omega \to \mathbb{R}^d$ is measurable. As for random processes above, this just means that $\omega \mapsto X(t,\omega)$ is a measurable map from Ω to $\mathbb{R}^d$ for all $t \in [0,T]$.

We will prove next, that the Borel σ-algebra $\mathscr{B}(C([0,T];\mathbb{R}^d))$ is generated by the point evaluations.

PROPOSITION B.2. *$\mathscr{B}(C([0,T];\mathbb{R}^d)) = \sigma(\pi_t : t \in [0,T])$.*

PROOF. A *cylinder set* is a set $A \subset C([0,T];\mathbb{R}^d)$ such that there exist $t_1, \dots, t_n \in [0,T]$ and $A_1, \dots, A_n \in \mathscr{B}(\mathbb{R}^d)$ such that

$$A = \{f \in C([0,T];\mathbb{R}^d) : f(t_j) \in A_j\, j = 1,\dots,n\} = \bigcap_{j=1}^n \pi_{t_j}^{-1}(A_j)\,.$$

The latter shows that cylinder sets belong to the Borel σ-algebra $\mathscr{B}(C([0,T];\mathbb{R}^d)$; they also belong to every σ-algebra with respect to which the point evaluations are measurable. It thus suffices to show that the cylinder sets generate $\mathscr{B}(C([0,T];\mathbb{R}^d))$. Actually, by Lemma B.1, it suffices to show that closed balls $\bar{B}(g,r)$ belong to the σ-algebra generated by the cylinder sets.

To see this, fix g and r and let $\{t_k : k\in\mathbb{N}\} = [0,T]\cap\mathbb{Q}$. For $x\in\mathbb{R}^d$ and $r>0$, we put $B_{r,x} := \prod_{k=1}^d [x_k - r, x_k + r]$ and claim

$$\bar{B}(g,r) = \bigcap_{j\in\mathbb{N}} \pi_{t_j}^{-1}(B_{r,g(t_j)})\,.$$

Note that the right-hand side certainly belongs to the σ-algebra generated by the cylinder sets. Thus, given the claim, we are done.

To prove the claim, note that the set on the right-hand side consists of all $f\in C([0,T];\mathbb{R}^d)$ such that $|f_k(t_j) = g_k(t_j)|\le r$ for all $j\in\mathbb{N}$, i.e. for all rationals in $[0,T]$, and $k=1,\dots,d$. As the components f_k and g_k are continuous for all $k=1,\dots,d$, it follows that the set in fact consists of all $f\in C([0,T];\mathbb{R}^d)$ with $|f_k(t)-g_k(t)|\le r$ for all $k=1,\dots,d$. Hence, the set is $\bar{B}(g,r)$ as claimed. □

Proposition B.2 has important consequences:

COROLLARY B.3. *Let $(\Omega,\Sigma,\mathbb{P})$ be a probability space and $X,Y:\Omega\to C([0,T];\mathbb{R}^d)$.*

(1) *X is measurable if and only if $X(t):\Omega\to\mathbb{R}^d$ is measurable for all $t\in[0,T]$, i.e. if and only if $(\pi_t(X))_{t\in[0,T]}$ is a stochastic process.*
(2) *If both X and Y are measurable, then X and Y have the same distribution if and only if X and Y have the same finite dimensional distribution, i.e. for all $n\in\mathbb{N}$ and $t_1,\dots,t_n$ the random vectors*

$$(X(t_1),\dots,X(t_n)),(Y(t_1),\dots,Y(t_n)):\Omega\to\mathbb{R}^n$$

have the same distribution.

Finally, we define some spaces of function-valued random elements. For a measurable map $X:\Omega\to C([0,T];\mathbb{R}^d)$, we write $[X]$ for the equivalence class modulo equality almost everywhere, i.e.

$$[X] := \{Y:\Omega\to C([0,T];\mathbb{R}^d) \text{ measurable} : X = Y\ \mathbb{P}\text{-a.e.}\}\,.$$

By $L^0(\Omega,\Sigma,\mathbb{P};C([0,T];\mathbb{R}^d))$, we denote the vector space space of all equivalence classes $[X]$ of measurable maps $X\Omega\to C([0,T];\mathbb{R}^d))$. As is customary, we will immediately drop the equivalence classes again from our notation and say things like $X\in L^0(\Omega;C([0,T];\mathbb{R}^d))$ or $X=Y$ with the understanding that equalities are only valid almost everywhere.

Note that since $\|\cdot\|_\infty : C([0,T];\mathbb{R}^d)$ is continuous, also $\|X\|_\infty:\Omega\to\mathbb{R}$ is measurable for measurable $X:\Omega\to C([0,T];\mathbb{R}^d)$. Hence, we may define

$$L^p(\Omega;C([0,T];\mathbb{R}^d)) := \{X\in L^0(\Omega;C([0,T];\mathbb{R}^d)) : \|X\|\in L^p(\Omega;\mathbb{R})\}\,.$$

PROPOSITION B.4. *Endowed with the norm $\|\cdot\|_{L^p(\Omega;C([0,T];\mathbb{R}^d))}$, defined by*

$$\|X\|^p_{L^p(\Omega;C([0,T];\mathbb{R}^d))} := \int_\Omega \|X(\omega)\|_\infty^p\,d\mathbb{P}$$

the space $L^p(\Omega;C([0,T];\mathbb{R}^d))$ is a Banach space.

The proof is basically a copy of the scalar-valued case, but we include it for completeness.

PROOF. We write $|||\cdot|||$ instead of $\|\cdot\|_{L^p(\Omega;C([0,T];\mathbb{R}^d))}$ for convenience. It suffices to prove that every Cauchy sequence in $L^p(\Omega;C([0,T];\mathbb{R}^d))$ has a convergent subsequence. Given a Cauchy sequence X_n, there exists a subsequence X_{n_k} with $|||X_{n_{k+1}} - X_{n_k}|||\le 2^{-k}$.

We put

$$Y_k := X_{n_{k+1}} - X_{n_k} \quad \text{and} \quad f := \sum_{k=1}^{\infty} \|Y_k\|_\infty .$$

For the scalar valued functions $\|Y_k\|_\infty$, we have

$$\Big\| \sum_{k=1}^{N} \|Y_k\|_\infty \Big\|_p \le \sum_{k=1}^{n} |||X_{n_{k+1}} - X_{n_k}||| \le \sum_{k=1}^{N} 2^{-k} \le 1 .$$

By monotone convergence, $\sum_{k=1}^N \|Y_k\|_\infty \uparrow f$ almost surely (say for $\omega \in N^c$ for some $N \in \Sigma$ with $\mathbb{P}(N) = 0$) and $f \in L^p(\Omega)$. By completeness of $C([0,T];\mathbb{R}^d)$, for $\omega \in \Omega \setminus N$ the series $\sum_{k=1}^\infty Y_k(\omega)$ converges to some $Y(\omega)$. Extend Y to Ω by setting it 0 on N. Then Y is measurable and $X_{n_k} = X_{n_1} + \sum_{j=1}^{k-1} Y_j$ converges almost surely to $Y + X_{n_1} =: X$. In particular, $\|X_{n_k} - X\|_\infty \to 0$ almost surely.

Noting that

$$\|X_{n_k} - X\|_\infty \le \sum_{j=1}^{k-1} \|Y_j\|_\infty + \|X_{n_1}\|_\infty + |X\|_\infty \le 2f + \|X_{n_1}\|_\infty$$

where the latter belongs to $L^p(\Omega;\mathbb{R})$, it follows from dominated convergence that

$$|||X_{n_k} - X|||^p = \int_\Omega \|X_{n_k} - X\|_\infty^p \, d\mathbb{P} \to 0$$

as $k \to \infty$. □

We note that

$$d(X,Y) := \int_\Omega \|X - Y\|_\infty \wedge 1 \, d\mathbb{P}$$

defines a metric on $L^0(\Omega; C([0,T];\mathbb{R}^d))$ which turns it into a complete metric space. Convergence with respect to this metric is called *convergence in probability.* As in the scalar case, one shows that a sequence X_n converges to X in probability if and only if every subsequence of X_n has a further subsequence converging to X almost surely.

APPENDIX C

Stieltjes Integrals

Let μ be a measure on $([0,T];\mathscr{B}([0,T]))$. Then

$$F_\mu(t) := \mu([0,t])$$

is an increasing (in the sense of non-decreasing), right-continuous function which determines μ uniquely, as $F_\mu(0) = \mu\{0\}$ and $\mu((a,b]) = F_\mu(b) - F_\mu(a)$ for all $a < b$ and these intervals, together with the singleton $\{0\}$ form a generator of $\mathscr{B}([0,T])$ which is stable under intersections.

It is natural to ask whether every increasing, right-continuous function arises in this way. This is indeed the case.

THEOREM C.1. *Let $F : [0,T] \to \mathbb{R}$ be a positive, increasing, right-continuous function. Then there exists a unique measure μ_F on $([0,T];\mathscr{B}([0,T]))$ such that $\mu_F((a,b]) = F(b) - F(a)$ and $\mu(\{0\}) = F(0)$.*

PROOF. Let $\mathcal{R}$ be the ring generated by intervals of the form $(a,b]$ with $0 \leq a < b \leq T$ and the singleton $\{0\}$. A typical element of $\mathcal{R}$ is of the form

$$\bigcup_{k=1}^n (a_k, b_k] \quad \text{or} \quad \{0\} \cup \bigcup_{k=1}^n (a_k, b_k]$$

for some $n \in \mathbb{N}$ and sets $a_k < b_k$.

We put

$$\mu\Big(\bigcup_{k=1}^n (a_k, b_k]\Big) = \sum_{k=1}^n F(b_k) - F(a_k) \quad \text{and} \quad \mu\Big(\{0\} \cup \bigcup_{k=1}^n (a_k, b_k]\Big) = F(0) + \sum_{k=1}^n F(b_k) - F(a_k)\,.$$

Note that this is well-defined. Indeed, if $(a,b] = \bigcup_{k=1}^n (a_k, b_k]$ as disjoint union, then, after relabeling we may assume that $a = a_1 < b_1 = a_2 < b_2 = \ldots < b_{n-1} = a_n < b_n = b$. We thus have

$$\mu((a,b]) = F(b) - F(a) = \sum_{k=1}^n F(b_k) - F(a_k) = \sum_{k=1}^n \mu((a_k, b_k])\,.$$

Similarly, one also sees that μ is finitely additive.

It suffices to prove that μ is a pre-measure on $\mathcal{R}$. Indeed, in this case, it follows from Carathéodory's theorem, that there exists a unique measure μ_F on $\sigma(\mathcal{R}) = \mathscr{B}([0,T])$ which extends μ.

Thus, let a pairwise disjoint sequence $A_k \in \mathcal{R}$ and an element $A \in \mathcal{R}$ be given with $A = \bigcup_{k\in\mathbb{N}} A_k$. Using the finite additivity, we see that

$$\mu(A) = \sum_{k=1}^n \mu(A_k) + \mu\Big(\bigcup_{k\geq n+1} A_k\Big) \geq \sum_{k=1}^n \mu(A_k) \uparrow \sum_{k=1}^\infty \mu(A_k)\,.$$

This prove that $\mu(A) \geq \sum_{k=1}^\infty \mu(A_k)$.

For the converse inequalities, it suffices to consider the cases where $A = (a,b]$ or $A = \{0\}$ for otherwise, A is a finite union of such sets and intersecting, we obtain partitions of the parts into countably many sets. The case $A = \{0\}$ is trivial.

In the case where $A = (a, b]$ we find, expanding the sets A_k $\bigcup_{k\in\mathbb{N}} A_k = \bigcup_{j\in\mathbb{N}}(a_j, b_j] = (a, b]$. Given $\varepsilon > 0$, we pick δ and δ_j such that $F(a+\delta) - F(a) \leq \varepsilon$ and $F(a_j + \delta_j) - F(a_j) \leq \varepsilon 2^{-j}$ for all $j \in \mathbb{N}$: Then

$$[a + \delta, b] \subset (a, b] \subset \bigcup_{j=1}^{\infty}(a_j, b_j + \delta_j)\,.$$

By compactness, $[a + \delta, b]$ is covered by finitely many of the $(a_j, b_j + \delta_j)$'s, say $[a + \delta, b] \subset \bigcup_{m=1}^{N}(a_{j_m}, b_{j_m} + \delta_{j_m})$. For convenience, we relabel $j_m = m$. Moreover, rearranging and removing unnecessary intervals, we may assume that $a_{m+1} > a_m$ for $m = 1, \ldots, N - 1$.

We find

$$\begin{aligned}
\mu((a,b]) &= F(b) - F(a) \leq F(b) - F(a+\delta) + \varepsilon \\
&\leq F(b_N + \delta_N) - F(a+\delta) + \varepsilon \leq F(b_n) - F(a_1) + 2\varepsilon \\
&= F(b_n) - F(a_N) + \sum_{m=1}^{N-1}\big[F(a_{m+1}) - F(a_m)\big] + 2\varepsilon \\
&\leq F(b_N) - F(a_N) + \sum_{m=1}^{N-1}\big[F(b_m + \delta_m) - F(a_m)\big] + 2\varepsilon \\
&\leq F(b_N) - F(a_N) + \sum_{m=1}^{N-1}\big[F(b_m) - F(a_m) + \varepsilon 2^{-j_m}\big] + 2\varepsilon \\
&\leq \sum_{k=1}^{\infty}\big[F(b_k) - F(a_k)\big] + 3\varepsilon = 3\varepsilon + \sum_{k=1}^{\infty}\mu((a_k, b_k])\,.
\end{aligned}$$

As $\varepsilon > 0$ was arbitrary, the converse inequality is proved. □

Now, let $\varphi : [0, T] \to \mathbb{R}$ be of bounded variation. It is well-known, that in this case, φ can be written as the difference of two increasing functions. If φ is right-continuous, then these functions can be chosen to be right-continuous as well.

Thus if $\varphi = \varphi_1 - \varphi_2$ is such that φ_1, φ_2 are right-continuous, increasing functions, then there exists unique measures μ_{φ_1} and μ_{φ_2} with $\mu_{\varphi_j}((a, b]) = \varphi_j(b) - \varphi_j(a)$ and $\mu_{\varphi_j}(\{0\}) = \varphi_j(0)$.

We now define

$$\mu_\varphi := \mu_{\varphi_1} - \mu_{\varphi_2}\,.$$

Then μ_φ is a signed measure. Note that it does not depend on the specific decomposition of φ into $\varphi_1 - \varphi_2$. Indeed, if $\varphi_1 - \varphi_2 = \psi_1 - \psi_2$ and we put $\tilde{\mu}_\varphi := \mu_{\psi_1} - \mu_{\psi_2}$ then

$$\mu_\varphi((a,b])\mu_{\varphi_1}((a,b]) - \mu_{\varphi_2}((a,b]) = \varphi(b) - \varphi(a) = \mu_{\psi_1}((a,b]) - \mu_{\psi_2}((a,b]) = \tilde{\mu}_\varphi((a,b])\,.$$

Similarly, $\mu_\varphi(\{0\}) = \varphi(0) = \tilde{\mu}_\varphi(\{0\})$. Since μ_φ and $\tilde{\mu}_\varphi$ are signed measures, the set of all A with $\mu_\varphi(A) = \tilde{\mu}_\varphi(A)$ is easily seen to be a Dynkin system. As it contains a generator of $\mathscr{B}([0, T])$ which is stable under intersections, it is already $\mathscr{B}([0, T])$, i.e. $\mu_\varphi = \tilde{\mu}_\varphi$. We may hence define

Definition C.2. Let φ be a function of bounded variation which is right-continuous and μ_φ be defined as above. Then we define for $f \in C([0, T])$

$$\int_A f(t)\, d\varphi(t) := \int_A f\, d\mu_\varphi$$

for all $A \in \mathscr{B}([0, T])$. For $A = [a, b], (a, b], [b, a)$ resp. (a, b), we use the notations

$$\int_a^b f(t)\, d\varphi(t), \quad \int_{a+}^b f(t)\, d\varphi(t), \quad \int_a^{b-} f(t)\, d\varphi(t) \quad \text{resp.} \quad \int_{a+}^{b-} f(t)\, d\varphi(t)\,.$$

Of course, these definitions extend to $f \in L^1([0,T];|\mu_\varphi|)$, in particular to bounded, measurable functions.

We now establish some properties of Stieltjes integrals:

PROPOSITION C.3. *Let $\varphi : [0,T] \to \mathbb{R}$ be right-continuous and of bounded variation and $f \in C([0,T]$.*

(1) *If φ is continuously differentiable, then*

$$\int_A f(t)\,d\varphi(t) = \mathbb{1}_A(0)\varphi(0) + \int_A f(t)\varphi'(t)\,dt$$

(2) *If $\pi_n := (0 = t_1^{(n)} < t_2^{(n)} < \cdots < t_{k_n}^{(n)} = T)$ is a sequence of partitions with $|\pi_n| \to 0$, then*

$$\int_0^T f(t)\,d\varphi(t) = \lim_{n\to\infty} \sum_{j=1}^{k_n} f(t_{j-1}^{(n)})(\varphi(t_j^{(n)}) - \varphi(t_{j-1}^{(n)})\,.$$

PROOF. (1) follows from the uniqueness theorem of measures since μ_φ and $\varphi(0)\delta_0 + \varphi' dt$ are two measure which agree on a generator of the Borel σ-algebra which is stable under intersections.

(2) Put

$$f_n(t) := \sum_{j=1}^{k_n} f(t_j^{(n)}) \mathbb{1}_{[t_{j-1}^{(n)}, t_j^{(n)})}\,.$$

Then f_n converges pointwise to f and is bounded by the constant function $\|f\|_\infty$ which is integrable with respect to $|\mu_\varphi|$. The assertion now follows from the definition of integrals of simple functions and dominated convergence. □

Next, we prove the integration by parts formula.

PROPOSITION C.4. *Let φ, ψ be right-continuous functions of bounded variation. Then for $t \in [0,T]$*

$$\varphi(t)\psi(t) = \varphi(0)\psi(0) + \int_0^t \varphi(s)\,d\psi(s) + \int_0^t \psi(s-)\,d\varphi(s)\,.$$

PROOF. Let μ resp. ν be the singed measures associated with φ resp. ψ.

Decomposition $[0,t]^2 = \{(0,0)\} + L_t + U_t$, where $L_t = \{(x,y)\,:\, 0 \le x < y <\le t\}$ and $U_t := \{(x,y)\,:\, 0 \le y \le x \le t\}$, we find

$$\varphi(t)\psi(t) = \mu\otimes\nu([0,t]^2) = \varphi(0)\psi(0) + \mu\otimes\nu(L_t) + \mu\otimes\nu(U_t)\,.$$

As a consequence of Fubini's theorem

$$\mu\otimes\nu(L_t) = \int_0^t \int_0^{(s-)} d\nu(r)\,d\mu(s) = \int_0^t \psi(s-)\,d\varphi(s)$$

and

$$\mu\otimes\nu(U_t) = \int_0^t \int_0^s d\mu(r)\,d\nu(s) = \int_0^t \varphi(s)\,d\psi(s)\,.$$

Assembling, the result is proved. □

Observe that if μ is a positive measure and $f \ge 0$ be integrable with respect to μ. Then

$$t \mapsto \int_0^t f(s)\,d\mu(s)$$

is a right-continuous, increasing function. Consequently, if φ is of bounded variation and $f \in L^1([0,T],|\mu_\varphi|)$, decomposing f in positive and negative part and writing φ as the difference of two increasing functions, it follows that $\int_0^t f(s)\,d\mu(s)$ is a right-continuous function of bounded variation.

How does one integrate with respect to this bounded variation function?

PROPOSITION C.5. *Let* φ *be a right-continuous function of bounded variation and* $f \in L^1([0,T],|\mu_\varphi|)$. *Put* $\psi(t) := \int_0^t f(s)\,d\varphi(s)$. *Then* $g \in L^1([0,T];|\mu_\psi|)$ *if and only if* $gf \in L^1([0,T],|\mu_\varphi|)$. *In this case,*

$$\int_0^t g(s)\,d\psi(s) = \int_0^t g(s)f(s)\,d\varphi(s)\,.$$

PROOF. This follows immediately from the observation that the measure μ_ψ has density f with respect to μ_φ. □

APPENDIX D

Measures on Topolgical Spaces

In this appendix, we review some properties of probability measures defined on the Borel σ-algebra of a complete, separable metric space. Such a measure is called *Borel probability measure*.

As a first result, we show that Borel probability measures are tight, i.e. $\mu(M) = \sup \mu(K)$ where the supremum runs over all compact subsets of M.

Lemma D.1. *Let μ be a probability measure on $(M, \mathscr{B}(M))$, where M is a complete, separable metric space. By $\mathscr{K}$, we denote the collection of all compact subsets of M. Then $\mu(M) = \sup_{K \in \mathscr{K}} \mu(K)$.*

Proof. It suffices to produce given $\varepsilon > 0$ a compact set K such that $\mu(M \setminus K) \leq \varepsilon$.

Since M is separable, there exists a countable, dense subset $\{x_n : n \in \mathbb{N}\}$. We write $B_{n,k}$ for the closed ball of radius $\frac{1}{k}$ centered at x_n. Then for every $k \in \mathbb{N}$ we have $M = \bigcup_{n \in \mathbb{N}} B_{n,k}$. Consequently, given $\varepsilon > 0$, we find an index m_k such that

$$\mu\Big(M \setminus \bigcup_{n=1}^{M_k} B_{n,k}\Big) \leq \frac{\varepsilon}{2^k}\,.$$

Put $K := \bigcap_{k \in \mathbb{N}} \bigcup_{n=1}^{m_k} B_{n,k}$. Then K is closed and totally bounded, hence, as M is complete, K is compact. Moreover, $\mu(M \setminus K) \leq \sum_{k \in \mathbb{N}} \mu(M \setminus \bigcup_{n=1}^{M_k} B_{n,k}) \leq \varepsilon$. ☐

This gives a first consequence of topological properties of M for the measure-theortic properties of Borel measures. As a matter of fact, we can use Lemma D.1 to prove that every probablity measure μ on a complete, separable metric space is *regular*, i.e. for every set $A \in \mathscr{B}(M)$ we have

$$\mu(A) = \inf\{\mu(U) \,:\, A \subset U\,,\, U \,\text{open}\,\} \quad \text{and} \quad \mu(A) = \sup\{\mu(K) \,:\, K \subset A\,,\, UK \,\text{compact}\,\}\,.$$

We prove this next.

Lemma D.2. *Let μ be a probability measure on $\mathscr{B}(M)$, where M is a complete, separable metric space. Then μ is regular.*

Proof. Let us denote the open sets in M by $\mathscr{O}$ and the compact sets by $\mathscr{C}$. We put

$$\mathscr{G} := \big\{A \in \mathscr{B}(M) \,:\, \mu(A) = \inf_{A \subset U, U \in \mathscr{O}} \mu(U) = \inf_{K \subset A, K \in \mathscr{C}} \mu(K)\big\}\,.$$

Then every closed subset F of M is contained in $\mathscr{G}$. Indeed, given $\varepsilon > 0$ we find by Lemma D.1 a compact set K with $\mu(M \setminus K) \leq \varepsilon$. Then $\tilde{K} := K \cap F$ is a compact set with $\mu(F \setminus \tilde{K}) \leq \varepsilon$. On the other hand, in a metric space every closed set is a G_δ-set, i.e. a countable intersection of open sets. Continuity of the measure implies that $\mu(F)$ is the infimum of the measures of the open sets which contain F.

We claim that $\mathscr{G}$ is a Dynkin system.

Indeed, M, being closed is obviously a member of $\mathscr{G}$. If A belongs to $\mathscr{G}$, then so does A^c. To see this, let $\varepsilon > 0$ be given. Then if $K \subset A$ is a compact set with $\mu(K) \geq \mu(A) - \varepsilon$, then K^c is an open set containing A with $\mu(A^c) > \mu(K^c) - \varepsilon$. Similarly, if U is an open set containing A, with $\mu(A) > \mu(U) - \varepsilon$, then U^c is a closed set contained in A^c with

$\mu(U^c) > \mu(A^c) - \varepsilon$. Intersecting U^c with a suitable compact set, we obtain a compact set contained in A^c with measure greater thatn $\mu(A^c) - 2\varepsilon$.

Finally, let A_k be a sequence of disjoint sets in $\mathscr{G}$. Given $\varepsilon > 0$, pick U_k open with $A_k \subset U_k$ and $\mu(A_k) > \mu(U_k) - \varepsilon 2^{-k}$. Then $\bigcup_k U_k$ is an open set containing $\bigcup_k A_k$ with $\mu(\bigcup_k U_k) \leq \mu(\bigcup_k A_k) + \varepsilon$.

For approximation from within, first pick k_0 with $\mu(\bigcup_{k \geq k_0} A_k) \leq \varepsilon/2$ and then $K_j \subset A_j$ compact with $\mu(A_j \setminus K_j) \leq \varepsilon(2k_0)^{-1}$ for $j = 1, \ldots, k_0$. Then $K := K_1 \cup \cdots \cup K_{k_0}$ is a compat set contained in $\bigcup_k A_k$ with $\mu(\bigcup_k A_k \setminus K) \leq \varepsilon$.

It follows that $\bigcup_k A_k$ is an element of $\mathscr{G}$, hence $\mathscr{G}$ is a Dynkin system as claimed.

By Dynkins π-λ theorem, $\mathscr{G} = \mathscr{B}(M)$, proving that μ is regular.

□

Next, we introduce the concept of weak convergence.

Definition D.3. Let μ_n and μ be probability measures on $(M, \mathscr{B}(M))$. We say that μ_n *converges weakly* to μ if

$$\int_M f\, d\mu_n \to \int_M f\, d\mu \quad \text{for all } f \in C_b(M)\,.$$

Weak convergence is induced by the so-called *weak topology* induced on the finite Borel measures by the bounded continuous functions.[1] This topology has a basis sets of the form

$$U_{f_1,\ldots,f_n;\varepsilon}(\mu) := \left\{\nu : \left|\int f_j\, d\mu - \int f_j\, d\nu\right| < \varepsilon \quad \forall\, j = 1, \ldots, n\right\}$$

where $f_1, \ldots, f_n$ are bounded, continuous functions.

It can be proved that if M is a complete, separable metric space, then the restriction of this topology to the Borel probability measures, is metrizable through a complete, separable metric, i.e. in the weak topology, the Borel probability measures on a complete, separable metric space are a complete separable metric space in their own right.

Important for applications is a characterization of *compact* subsets of measures (endowed with the weak topology). By the above results about the metrizability of the weak topology, a set is compact if and only if it is sequentially compact. We thus state and prove this result only for sequences.

Definition D.4. Let μ_n be a sequence of Borel probability measures. This sequence is called *relatively weakly compact* if each of its subsequences has a weakly convergent subsequence. It is called *tight* if given $\varepsilon > 0$ there exists a compact subset K of M such that $\mu_n(M \setminus K) \leq \varepsilon$ for all $n \in \mathbb{N}$.

Theorem D.5. *(Prokhorov)*

Let μ_n be a sequence of Borel probability measures on a complete, separable metric space M. Then the sequence μ_n is relatively weakly compact if and only if it is tight.

Proof. See [**1**, Theorem 8.6.2]. □

[1] If you know something about Banach spaces and their dual spaces, note that this is *not* the weak topology in the Banach space sense

Bibliography

[1] V. I. Bogachev. *Measure theory. Vol. I, II.* Springer-Verlag, Berlin, 2007.

[2] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. Characterization and convergence.

[3] Avner Friedman. *Partial differential equations of parabolic type.* Prentice-Hall Inc., Englewood Cliffs, N.J., 1964.

[4] Olav Kallenberg. *Foundations of modern probability.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[5] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, second edition, 1991.

[6] Bernt Øksendal. *Stochastic differential equations.* Universitext. Springer-Verlag, Berlin, sixth edition, 2003. An introduction with applications.

[7] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, third edition, 1999.

[8] Daniel W. Stroock and S. R. Srinivasa Varadhan. *Multidimensional diffusion processes.* Classics in Mathematics. Springer-Verlag, Berlin, 2006. Reprint of the 1997 edition.

[9] Shinzo Watanabe and Toshio Yamada. On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.*, 11:155–167, 1971.

[10] Shinzo Watanabe and Toshio Yamada. On the uniqueness of solutions of stochastic differential equations. II. *J. Math. Kyoto Univ.*, 11:553–563, 1971.