BMS Basic Course

Stochastic Processes II/ Wahrscheinlichkeitstheorie III

Michael Scheutzow

Lecture Notes

Technische Universität Berlin

Sommersemester 2013

PRELIMINARY VERSION
SEPTEMBER 4TH 2014

Contents

Chapter 1

Processes, Filtrations, Martingales

1.1 Foundations

Most of this chapter can be found in the first chapter of the monograph [KS91] or in [RY94].

Definition 1.1. Let $(\Omega, \mathcal{F})$ and $(E, \mathcal{E})$ be measurable spaces and I a (non-empty) index set. A collection $\mathbb{X} = (X_t)_{t \in I}$ of measurable maps from $(\Omega, \mathcal{F})$ to $(E, \mathcal{E})$ is called (E-valued) *process*. If – in addition – a probability measure is specified on $(\Omega, \mathcal{F})$, then we call $\mathbb{X}$ a *stochastic process*.

Notation 1.2. If (E, d) is a metric space (or, more generally, E is a topological space), then we denote its Borel-σ-algebra (i.e. the smallest σ-algebra containing all open subsets of E) by $\mathcal{B}(E)$. Unless we say something different, we always assume that subsets of $\mathbb{R}^d$ are equipped with the Euclidean metric. Sometimes we write $\mathcal{B}[a, b]$ instead of $\mathcal{B}([a, b])$ etc.

Notation 1.3. Let $(E, \mathcal{E})$ be a measurable space. $\mathcal{E}^I$ denotes the smallest σ-algebra on E^I such that for each $i \in I$ the projection map $\pi_i : E^I \to E$ defined as $\pi_i(x) := x_i$, $x \in E^I$ is measurable. $\mathcal{E}^I$ is called *product σ-algebra*.

The following proposition is easy to prove.

Proposition 1.4. *Let $(X_t)_{t \in I}$ be an $(E, \mathcal{E})$-valued process. Let $\mathbb{X} : (\Omega, \mathcal{F}) \to (E^I, \mathcal{E}^I)$ be defined by $\mathbb{X}(\omega)(t) := X_t(\omega)$. Then $\mathbb{X}$ is measurable.*

Remark 1.5. If I is uncountable then the σ-algebra $\mathcal{E}^I$ is rather crude. Subsets of E^I which are not determined by countably many $i \in I$ are *never* in $\mathcal{E}^I$. For example $\{f \in \mathbb{R}^{[0,1]} : \sup_{s \in [0,1]} f(s) \leq 1\} \notin \mathcal{B}^{[0,1]}$, no matter which σ-algebra $\mathcal{B}$ on $\mathbb{R}$ we choose.

Definition 1.6. Let $\mathbb{X}$ be an $(E, \mathcal{E})$-valued stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$. Then $\mathbb{P}_{\mathbb{X}} := \mathbb{P}\mathbb{X}^{-1}$ is called the *distibution* or the *law* of $\mathbb{X}$. We often write $\mathcal{L}(\mathbb{X})$ instead of $\mathbb{P}_{\mathbb{X}}$.

Remark 1.7. Note that $\mathbb{P}_{\mathbb{X}}$ is a probability measure on $(E^I, \mathcal{E}^I)$.

Definition 1.8. Let $\mathbb{X}$ and $\mathbb{Y}$ be two stochastic processes on $(\Omega, \mathcal{F}, \mathbb{P})$ with the same index set I and the same target space $(E, \mathcal{E})$.

a) $\mathbb{X}$ and $\mathbb{Y}$ are called *modifications* if for all $t \in I$ we have $\mathbb{P}(\{\omega : X_t(\omega) = Y_t(\omega)\}) = 1$.

b) $\mathbb{X}$ and $\mathbb{Y}$ are called *indistinguishable* (or *equivalent*) if $\mathbb{P}(\{\omega : X_t(\omega) = Y_t(\omega) \text{ for all } t \in I\}) = 1$.

Remark 1.9. a) Strictly speaking, the previous definition should be written more carefully. It is possible that for processes $\mathbb{X}$ and $\mathbb{Y}$ the set $\{\omega : X_t(\omega) = Y_t(\omega)\}$ is not in $\mathcal{F}$ (when the space $(E, \mathcal{E})$ is a bit pathological – we will provide an example in class). Therefore the phrase "we have ..." in part a) should be replaced by "the set $\{\omega : X_t(\omega) = Y_t(\omega)\}$ contains a set $F \in \mathcal{F}$ such that $\mathbb{P}(F) = 1$". A similar remark applies to part b). As an example, take $I = \{0\}$, $(\Omega, \mathcal{F}) = (E, \mathcal{E}) := (\{a, b\}, \{\emptyset, E\})$, $X(a) = a, X(b) = b, Y(a) = Y(b) = a$. Then $\{\omega : X(\omega) = Y(\omega)\} = \{a\}$ which is not in $\mathcal{F}$.

b) If $\mathbb{X}$ and $\mathbb{Y}$ are modifications then $\mathbb{P}_{\mathbb{X}} = \mathbb{P}_{\mathbb{Y}}$.

c) If $\mathbb{X}$ and $\mathbb{Y}$ are indistinguishable then they are modifications.

d) If $\mathbb{X}$ and $\mathbb{Y}$ are modifications and I is countable (i.e. finite or countably infinite) then they are indistinguishable.

Example 1.10. $\Omega = [0,1], \mathcal{F} = \mathcal{B}[0,1], \mathbb{P} = \lambda|_{[0,1]}, I = [0,1]$ (where $\lambda|_{[0,1]}$ is Lebesgue measure on the Borel σ-Algebra $\mathcal{B}[0,1]$ on $[0,1]$).

$$X_t(\omega) := \begin{cases} 1, & t = \omega \\ 0, & t \neq \omega \end{cases} \qquad Y_t(\omega) \equiv 0.$$

Then $\mathbb{X}$ and $\mathbb{Y}$ are modifications but not indistinguishable.

Definition 1.11. If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, then a set $A \in \mathcal{F}$ is called $\mathbb{P}$-*negligible* if $\mathbb{P}(A) = 0$.

1.2 Filtrations and stopping times

In this section we consider stochastic processes and filtrations indexed by the interval $[0, \infty)$. We could formulate these concepts for more general totally or even partially ordered index sets but we prefer not to be too general. All (stochastic) processes are assumed to have index set $I = [0, \infty)$ unless we explicitly say something different. Note that we will usually write X instead of $\mathbb{X}$ (and similarly for other letters denoting (stochastic) processes).

Definition 1.12. An $(E, \mathcal{E})$-valued process X on $(\Omega, \mathcal{F})$ is called *measurable* if, for every $A \in \mathcal{E}$, we have $X^{-1}(A) \in \mathcal{B}[0, \infty) \otimes \mathcal{F}$ (where $X^{-1}(A) := \{(t, \omega) \in [0, \infty) \times \Omega : X_t(\omega) \in A\}$).

Definition 1.13. Let $(\Omega, \mathcal{F})$ be a measurable space. $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ is called a *filtration* (on $(\Omega, \mathcal{F})$) if for each $t \in [0, \infty)$, $\mathcal{F}_t$ is a sub-$\sigma-$algebra of $\mathcal{F}$ such that $\mathcal{F}_s \subseteq \mathcal{F}_t$ whenever $0 \leq s \leq t < \infty$. In this case, $(\Omega, \mathcal{F}, \mathbb{F})$ is called a *filtered measurable space* (FMS). If, moreover, $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, then $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ is called a *filtered probability space* (FPS). We define $\mathcal{F}_\infty$ as the smallest σ-algebra which contains all $\mathcal{F}_t$.

Definition 1.14. An $(E, \mathcal{E})$-valued process X on the FMS $(\Omega, \mathcal{F}, \mathbb{F})$ is called *adapted* (or $\mathbb{F}$-*adapted*) if for each $t \geq 0$, X_t is $(\mathcal{F}_t, \mathcal{E})$-measurable. The filtration $\mathcal{F}_t := \sigma(X_s, s \leq t)$ is called the *natural filtration* of $\mathbb{X}$ or the filtration *generated by* $\mathbb{X}$.

Definition 1.15. If $(\Omega, \mathcal{F}, \mathbb{F})$ is a FMS, then we define

$$\mathcal{F}_s^+ := \mathcal{F}_{s^+} := \bigcap_{t > s} \mathcal{F}_t, \; s \geq 0 \text{ and } \mathbb{F}^+ := (\mathcal{F}_{s^+})_{s \geq 0}.$$

It is easy to check that $\mathbb{F}^+$ is also a filtration and that $(\mathbb{F}^+)^+ = \mathbb{F}^+$.

Definition 1.16. If $(\Omega, \mathcal{F}, \mathbb{F})$ is a FMS such that $\mathbb{F} = \mathbb{F}^+$, then $\mathbb{F}$ is called *right continuous*.

The following assumption about a filtration is common in stochastic analysis but not so common in the theory of Markov processes.

Definition 1.17. A filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ is said to satisfy the *usual conditions* (or to be *normal*) if the following hold:

(i) $\mathcal{F}_0$ contains all $\mathbb{P}$-negligible sets in $\mathcal{F}$,

(ii) $\mathbb{F} = \mathbb{F}^+$, i.e. $\mathbb{F}$ is right continuous.

Remark 1.18. Some authors require in addition that the space $(\Omega, \mathcal{F}, \mathbb{P})$ is *complete*, i.e. that each subset of a $\mathbb{P}$-negligible set belongs to $\mathcal{F}$. Here we follow [KS91] who do not require completeness.

Definition 1.19. Let $(\Omega, \mathcal{F}, \mathbb{F})$ be a FMS. A map $\tau : \Omega \to [0, \infty]$ is called a *weak* $\mathbb{F}$*-stopping time* or *optional time*, if for all $t \geq 0$ we have $\{\omega \in \Omega : \tau(\omega) \leq t\} \in \mathcal{F}_t^+$. τ is called a *strict* $\mathbb{F}$*-stopping time* or just $\mathbb{F}$*-stopping time* if for all $t \geq 0$ we have $\{\omega \in \Omega : \tau(\omega) \leq t\} \in \mathcal{F}_t$. The specification "$\mathbb{F}$" is often omitted when the choice of the filtration is clear from the context.

Remark 1.20. A strict $\mathbb{F}$-stopping time is a weak $\mathbb{F}$-stopping time and a weak $\mathbb{F}$-stopping time is the same as an $\mathbb{F}^+$-stopping time. If the filtration $\mathbb{F}$ is right continuous, then a weak $\mathbb{F}$-stopping time is automatically an $\mathbb{F}$-stopping time.

Proposition 1.21. *Let $(\Omega, \mathcal{F}, \mathbb{F})$ be a FMS and $\tau : \Omega \to [0, \infty]$ a map. The following are equivalent:*

(i) *τ is a weak stopping time*

(ii) *$\{\omega : \tau(\omega) < t\} \in \mathcal{F}_t$ for all $t \geq 0$.*

Proof. (i) $\Rightarrow$ (ii): Let τ be a weak stopping time and $t \geq 0$. Then

$$\{\tau < t\} = \bigcup_{n \in \mathbb{N}} \{\tau \leq t - \frac{1}{n}\} \in \mathcal{F}_t.$$

(ii) $\Rightarrow$ (i): Let τ satisfy (ii) and $t \geq 0$. Then

$$\{\tau \leq t\} = \bigcap_{n \in \mathbb{N}} \{\tau < t + \frac{1}{n}\} \in \mathcal{F}_t^+.$$

□

Example 1.22. Let $\Omega = \{a, b\}$, $\mathcal{F} = 2^\Omega$ equipped with the uniform measure $\mathbb{P}$. Define $X_t(a) := t$ and $X_t(b) = -t$ for $t \geq 0$. Let $\mathbb{F}$ be the filtration generated by X, i.e. $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_t = 2^\Omega$ for $t > 0$. Then $\tau(\omega) := \inf\{t \geq 0 : X_t(\omega) > 0\}$ is clearly an optional time but not a stopping time. Note that $\mathbb{F}$ is not right continuous.

Warning 1.23. For each stopping time τ on a FMS $(\Omega, \mathcal{F}, \mathbb{F})$ we have for all $t \geq 0$

$$\{\tau = t\} = \{\tau \leq t\} \setminus \{\tau < t\} \in \mathcal{F}_t \tag{1.2.1}$$

by the previous proposition, but the converse is not true, i.e. a map $\tau : \Omega \to [0, \infty]$ satisfying (1.2.1) for every $t \geq 0$ is not necessarily a stopping time – not even a weak one. As an example, take $(\Omega, \mathcal{F}, \mathbb{P}) := ([0,1], \mathcal{B}[0,1], \lambda|_{[0,1]})$ and define $X_t(\omega) = 1$ if $t = \omega$ and $X_t(\omega) = 0$ otherwise. Let $\mathbb{F}$ be the filtration generated by X, i.e. $\mathcal{F}_t := \sigma(\{\omega\}, \omega \leq t)$. Then $\tau(\omega) := \inf\{s : X_s(\omega) = 1\}$ is not a stopping time (not even a weak one) since $\{\tau < t\}$ is not in $\mathcal{F}_t$ for any $t \in (0,1)$ but $\{\tau = t\} \in \mathcal{F}_t$.

Definition 1.24. Let $(\Omega, \mathcal{F}, \mathbb{F})$ be a FMS and $\tau : \Omega \to [0, \infty]$ a map.

$$\begin{aligned}
\mathcal{F}_\tau &:= \{A \in \mathcal{F}_\infty : A \cap \{\tau \leq t\} \in \mathcal{F}_t \text{ for all } t \geq 0\}, \\
\mathcal{F}_\tau^+ &:= \{A \in \mathcal{F}_\infty : A \cap \{\tau \leq t\} \in \mathcal{F}_t^+ \text{ for all } t \geq 0\}.
\end{aligned}$$

The proof of the following lemma is very easy and is left to the reader.

Lemma 1.25. *Let $(\Omega, \mathcal{F}, \mathbb{F})$ be a FMS and $\tau : \Omega \to [0, \infty]$ an $\mathbb{F}$-stopping time.*

a) $\mathcal{F}_\tau$ is a σ-algebra.

b) If $\tau \equiv s \in [0, \infty]$, then $\mathcal{F}_\tau = \mathcal{F}_s$.

c) τ is $\mathcal{F}_\tau$-measurable.

d) If $0 \leq \sigma \leq \tau$, then $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$ (even if σ is not a stopping time).

e) If $(\sigma_n)_{n \in \mathbb{N}}$ is a sequence of stopping times, then so is $\bar{\sigma} := \sup_n \sigma_n$ and $\underline{\sigma} := \inf_n \sigma_n$ is a weak stopping time.

f) If σ is another stopping time then $\mathcal{F}_{\sigma \wedge \tau} = \mathcal{F}_\sigma \cap \mathcal{F}_\tau$.

Remark 1.26. If τ is a weak stopping time then $\mathcal{F}_\tau$ need not be a σ-algebra since it does not necessarily contain the set Ω. In fact $\mathcal{F}_\tau$ is a σ-algebra if and only if τ is a strict stopping time.

Next, we want to investigate under which conditions first entrance times of a process X are stopping times.

Definition 1.27. Let X be an $(E, \mathcal{E})$-valued process on a FMS $(\Omega, \mathcal{F}, \mathbb{F})$ and $G \subseteq E$. Then the maps S^G resp. T^G from Ω to $[0, \infty]$ defined as

$$\begin{aligned}
S^G(\omega) &: = \inf\{t > 0 : X_t(\omega) \in G\} \\
T^G(\omega) &: = \inf\{t \geq 0 : X_t(\omega) \in G\}
\end{aligned}$$

are called *hitting time* resp. *debut time* of X in G.

Proposition 1.28. *Let (E, d) be a metric space with Borel σ-algebra $\mathcal{E}$ and let X be an E-valued adapted process.*

a) *If $G \subseteq E$ is open and X has right (or left) continuous paths, then T^G and S^G are weak stopping times.*

b) *If $G \subseteq E$ is closed and X has continuous paths, then T^G is a strict stopping time.*

Proof. a) Let $t > 0$. Then

$$\{T^G < t\} = \bigcup_{0 \le s < t} \{X_s \in G\} = \bigcup_{s \in [0,t) \cap \mathbb{Q}} \{X_s \in G\} \in \mathcal{F}_t,$$

and analogously for S^G, where we used the facts that G is open and the continuity assumption in the second equality.

b) For $n \in \mathbb{N}$, let $G_n := \{x \in E : d(x, G) < \frac{1}{n}\}$. Then G_n is open for each n and $T^{G_n} < T^G$ on $\{T^G \in (0, \infty)\}$. To see this, observe that the fact that G is closed and X is right continuous implies $X_{T^G} \in G$ on the set $\{T^G < \infty\}$ and left continuity of X then shows that $T^{G_n} < T^G$ on $\{T^G \in (0, \infty)\}$. The sequence T^{G_n} is clearly increasing and therefore converges to some $T(\omega) \le T^G(\omega)$. Using again continuity of X we see that $T(\omega) = T^G(\omega)$ for all $\omega \in \Omega$. From this it follows that $\{T^G \le t\} = \cap_{n=1}^{\infty} \{T^{G_n} < t\} \in \mathcal{F}_t$ by part a) since G_n is open.

□

We will see later in Theorem 1.37 that under slight additional assumptions on the filtration we get a much better result.

Definition 1.29. Let $(\Omega, \mathcal{F}, \mathbb{F})$ be a FMS.

a) The family

$$\{A \subset [0, \infty) \times \Omega : A \cap ([0, t] \times \Omega) \in \mathcal{B}[0, t] \otimes \mathcal{F}_t \text{ for all } t \ge 0\}$$

is called the *progressive* σ-algebra.

b) An $(E, \mathcal{E})$-valued process X is called *progressive* or *progressively measurable* if $X : [0, \infty) \times \Omega \to E$ is measurable with respect to the progressive σ-algebra.

Remark 1.30. Note that the progressive σ-algebra is a σ-algebra. Further note that a progressive set A satisfies $A \in \mathcal{B}[0, \infty) \otimes \mathcal{F}_\infty$ and that a progressive process X is adapted and measurable. On the other hand an adapted and measurable process X is *not* necessarily progressive. We will provide an example at the end of this section (Example 1.38).

Proposition 1.31. *If X is an $(E, \mathcal{E})$-valued progressive process on a FMS $(\Omega, \mathcal{F}, \mathbb{F})$ and T is a stopping time. Then*

a) *X_T is $\mathcal{F}_T$-measurable, i.e. for $B \in \mathcal{E}$ we have $\{\omega : T(\omega) < \infty \text{ and } X_{T(\omega)}(\omega) \in B\} \in \mathcal{F}_T$.*

b) *The process $(X_{T \wedge t})_{t \ge 0}$ is progressive.*

Proof. a) Fix $t \in [0, \infty)$. Then the map $\varphi : \Omega \to [0, t] \times \Omega$ defined as $\varphi(\omega) := (T(\omega) \wedge t, \omega)$ is $(\mathcal{F}_t, \mathcal{B}[0, t] \otimes \mathcal{F}_t)$-measurable since for $0 \le a \le t$ and $A \in \mathcal{F}_t$ we have

$$\varphi^{-1}([a, t] \times A) = A \cap \{\omega : T(\omega) \ge a\} \in \mathcal{F}_t$$

and the sets $[a, t] \times A$ of this form generate $\mathcal{B}[0, t] \otimes \mathcal{F}_t$.

Therefore, for $B \in \mathcal{E}$,

$$\{T < \infty, X_T \in B\} \cap \{T \leq t\} = \{X_{T \wedge t} \in B\} \cap \{T \leq t\} = \{X \circ \varphi \in B\} \cap \{T \leq t\} \in \mathcal{F}_t$$

and

$$\{T < \infty, X_T \in B\} = \cup_{n=1}^{\infty} \{T \leq n, X_T \in B\} \in \mathcal{F}_{\infty}.$$

b) The proof of this part is similar.

□

In the proof of the following proposition we will need the following lemma.

Lemma 1.32. *If (E,d) is a metric space with Borel-σ-algebra $\mathcal{E}$ and $X_n : (\tilde{\Omega}, \tilde{\mathcal{F}}) \to (E, \mathcal{E})$, $n \in \mathbb{N}$ are measurable and $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ for all $\omega \in \tilde{\Omega}$, then X is also measurable.*

Proof. Let $A \subseteq E$ be closed and $A_n := \{x \in E : d(x, A) < \frac{1}{n}\}$. Then A_n is open and

$$X^{-1}(A) = \bigcap_{n=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcap_{m=k}^{\infty} X_m^{-1}(A_n) \in \tilde{\mathcal{F}}.$$

□

Remark 1.33. The previous lemma does not hold on arbitrary topological spaces E – even if E is compact Hausdorff! For example, if $I = [0,1]$ carries the usual topology and $E := I^I$ is equipped with the product topology (which is compact and Hausdorff), then there exists a sequence of continuous (and hence measurable) $X_n : I \to E$ which converge pointwise to a map X which is not measurable! The reader is invited to check this and/or to construct a different counterexample.

Proposition 1.34. *If X is a right (or left) continuous adapted process on a FMS $(\Omega, \mathcal{F}, \mathbb{F})$ taking values in a metric space (E,d) with Borel σ-algebra $\mathcal{E}$, then X is progressive.*

Proof. Assume that X is right continuous and adapted. Fix $t \geq 0$. For $n \in \mathbb{N}$ define

$$X_s^n := \begin{cases} X_{\frac{k+1}{2^n}t}, & s \in \left[\frac{k}{2^n}t, \frac{k+1}{2^n}t\right), \ 0 \leq k < 2^n \\ X_t & s = t. \end{cases}$$

For any $B \in \mathcal{E}$ we have

$$\begin{aligned} &\{(s,\omega) \in [0,t] \times \Omega : X_s^n(\omega) \in B\} \\ &\quad = \bigcup_{0 \leq k < 2^n} \left(\left[\frac{k}{2^n}t, \frac{k+1}{2^n}t\right) \times \left\{ X_{\frac{k+1}{2^n}t}(\omega) \in B \right\} \right) \cup (\{t\} \times \{X_t(\omega) \in B\}) \in \mathcal{B}[0,t] \otimes \mathcal{F}_t, \end{aligned}$$

so X^n is progressive. Further, $\lim_{n\to\infty} X_s^n(\omega) = X_s(\omega)$ for every $\omega \in \Omega$ and $s \in [0,t]$ by right continuity. The result follows using Lemma 1.32. The result for left continuous processes follows similarly. In this case we define

$$X_s^n := \begin{cases} X_{\frac{k}{2^n}t}, & s \in \left(\frac{k}{2^n}t, \frac{k+1}{2^n}t\right], \ 0 \leq k < 2^n \\ X_0 & s = 0. \end{cases}$$

The rest of the proof is as in the right continuous case. □

Definition 1.35. Let $(\Omega, \mathcal{F})$ be a measurable space and $\mu \in \mathcal{M}_1(\Omega)$. We denote by $\mathcal{F}^\mu$ the completion of $\mathcal{F}$ with respect to μ, i.e. the smallest σ-algebra which contains $\mathcal{F}$ and all subsets of μ-negligible sets in $\mathcal{F}$. Then $\mathcal{F}^u := \cap_{\mu \in \mathcal{M}_1(\Omega,\mathcal{F})} \mathcal{F}^\mu$ is called the *universal completion* of $\mathcal{F}$.

We state the following important and deep *projection theorem*. For a proof, see e.g. [He79] or [DM78].

Theorem 1.36. *Let $(\Omega, \mathcal{F})$ be a measurable space and let $\mathcal{R}$ be the Borel-σ-algebra of a Polish (i.e. complete separable metric) space R. If $A \in \mathcal{R} \otimes \mathcal{F}$, then*

$$\mathrm{pr}_\Omega(A) \in \mathcal{F}^u,$$

where $\mathrm{pr}_\Omega(A) := \{\omega \in \Omega : \exists s \in R : (s, \omega) \in A\}$ *is the* projection *of A onto Ω.*

Theorem 1.37. *Let $(\Omega, \mathcal{F}, \mathbb{F})$ be a FMS such that $\mathcal{F}_t = \mathcal{F}_t^u$ for each $t \geq 0$. Let X be $(E, \mathcal{E})$-valued progressive and let $A \in \mathcal{E}$. Then both S^A and T^A are weak stopping times.*

Proof. For each $t > 0$ we have to show that $\{S^A < t\}$ and $\{T^A < t\}$ are in $\mathcal{F}_t$. To show this for T^A define

$$A_t := \{(s, \omega) : 0 \leq s < t,\, X_s(\omega) \in A\} = \bigcup_{n \in \mathbb{N}} \big\{(s, \omega) : 0 \leq s \leq t - \frac{1}{n},\, X_s(\omega) \in A\big\},$$

which is in $\mathcal{B}[0, t] \otimes \mathcal{F}_t$. Therefore,

$$\{T^A < t\} = \mathrm{pr}_\Omega(A_t) \in \mathcal{F}_t$$

by Theorem 1.36. The proof for S^A is similar: just consider

$$\tilde{A}_t := \{(s, \omega) : 0 < s < t,\, X_s(\omega) \in A\} = A_t - \{(0, \omega) : w \in \Omega\}$$

which is also in $\mathcal{B}[0, t] \otimes \mathcal{F}_t$. □

Now we provide an example of an adapted and measurable process which is not progressive.

Example 1.38. Let $(\Omega, \mathcal{F}) = ([0, 1], \mathcal{L})$, where $\mathcal{L}$ denotes the σ-algebra of Lebesgue sets in $[0, 1]$ which – by definition – is the completion of $\mathcal{B}[0, 1]$ with respect to Lebesgue measure. Let $\mathcal{L}_0$ be the σ-algebra on $[0, 1]$ containing all sets in $\mathcal{L}$ which have Lebesgue measure 0 or 1. Define $\mathcal{F}_t := \mathcal{L}_0$ for all $t \geq 0$. Define the set $A \subset [0, \infty) \times \Omega$ by $A = \{(x, x) : x \in [0, 1/2]\}$. Then $A \in \mathcal{B}[0, \infty) \otimes \mathcal{F}$ but for each $t > 0$, $A \cap ([0, t] \times \Omega) \notin \mathcal{B}[0, t] \otimes \mathcal{F}_t$ (otherwise the projection of the intersection onto Ω would be in $\mathcal{F}$ by Theorem 1.36 which is however not the case). Therefore the indicator of A is measurable and (check!) adapted but not progressive.

1.3 Martingales

We will assume that the reader is familiar with discrete time martingales.

Definition 1.39. Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a FPS. A real-valued stochastic process $M = (M_t)_{t \geq 0}$ is called an $\mathbb{F}$-*martingale* (resp. $\mathbb{F}$-*submartingale* resp. $\mathbb{F}$-*supermartingale*) if

(i) $(M_t)_{t \geq 0}$ is adapted.

(ii) $\mathbb{E}|M_t| < \infty$ for each $t \geq 0$.

(iii) For each $0 \leq s \leq t < \infty$ we have

$$\begin{aligned} \mathbb{E}(M_t|\mathcal{F}_s) &= M_s \quad \text{a.s.} \\ (\textit{resp.} &\geq \\ \textit{resp.} &\leq) \end{aligned}$$

Example 1.40. Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a FPS. An $\mathbb{F}$-adapted real-valued process B is called an $\mathbb{F}$-Brownian motion if

(i) $\mathbb{P}(B_0 = 0) = 1$

(ii) For every $0 \leq s < t < \infty$, $B_t - B_s$ is independent of $\mathcal{F}_s$.

(iii) For every $0 \leq s < t < \infty$, $\mathcal{L}(B_t - B_s) = \mathcal{N}(0, t - s)$.

(iv) B has continuous paths.

Note that a Brownian motion as defined in *Wahrscheinlichkeitstheorie II* (hereafter abbreviated WT2) is an $\mathbb{F}$-Brownian motion with respect to the filtration generated by B. Any $\mathbb{F}$-Brownian motion is an $\mathbb{F}$-martingale (even if it only satisfies (i)-(iii) and not (iv)): for $0 \leq s < t < \infty$ we have (almost surely)

$$\mathbb{E}(B_t|\mathcal{F}_s) = \mathbb{E}(B_t - B_s + B_s|\mathcal{F}_s) = \mathbb{E}(B_t - B_s|\mathcal{F}_s) + \mathbb{E}(B_s|\mathcal{F}_s) = 0 + B_s = B_s.$$

Further, the process $M_t := B_t^2 - t$, $t \geq 0$ is a martingale, since for $0 \leq s \leq t < \infty$, we have

$$\mathbb{E}(M_t|\mathcal{F}_s) = \mathbb{E}\big((B_t - B_s + B_s)^2|\mathcal{F}_s\big) - t = \mathbb{E}(B_t - B_s)^2 + B_s^2 + 2B_s\mathbb{E}(B_t - B_s) - t = M_s.$$

Another example, the compensated Poisson process N, will be introduced in class.

The following theorem (which is essentially the same as Theorem 3.8 in [KS91]) states some fundamental results for right continuous submartingales which were shown for discrete time submartingales in WT2. We will answer the question whether (and if so in which sense) every submartingale admits a right continuous modification afterwards.

For a function $f : [0, \infty) \to \mathbb{R}$ and a finite subset $F \subset [0, \infty)$ and $-\infty < a < b < \infty$ we let $U_F(a, b; f)$ be the number of *upcrossings* of f restricted to F of the interval $[a, b]$ as defined in WT2. For a general subset $I \subseteq [0, \infty)$, we define

$$U_I(a, b; f) := \sup\{U_F(a, b; f);\ F \subset I,\ F \text{ finite}\}.$$

Theorem 1.41. *Let X be a submartingale on $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ with right continuous paths and let $[s, u] \subset [0, \infty)$. For real numbers $a < b$, $\lambda > 0$ and $p > 1$ we have*

(i) *Maximal inequality:*

$$\lambda\mathbb{P}\big\{\sup_{s\leq t\leq u} X_t \geq \lambda\big\} \leq \mathbb{E}(X_u^+).$$

(ii) *Upcrossing inequality:*

$$\mathbb{E}U_{[s,u]}(a, b; X(\omega)) \leq \frac{\mathbb{E}(X_u^+) + |a|}{b - a}$$

(iii) *Doob's L^p-inequality: if – in addition – X is nonnegative, then*

$$\mathbb{E}\Big(\sup_{s\le t\le u} X_t\Big)^p \le \Big(\frac{p}{p-1}\Big)^p \mathbb{E}(X_u^p).$$

Proof. (i) Let F be a finite subset of $[s,u]\cap\mathbb{Q}$ which contains s and u. From WT2 we know that for any $\mu>0$ we have

$$\mu\mathbb{P}\big\{\sup_{t\in F} X_t > \mu\big\} \le \mathbb{E}(X_u^+). \tag{1.3.1}$$

Take an increasing sequence F_n of such sets whose union equals $G := ([s,u]\cap\mathbb{Q})\cup\{s,u\}$. The continuity of $\mathbb{P}$ implies that (1.3.1) holds true if F is replaced by G. Further, by right continuity of X the same estimate holds if G is replaced by $[s,u]$. The claim follows if we apply this to $\mu := \lambda - \frac{1}{n}$ and let $n\to\infty$.

(ii) For any finite subset $F\subset[s,u]$ containing u, the inequality

$$\mathbb{E}U_F(a,b;X(\omega)) \le \frac{\mathbb{E}(X_u^+)+|a|}{b-a}$$

was proved in WT2. Defining F_n and G as in part (i) and using right continuity of X the claim follows.

(iii) Again this follows from the corresponding discrete time result as in parts (i) and (ii). □

Obviously the statements of the previous theorem remain true if only *almost all* paths of X are right continuous.

Next, we state and prove the submartingale convergence theorem.

Theorem 1.42. *Let X be a right continuous $\mathbb{F}$-submartingale and assume that $\sup_{t\ge0}\mathbb{E}(X_t^+)<\infty$. Then $X_\infty(\omega) := \lim_{t\to\infty} X_t(\omega)$ exists for almost all $\omega\in\Omega$ and $\mathbb{E}|X_\infty|<\infty$.*

Proof. The proof is as in the discrete time case using the upcrossing inequality in Theorem 1.41 (ii). □

Now we state the optional sampling theorem.

Theorem 1.43 (Optional Sampling)**.** *Let X be a right continuous $\mathbb{F}$-submartingale which can be extended to a submartingale on $[0,\infty]$ (i.e. there exists some $\mathcal{F}_\infty$-measurable integrable X_∞ such that $X_t \le \mathbb{E}(X_\infty|\mathcal{F}_t)$ a.s.). Let $S\le T$ be $\mathbb{F}$-stopping times. Then*

$$\mathbb{E}(X_T|\mathcal{F}_S)\ge X_S \text{ and } \mathbb{E}(X_T|\mathcal{F}_S^+)\ge X_S \text{ a.s.}$$

In the martingale case the first and last inequalities are replaced by equalities.

Proof. Consider the random times

$$S_n(\omega) := \begin{cases} \frac{k}{2^n}, & \text{if } S(\omega)\in\Big[\frac{k-1}{2^n},\frac{k}{2^n}\Big),\ k\in\{1,\dots,n2^n\},\\ \infty & \text{if } S(\omega)\ge n,\end{cases}$$

and analogously for T_n. Both S_n and T_n are decreasing sequences of stopping times and converge to T resp. S for each $\omega \in \Omega$. The discrete time optional sampling theorem tells us that for each $n \in \mathbb{N}$

$$\mathbb{E}(X_{T_n}|\mathcal{F}_{S_n}) \geq X_{S_n},$$

i.e. for each $A \in \mathcal{F}_{S_n}$ (and hence for each $A \in \mathcal{F}_S^+$ since $S_n > S$ on the set $\{S < \infty\}$ implies $\mathcal{F}_S^+ \subseteq \mathcal{F}_{S_n}$)

$$\int_A X_{T_n} \,\mathrm{d}\mathbb{P} \geq \int_A X_{S_n} \,\mathrm{d}\mathbb{P}. \tag{1.3.2}$$

By right continuity of X the integrands converge to X_T resp. X_S. If we were allowed to interchange the limit $n \to \infty$ and the integral on both sides of (1.3.2), then the result would follow. A sufficient condition for the interchange to be legal is that the families X_{T_n} resp. X_{S_n} are uniformly integrable (see WT2). This property is indeed true. For a proof see [KS91], Theorem 1.3.22 using Problem 3.11 the solution of which is given on page 41. □

The following result shows that the right continuity assumption of the filtration is in some sense without loss of generality.

Corollary 1.44. *A right continuous $\mathbb{F}$-submartingale X is also an $\mathbb{F}^+$-submartingale and a right continuous $\mathbb{F}$-martingale X is also an $\mathbb{F}^+$-martingale.*

Proof. It is clear that X is $\mathbb{F}^+$-adapted. To prove the (sub-)martingale property, let $0 \leq s < t < \infty$ and apply the previous theorem to $S = s$ and $T = t$ (note that the previous theorem can be applied even if X cannot be extended to a submartingale on $[0, \infty]$; why?). □

Finally, we state a result about the existence of a right continuous modification without proof. The proof can for example be found in [KS91].

Theorem 1.45. *Let X be an $\mathbb{F}$-submartingale and assume that $\mathbb{F}$ is right continuous. If the function $t \mapsto \mathbb{E}X_t$ is right continuous, then X has a right continuous modification $\overline{X}$ which is also an $\mathbb{F}$-submartingale and for which almost all paths have a finite left limit at every point $t > 0$.*

Example 1.46. Here is an example of an $\mathbb{F}$-martingale which does not have a right continuous modification (and for which therefore $\mathbb{F}$ is not right continuous): let $\Omega = \{a, b\}$ (equipped with the σ-algebra $\mathcal{F}$ containing all subsets of Ω) and let $\mathbb{P}$ be the uniform measure on $(\Omega, \mathcal{F})$. Define $X_t(\omega) = 0$ for $\omega \in \Omega$ and $t \in [0, 1]$, and $X_t(a) = 1$ and $X_t(b) = -1$ for $t > 1$. Let $\mathbb{F}$ be the filtration generated by X. Clearly, $\mathbb{X}$ is a martingale and X does not admit a right continuous modification. Note that $\mathbb{F}$ is not right continuous and that X is *not* an $\mathbb{F}^+$-martingale.

1.4 Semimartingales, Quadratic Variation

For the rest of this chapter and the next chapter on stochastic integration, we will only consider martingales (or semimartingales) with continuous paths. Throughout this section, $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ is a given FPS.

Definition 1.47.
- The class of all real-valued adapted stochastic processes with non-decreasing continuous paths is denoted by $\mathcal{A}^+$.

- The class of all real-valued adapted stochastic processes X with continuous paths of locally bounded variation is denoted by $\mathcal{A}$. Recall that $\mathbb{X}$ has paths of locally bounded variation if
$$V_t(\omega) := \sup_{\Delta} \sum_i |X_{t_i \wedge t}(\omega) - X_{t_{i-1} \wedge t}(\omega)| < \infty$$
holds for all $t \geq 0$ and all $\omega \in \Omega$, where the supremum is taken over all finite subsets $\Delta = \{t_0, ..., t_n\}$.

We call V the *variation process* of $\mathbb{X}$.

Remark 1.48. $X \in \mathcal{A}$ implies $V \in \mathcal{A}^+$.

Lemma 1.49. *$X \in \mathcal{A}$ iff there exist S, $W \in \mathcal{A}^+$ such that $X_t(\omega) = S_t(\omega) - W_t(\omega)$ for all $\omega \in \Omega$.*

Proof. If $X \in \mathcal{A}$, then $S_t := \frac{1}{2}(V_t + X_t)$ and $W_t := \frac{1}{2}(V_t - X_t)$ are both in $\mathcal{A}^+$ and $X_t = S_t - W_t$ holds for all $t \geq 0$. The converse is clear. □

Definition 1.50. An adapted real-valued stochastic process $M = (M_t)_{t \geq 0}$ is called an $\mathbb{F}$-*local martingale*, if there exists a sequence of stopping times $(\tau_n)_{n \in \mathbb{N}}$ such that $\tau_n \uparrow \infty$ almost surely, and such that for every $n \in \mathbb{N}$
$$N_t^n := M_{t \wedge \tau_n}, t \geq 0$$
is a martingale. We denote the family of all local martingales with continuous paths by $\mathcal{M}_{loc}$ and the family of all $M \in \mathcal{M}_{loc}$ for which $M_0(\omega) = 0$ for all $\omega \in \Omega$ by $\mathcal{M}_{loc}^0$. Further, $\mathcal{M}$ denotes the family of all martingales with continuous paths.

Remark 1.51. Every martingale is a local martingale but the converse is not true. One example of a continuous local martingale which is not a martingale is $M_t := 1/|x + B_t|$, where $x \in \mathbb{R}^3 \backslash \{0\}$ and B_t is 3-dimensional Brownian motion (starting at 0). It is however not so easy to prove these facts now (only after we have Itô's formula at our disposal).

The following criterion for a continuous local martingale to be a martingale will be often used in the sequel.

Proposition 1.52. *If $M \in \mathcal{M}_{loc}$ satisfies $\mathbb{E}(\sup_{s \geq 0} |M_s|) < \infty$, then $M \in \mathcal{M}$.*

Proof. Let τ_n, $n \in \mathbb{N}$ and N^n be defined as in Definition 1.50. Then for each $t \geq 0$, $\lim_{n \to \infty} N_t^n = M_t$ almost surely and in L^1 (the latter by dominated convergence since $\mathbb{E}(\sup_{s \geq 0} |M_s|) < \infty$). Hence, for $0 \leq s \leq t < \infty$,
$$\mathbb{E}(M_t | \mathcal{F}_s) = \lim_{n \to \infty} \mathbb{E}(N_t^n | \mathcal{F}_s) = \lim_{n \to \infty} N_s^n = M_s,$$
and the claim follows. □

Remark 1.53. The assumptions of the previous proposition are in particular satisfied when $M \in \mathcal{M}_{loc}$ is bounded in (t, ω). There exist however examples of L^p-bounded $M \in \mathcal{M}_{loc}$ (for arbitrarily large p) which are not in $\mathcal{M}$. Observe however that the condition in the previous proposition can be weakened to "$\mathbb{E}(\sup_{t \geq s \geq 0} |M_s|) < \infty$ for each $t \geq 0$".

The following concept of a semimartingale is fundamental in stochastic analysis. Semimartingales will turn out to be exactly the right class of stochastic integrators.

Definition 1.54. A stochastic process X of the form $X_t = M_t + A_t$ with $M \in \mathcal{M}_{loc}$ and $A \in \mathcal{A}$ is called a (continuous) semimartingale. We denote the family of all continuous semimartingales by $\mathcal{S}$.

The following theorem is very important. Its proof largely follows [Re13].

Theorem 1.55.

$$\mathcal{M}_{loc}^0 \cap \mathcal{A} = \{0\}.$$

In particular, the decomposition of $X \in \mathcal{S}$ in the previous definition is unique if we require in addition that $M_0 = 0$.

Proof. Let $X \in \mathcal{M}_{loc}^0 \cap \mathcal{A}$ and let V_t, $t \geq 0$ be the associated variation process. Let us first assume that – in addition – there exist $C_1, C_2 > 0$ such that

$$|X_t(\omega)| \leq C_1, \quad V_t(\omega) \leq C_2$$

for all $t \geq 0$ and all $\omega \in \Omega$. Let $\varepsilon > 0$, $T_0 := 0$ and

$$T_{i+1} := \inf\{t \geq T_i : |X_t - X_{T_i}| \geq \varepsilon\}, \ i \in \mathbb{N}.$$

Fix $t > 0$ and define $S_i := T_i \wedge t$ (these are stopping times). Then, for $n \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{E}(X_{S_n}^2) &= \mathbb{E}\Big(\sum_{i=0}^{n-1} \big(X_{S_{i+1}}^2 - X_{S_i}^2\big)\Big) = \mathbb{E}\Big(\sum_{i=0}^{n-1} \big(X_{S_{i+1}} - X_{S_i}\big)^2\Big) + 2\mathbb{E}\Big(\sum_{i=0}^{n-1} X_{S_i}\big(X_{S_{i+1}} - X_{S_i}\big)\Big) \\
&= \mathbb{E}\Big(\sum_{i=0}^{n-1} \big(X_{S_{i+1}} - X_{S_i}\big)^2\Big) \leq \varepsilon \mathbb{E}\Big(\sum_{i=0}^{n-1} \big|X_{S_{i+1}} - X_{S_i}\big|\Big) \leq \varepsilon \mathbb{E} V_t \leq \varepsilon C_2,
\end{aligned}$$

where we used the fact that $X \in \mathcal{M}$ (by Proposition 1.52) and therefore the second term after the second equality sign is zero by the optional sampling theorem. Therefore, since $T_n \to \infty$ almost surely, we get – using dominated convergence –

$$\mathbb{E} X_t^2 \leq C_2 \varepsilon$$

for all $t > 0$ and all $\varepsilon > 0$ which implies $X \equiv 0$ almost surely.

For the general case, let τ_n be as in Definition 1.50 and define

$$\begin{aligned}
\tilde{T}_n &: = \inf\{s \geq 0 : |X_s| \geq n\} \\
\hat{T}_n &: = \inf\{s \geq 0 : V_s \geq n\} \\
\check{T}_n &: = \inf\{\tau_n, \tilde{T}_n, \hat{T}_n\}.
\end{aligned}$$

Then, for each $n \in \mathbb{N}$, the process $t \mapsto X_{t \wedge \check{T}_n}$ satisfies the assumptions of the first case with $C_1 = C_2 = n$ which implies $X_{t \wedge \check{T}_n} = 0$ for all $t \geq 0$ almost surely. Since $\check{T}_n \to \infty$ almost surely, the assertion follows. □

Remark 1.56. In the previous theorem, the assumption that the involved processes have continuous paths is important. If we replace *continuous* by *right continuous* in all definitions of this section, then the uniqueness of the decomposition does no longer hold in general: As an example consider the Poisson process N_t which can be decomposed as $N_t = (N_t - t) + t$ and $N_t = 0 + N_t$.

Definition 1.57. Let $\mathcal{D}$ be the family of all $\Delta = \{0 = t_0 < t_1 < ...\}$ such that $t_n \to \infty$. For a stochastic process X and $\Delta \in \mathcal{D}$, we define

$$T_t^\Delta(X) := \sum_{i=0}^{k-1} \left(X_{t_{i+1}} - X_{t_i}\right)^2 + \left(X_t - X_{t_k}\right)^2,\ t \geq 0$$

where k is such that $t_k \leq t < t_{k+1}$. We say that X has *finite quadratic variation* if there exists a process $\langle X, X\rangle$ such that for each $t \geq 0$, T_t^Δ converges in probability to $\langle X, X\rangle_t$ as $|\Delta| \to 0$, where $|\Delta| := \sup\{t_{n+1} - t_n,\ n \in \mathbb{N}_0\}$. We also write $\langle X\rangle_t$ instead of $\langle X, X\rangle_t$.

Remark 1.58. Recall that for Brownian motion B we showed that $B_t^2 - t$ is a martingale. It therefore follows from the next theorem (or rather Theorem 1.62) that $\langle B, B\rangle_t = t$ almost surely. We emphasize however that nevertheless for almost every $\omega \in \Omega$ we can find a sequence $\Delta_n(\omega) \subset \mathcal{D}$ for which $|\Delta_n(\omega)| \to 0$ such that $T_t^{\Delta_n(\omega)}(B(\omega)) \to \infty$. Even for a deterministic sequence $(\Delta_n)_{n\in\mathbb{N}}$ it may happen that $\limsup_{n\to\infty} T_t^{\Delta_n}(B(\omega)) = \infty$ almost surely.

Theorem 1.59. *Let $M \in \mathcal{M}_{loc}^0$ be bounded in (t,ω) (and hence $M \in \mathcal{M}$). Then M has finite quadratic variation and $\langle M, M\rangle$ is the unique process in $\mathcal{A}^+$ vanishing at zero such that $t \mapsto M_t^2 - \langle M, M\rangle_t$ is a martingale.*

Proof. The proof is largely taken from [RY94] (or [Re13]). Uniqueness follows immediately from Theorem 1.55.

Now we show existence of such a process. For $\Delta \in \mathcal{D}$ and $t_i \leq s \leq t_{i+1}$ we have

$$\mathbb{E}\big((M_{t_{i+1}} - M_{t_i})^2|\mathcal{F}_s\big) = \mathbb{E}\big((M_{t_{i+1}} - M_s)^2|\mathcal{F}_s\big) + (M_s - M_{t_i})^2.$$

Therefore, for $0 \leq s < t < \infty$,

$$\begin{aligned}\mathbb{E}\big(T_t^\Delta(M)|\mathcal{F}_s\big) &= T_s^\Delta(M) + \mathbb{E}\big((M_t - M_s)^2|\mathcal{F}_s\big)\\ &= T_s^\Delta(M) - M_s^2 + E\big(M_t^2|\mathcal{F}_s\big),\end{aligned}$$

so $t \mapsto M_t^2 - T_t^\Delta(M)$ is a continuous martingale. $T_t^\Delta(M)$ is not yet our claimed process since it is in general not nondecreasing.

We want to show that $T_t^\Delta(M)$ converges in probability as $|\Delta| \to 0$. Fix $a > 0$. Let $\Delta,\ \Delta' \in \mathcal{D}$ and denote the union of both subdivisions by $\Delta\Delta'$. Then $X_t := T_t^\Delta(M) - T_t^{\Delta'}(M)$ is in $\mathcal{M}$ and bounded on $[0,t] \times \Omega$ for each $t \geq 0$ and therefore

$$t \mapsto X_t^2 - T_t^{\Delta\Delta'}(X)$$

is in $\mathcal{M}$ and vanishes at 0. In particular, we have $\mathbb{E}\big((T_a^\Delta(M) - T_a^{\Delta'}(M))^2\big) = \mathbb{E}\big(X_a^2\big) = \mathbb{E}\big(T_a^{\Delta\Delta'}(X)\big)$. We want to show that

$$\mathbb{E}\big(X_a^2\big) \to 0 \text{ as } |\Delta| + |\Delta'| \to 0. \tag{1.4.1}$$

Assume for a moment that (1.4.1) has been shown (for each $a > 0$) – we refer the reader to [RY94] or [Re13] for the proof. Since $L^2(\Omega, \mathcal{F}, \mathbb{P})$ is complete, there exists a random variable $\langle M, M\rangle_a$ such that $\lim_{|\Delta|\to 0} T_a^\Delta(M) = \langle M, M\rangle_a$ in L^2. All that remains to be shown is that the process $t \mapsto \langle M, M\rangle_t$ has a modification in $\mathcal{A}^+$ which vanishes at 0 and that $t \mapsto M_t^2 - \langle M, M\rangle_t$ is a martingale. Doob's L^2-inequality shows that for each fixed $a > 0$ the processes $T_s^\Delta(M) - \langle M, M\rangle_s$

converge to zero even uniformly on $[0,a]$ in L^2 (i.e. the supremum converges to 0 in L^2) and therefore the limit has a continuous modification. It is also easy to see by approximation that we have $\langle M,M\rangle_0 = 0$ almost surely and that the process is nondecreasing. Being the limit of adapted processes it also has a modification which is adapted. Finally, the martingale property follows since L^2-limits of martingales are martingales. □

We aim at generalizing the previous theorem to arbitrary $M \in \mathcal{M}_{loc}$. To do that we need the following lemma.

Lemma 1.60. *For M as in the previous theorem and a stopping time T define $M_t^T := M_{t\wedge T}$. We have*

$$\langle M^T, M^T\rangle_t = \langle M,M\rangle_{t\wedge T}.$$

Proof. By the optional sampling theorem and Theorem 1.59, the process $t \mapsto M^2_{T\wedge t} - \langle M,M\rangle_{t\wedge T}$ is a martingale. The claim of the lemma now follow using the uniqueness statement in Theorem 1.59. □

Definition 1.61. If X^n and Y are real-valued stochastic processes (indexed by $[0,\infty)$), then we write $\lim_{n\to\infty} X^n = Y$ *ucp*, if for each $T > 0$, we have that $\sup_{0\le s\le T} |X_s^n - Y_s|$ converges to 0 in probability as $n \to \infty$ (*ucp* stands for *uniformly on compact sets in probability*).

Theorem 1.62. *Let $M \in \mathcal{M}^0_{loc}$. There exists a unique process $\langle M,M\rangle \in \mathcal{A}^+$ which vanishes at 0 such that*

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

Further, $\langle M,M\rangle = \lim_{|\Delta|\to 0} T^\Delta(M)$ ucp.

Proof. Uniqueness follows from Theorem 1.55. Define $T_n := \inf\{t \ge 0 : |M_t| = n\}$ and $M^n := M^{T_n}$. Then M^n is a bounded continuous martingale. By Theorem 1.59, there exists a process $\langle M^n, M^n\rangle$ such that

$$\big(M^n\big)^2 - \langle M^n, M^n\rangle$$

is a martingale. For $m \le n$, we have $(M^n)^{T_m} = M^m$ and therefore, by Lemma 1.60,

$$\langle M^n, M^n\rangle_{t\wedge T_m} = \langle M^m, M^m\rangle_t.$$

For fixed $t \ge 0$ and on the set $\{\omega : T_m(\omega) \ge t\}$ we define

$$\langle M,M\rangle_t := \langle M^m, M^m\rangle_t.$$

Note that $\langle M,M\rangle$ is well-defined and equals $\lim_{n\to\infty}\langle M^n,M^n\rangle$. Further, since $\big(M^n\big)^2 - \langle M^n, M^n\rangle$ is a martingale, by taking the limit $n\to\infty$, we see that $M^2 - \langle M,M\rangle \in \mathcal{M}_{loc}$.

To show the second part of the theorem, we fix $\delta > 0$ and $t > 0$. Defining T_n as above, we pick some $k \in \mathbb{N}$ such that

$$\mathbb{P}(T_k < t) \le \delta.$$

On the set $\{s \le T_k\}$, we have

$$\langle M,M\rangle_s = \langle M,M\rangle_{s\wedge T_k} = \langle M^{T_k}, M^{T_k}\rangle_s.$$

Therefore, for $\varepsilon > 0$,

$$\begin{aligned}
\mathbb{P}\Big(\sup_{s\le t}\big|T_s^\Delta(M) - \langle M,M\rangle_s\big| \ge \varepsilon\Big) &\le \delta + \mathbb{P}\Big(\sup_{s\le T_k}\big|T_s^\Delta(M) - \langle M,M\rangle_s\big| \ge \varepsilon,\, T_k \ge t\Big)\\
&= \delta + \mathbb{P}\Big(\sup_{s\le T_k}\big|T_s^\Delta(M^{T_k}) - \langle M^{T_k},M^{T_k}\rangle_s\big| \ge \varepsilon,\, T_k \ge t\Big)\\
&\le \delta + \mathbb{P}\Big(\sup_{s\le T_k}\big|T_s^\Delta(M^{T_k}) - \langle M^{T_k},M^{T_k}\rangle_s\big| \ge \varepsilon\Big).
\end{aligned}$$

Since M^{T_k} is a bounded martingale, Theorem 1.59 implies

$$\limsup_{|\Delta|\to 0}\mathbb{P}\Big(\sup_{s\le t}\big|T_s^\Delta(M) - \langle M,M\rangle_s\big| \ge \varepsilon\Big) \le \delta.$$

Since $\delta, \varepsilon > 0$ are arbitrary, the asserted ucp convergence follows. □

Next, we define the *bracket* of two local martingales.

Theorem 1.63. *Let $M, N \in \mathcal{M}^0_{loc}$. There exists a unique process $\langle M,N\rangle \in \mathcal{A}$ vanishing at 0 such that*

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

Further, we have

$$\lim_{|\Delta|\to 0} T^\Delta(M,N) = \langle M,N\rangle \text{ ucp},$$

where

$$T_s^\Delta(M,N) := \sum (M_{t_i\wedge s} - M_{t_{i-1}\wedge s})\cdot(N_{t_i\wedge s} - N_{t_{i-1}\wedge s}).$$

Proof. Define

$$\langle M,N\rangle := \frac{1}{4}\big(\langle M+N,M+N\rangle - \langle M-N,M-N\rangle\big).$$

Since

$$M\cdot N = \frac{1}{4}\big((M+N)^2 - (M-N)^2\big),$$

we obtain from Theorem 1.62 the fact that $M\cdot N - \langle M,N\rangle \in \mathcal{M}_{loc}$. Since $\langle M,N\rangle$ is the difference of two processes in $\mathcal{A}^+$ it belongs to $\mathcal{A}$ (but not necessarily to $\mathcal{A}^+$!). The rest follows as before. □

Proposition 1.64. *If $X \in \mathcal{S}$ has a decomposition $X = M + A$ with $M \in \mathcal{M}^0_{loc}$ and $A \in \mathcal{A}$, then $T^\Delta(X) \to \langle M,M\rangle$ ucp.*

Proof. We have

$$T_s^\Delta(X) = \sum_i \big(X_{t_i\wedge s} - X_{t_{i-1}\wedge s}\big)^2 = T_s^\Delta(M) + 2T_s^\Delta(M,A) + T_s^\Delta(A).$$

Now

$$\begin{aligned}
\big|T_s^\Delta(M,A)\big| &= \Big|\sum_i \big(M_{t_i\wedge s} - M_{t_{i-1}\wedge s}\big)\cdot\big(A_{t_i\wedge s} - A_{t_{i-1}\wedge s}\big)\Big|\\
&\le \sup_{t_i\in\Delta, t_i\le s}|M_{t_i} - M_{t_{i-1}}|\cdot\sum_i|A_{t_i\wedge s} - A_{t_{i-1}\wedge s}| \to 0
\end{aligned}$$

since M has continuous paths (in fact the convergence is almost surely uniform on compact sets). The fact that $T_s^{\Delta}(A) \to 0$ follows in the same way. The proposition now follows from Theorem 1.62. $\square$

Definition 1.65. Let
$$\mathcal{H} := \left\{ M \in \mathcal{M} : \sup_{t \geq 0} (\mathbb{E}|M_t|^2) < \infty \right\}$$
be the set of all L^2-bounded continuous martingales. The meaning of an additional upper index 0 should be clear. (Many authors use the symbol H^2 or $\mathcal{H}^2$ to denote this space).

Theorem 1.66. 1. *If $M \in \mathcal{H}$, then M converges to some M_∞ almost surely and in L^2.*

2. *$\mathcal{H}$ is a real Hilbert space (after forming equivalence classes) if equipped with the norm*
$$\|M\|_{\mathcal{H}} := \lim_{t \to \infty} \mathbb{E}(|M_t|^2)^{1/2} = \mathbb{E}(|M_\infty|^2)^{1/2}$$

3. *The following norm on $\mathcal{H}$ is equivalent to $\|.\|_{\mathcal{H}}$:*
$$\|M\| := \mathbb{E}(\sup_{t \geq 0} |M_t|^2)^{1/2}$$

4. *If $M \in \mathcal{H}^0$, then $\langle M, M\rangle_\infty := \lim_{t\to\infty} \langle M, M\rangle_t$ exists almost surely and $\|M\|_{\mathcal{H}}^2 = \mathbb{E}\langle M, M\rangle_\infty$.*

Proof. 1. This follows from the submartingale convergence theorem (Theorem 1.42) together with results from WT2 (L^2-boundedness of a martingale implies L^2-convergence).

2. The second equality follows from L^2-convergence. Since $L^2(\Omega, \mathcal{F}_\infty, \mathbb{P})$ is a Hilbert space, it only remains to show that $\mathcal{H}$ is closed (or complete). To see this pick a Cauchy sequence M^n in $\mathcal{H}$. Then $M_\infty := \lim_{n\to\infty} M_\infty^n$ exists in L^2. Using Doob's L^2-inequality (Theorem 1.41 (iii)), we see that $M_t := \mathbb{E}(M_\infty|\mathcal{F}_t)$ has a continuous modification and is therefore in $\mathcal{H}$.

3. This follows from $\mathbb{E}(\sup_{t\geq 0} M_t^2) \leq 4\mathbb{E}M_\infty^2 \leq 4\mathbb{E}(\sup_{t \geq 0} M_t^2)$, where the first inequality is an application of Doob's L^2-inequality (Theorem 1.41 (iii)).

4. We know from Theorem 1.62 that $M_t^2 - \langle M, M\rangle_t \in \mathcal{M}_{loc}^0$. Let $T_n := \inf\{t \geq 0 : |M_t| = n\}$. Then
$$\mathbb{E}M_{t\wedge T_n}^2 = \mathbb{E}\langle M, M\rangle_{t \wedge T_n}.$$
Letting $n \to \infty$ and $t \to \infty$, the right hand side converges to $\mathbb{E}\langle M, M\rangle_\infty$ by monotone convergence. Further $\mathbb{E}M_\infty^2 \leq \liminf_{n,t\to\infty} \mathbb{E}M_{t\wedge T_n}^2$ by Fatou's lemma and $\mathbb{E}M_\infty^2 \geq \limsup_{n,t\to\infty} \mathbb{E}M_{t\wedge T_n}^2$ follows from the optional sampling theorem applied to the submartingale M^2.

$\square$

Chapter 2

Stochastic Integrals and Stochastic Differential Equations

2.1 The stochastic integral

Throughout this section $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ will be a FPS. In addition we assume that the *usual conditions* introduced earlier hold. We aim at defining the stochastic integral process $t \mapsto \int_0^t f_s(\omega)\,\mathrm{d}S_s$ for $S \in \mathcal{S}^0$ and a reasonably large class of integrand processes f. It turns out that this cannot be done in a *pathwise* manner. Since each $S \in \mathcal{S}^0$ can be uniquely decomposed as $S = M + A$ with $M \in \mathcal{M}^0_{loc}$ and $A \in \mathcal{A}^0$, we "define"

$$\int_0^t f_s(\omega)\,\mathrm{d}S_s := \int_0^t f_s(\omega)\,\mathrm{d}M_s + \int_0^t f_s(\omega)\,\mathrm{d}A_s.$$

So far neither of the two integrals on the right hand side has been defined. The second one (a so called *Lebesgue-Stieltjes integral*) turns out to be much easier than the first, so we start by defining the second integral. This can be done pathwise, i.e. for each $\omega \in \Omega$ separately (only afterwards will we care about measurability properties with respect to ω).

Let $A \in \mathcal{A}^0$, i.e. $A \in \mathcal{A}$ and $A_0 = 0$. Think of ω being fixed, so A is just a continuous function from $[0, \infty)$ to $\mathbb{R}$ of locally bounded variation. A can be written as $A_t = A_t^+ - A_t^-$, where both A^+ and A^- are continuous, nondecreasing and vanishing at 0, so they correspond to σ-finite measures μ^+ and μ^- on $(0, \infty)$ via $A_t^+ = \mu^+(0, t]$ and $A^- = \mu^-(0, t]$. This allows us to define

$$\int_0^t f_s\,\mathrm{d}A_s := \int_0^t f_s\,\mathrm{d}\mu^+(s) - \int_0^t f_s\,\mathrm{d}\mu^-(s),$$

whenever f is measurable with respect to s and the right hand side is well-defined (i.e. is not of the form $\infty - \infty$ or $-\infty - (-\infty)$). We will in fact assume that f is chosen such that both integrals are finite. Note that the decomposition of $A = A^+ - A^-$ is not unique (and even the integrability of f may depend on the choice of the decomposition), but the value of the left hand side does not depend on the particular decomposition and there is always a unique minimal decomposition, i.e. one for which $V_t = A_t^+ + A_t^-$ is the variation process (cf. the proof of Lemma 1.49). The integral is then automatically continuous in t (since the measures μ^+ and μ^- are atomless). We want to ensure that the integral is also adapted. This property will hold when f is progressive and – for example – locally bounded for almost every $\omega \in \Omega$ (this is a version of Fubini's theorem which says that the integral of a jointly measurable function is measurable with respect to the remaining variable – it was shown in WT2 in connection with the proof of Ionesco Tulcea's theorem).

In the following we will also write $f \cdot A$ instead of $\int_0^{\cdot} f_s \, \mathrm{d}A_s$.

We will now proceed to define the *stochastic integral* $\int_0^t f_s(\omega) \, \mathrm{d}M_s$. This will be done in several steps. We first assume that $M \in \mathcal{H}^0$ and f is simple in the sense of the following definition. Then we extend to more general f and to general $M \in \mathcal{M}^0_{loc}$.

Definition 2.1. The stochastic process $f : [0, \infty) \times \Omega \to \mathbb{R}$ is called a *simple process*, in short $f \in \mathcal{L}_s$, if there exist $\Delta = \{0 = t_0 < t_1 < ...\}$ such that $t_i \to \infty$ and $C < \infty$ and $\mathcal{F}_{t_i}$-measurable ξ_i, $i = 0, 1, ...$ such that $\sup_{n \geq 0} |\xi_n(\omega)| \leq C$ for all $\omega \in \Omega$ and

$$f_t(\omega) = \xi_0(\omega) \mathbb{1}_{\{0\}}(t) + \sum_{i=0}^{\infty} \xi_i(\omega) \mathbb{1}_{(t_i, t_{i+1}]}(t); \quad 0 \leq t < \infty, \, \omega \in \Omega. \tag{2.1.1}$$

For $M \in \mathcal{H}^0$ and $f \in \mathcal{L}_s$ as in (2.1.1), we define the stochastic integral in an obvious way:

$$I_t(f) := \int_0^t f_s \, \mathrm{d}M_s := \sum_{i=0}^{\infty} \xi_i \big(M_{t \wedge t_{i+1}} - M_{t \wedge t_i} \big), \quad 0 \leq t < \infty.$$

We will also write $f \cdot M$ instead of $I_{\cdot}(f)$. Note that the value of f_0 has no relevance for the stochastic integral.

Proposition 2.2. *For $M \in \mathcal{H}^0$ and $f \in \mathcal{L}_s$ we have*

i) $f \cdot M \in \mathcal{H}^0$.

ii) $\langle f \cdot M \rangle_t = \int_0^t f_s^2 \, \mathrm{d}\langle M \rangle_s$.

iii) $\|f \cdot M\|_{\mathcal{H}}^2 = \mathbb{E} \int_0^\infty f_s^2 \, \mathrm{d}\langle M \rangle_s$.

Proof. i) $f \cdot M$ is clearly adapted, has continuous paths, vanishes at 0 and is L^2-bounded. To show the martingale property, take $0 \leq s < t < \infty$ and define k, n such that $s \in [t_{k-1}, t_k)$ and $t \in [t_n, t_{n+1})$. If $k \leq n$, Then

$$(f \cdot M)_t - (f \cdot M)_s = \sum_{i=k}^{n-1} \xi_i \big(M_{t_{i+1}} - M_{t_i} \big) + \xi_n \big(M_t - M_{t_n} \big) + \xi_{k-1} \big(M_{t_k} - M_s \big).$$

The conditional expectation of the right hand side given $\mathcal{F}_s$ is 0. If $k = n+1$, then

$$\mathbb{E}\big((f \cdot M)_t - (f \cdot M)_s | \mathcal{F}_s \big) = \xi_n \mathbb{E}\big(M_t - M_s | \mathcal{F}_s \big) = 0$$

and therefore the martingale property follows.

ii) Since the right hand side of the claimed equality is in $\mathcal{A}^+$ and vanishes at 0, it suffices to show that

$$t \mapsto (f \cdot M)_t^2 - \int_0^t f_s^2 \, \mathrm{d}\langle M \rangle_s$$

is a martingale. Both summands are integrable. For $0 \le s < t < \infty$ we have, assuming without loss of generality that s and t are in the partition, say $s = t_k$ and $t = t_{n+1}$,

$$\begin{aligned}\mathbb{E}\big((f\cdot M)_t^2 - (f\cdot M)_s^2|\mathcal{F}_s\big) &= \mathbb{E}\big(\big((f\cdot M)_t - (f\cdot M)_s\big)^2|\mathcal{F}_s\big)\\ &= \mathbb{E}\big(\big(\sum_{i=k}^n \xi_i(M_{t_{i+1}} - M_{t_i})\big)^2|\mathcal{F}_s\big)\\ &= \mathbb{E}\big(\sum_{i=k}^n \xi_i^2(M_{t_{i+1}} - M_{t_i})^2|\mathcal{F}_s\big)\\ &= \mathbb{E}\big(\int_s^t f_r^2\,\mathrm{d}\langle M\rangle_r|\mathcal{F}_s\big),\end{aligned}$$

so the claim follows.

iii) This follows from part ii) and part 4. of Theorem 1.66.

□

Property iii) of the previous proposition states that I is an isometry from $\mathcal{L}_s$ to $\mathcal{H}^0$ if $\mathcal{L}_s$ is equipped with the corresponding norm. To formulate this precisely, we introduce the *Doleans measure* and various function spaces.

Definition 2.3. Let $M \in \mathcal{H}^0$. The *Doleans measure* μ_M on $\big([0,\infty)\times\Omega,\, \mathcal{B}[0,\infty)\otimes\mathcal{F}\big)$ is defined as

$$\mu_M(A) := \mathbb{E}\int_0^\infty \mathbb{1}_A(s,\omega)\,\mathrm{d}\langle M\rangle_s.$$

Note that μ_M is a finite measure (with total mass $\mathbb{E}\langle M\rangle_\infty$).

Definition 2.4. Let $\bar{\mathcal{L}}(M) := \mathcal{L}^2\big([0,\infty)\times\Omega,\, \mathcal{B}[0,\infty)\otimes\mathcal{F},\, \mu_M\big)$.

Note that $\mathcal{L}_s$ is a subspace of $\bar{\mathcal{L}}(M)$. We will work with several subspaces of $\bar{\mathcal{L}}(M)$ below. As usual, we denote the corresponding spaces of equivalence classes by $L(M)$ etc. We will always assume that $M \in \mathcal{H}^0$ is given. Observe that $\bar{L}(M)$ is a Hilbert space. We will denote the corresponding norm by $\|.\|_M$. The following corollary is a reformulation of part iii) of Proposition 2.2.

Corollary 2.5. *$I : L_s \to \mathcal{H}^0$ defined above is an isometry (if L_s is equipped with the norm $\|.\|_M$).*

An isometry from some metric space to a complete metric space can always be extended uniquely to an isometry on the completion of the first space. Denote the completion of L_s in the Hilbert space $\bar{L}(M)$ by $L_*(M)$. We can (and will) simply *define* the stochastic integral $I_t(f)$ for $f \in L_*(M)$ as the image of the unique extension of I to a (linear) isometry from $L_*(M)$ to $\mathcal{H}^0$ which we again denote by I. It is of course of interest to have a more explicit description of the space $L_*(M)$. We will deal with this question soon. Observe however that it is definitely *not* true (in general) that $L_*(M) = \bar{L}(M)$: if $L_{**}(M)$ denotes the closed subspace of $\bar{L}(M)$ consisting of all progressive processes in $\bar{L}(M)$ (more precisely: consisting of all equivalence classes containing some progressive process) then $L_{**}(M)$ contains L_s (since left continuous adapted processes are progressive) and therefore also its completion $L_*(M)$. We point out that it may well happen that an equivalence class of $\bar{L}(M)$ contains both progressive and non-progressive processes: just think of the extreme case in which $M \equiv 0$ and therefore $\bar{L}(M) = \{0\}$ (then $\mu_M = 0$ and all processes in $\bar{\mathcal{L}}(M)$ are equivalent).

One approach to identifying the space $L_*(M)$ is to consider the sub-σ-algebra of $\mathcal{B}[0,\infty) \otimes \mathcal{F}$ which is generated by all processes in $\mathcal{L}_s$. This sub-σ-algebra is called the *predictable* σ-algebra. It is contained in the progressive σ-algebra (since each left continuous adapted process is progressive). It is then not hard to show that any predictable process in $\bar{\mathcal{L}}(M)$ can be approximated by a sequence of processes from $\mathcal{L}_s$ with respect to the norm on $\bar{L}(M)$ (or semi-norm on $\bar{\mathcal{L}}(M)$) – we will show this below – and therefore $L_*(M)$ is precisely the space of equivalence classes which contain at least one predictable process which is in $\mathcal{L}(M)$. Even though by far not every predictable process is progressive, Lemma 3.2.7 in [KS91] (the proof of which is neither hard nor short) shows that each equivalence class in $\bar{L}(M)$ which contains a predictable process also contains a progressive process (i.e. $L_*(M) = L_{**}(M)$) and in this sense we have defined the stochastic integral (also called *Itô's integral*) for all progressive integrands in $\bar{L}(M)$.

We point out that if one likes to introduce stochastic integrals with respect to general (right continuous but not necessarily continuous) semimartingales, then one has to restrict the class of integrands to predictable ones since then the corresponding equality $L_*(M) = L_{**}(M)$ does not hold in general.

Definition 2.6. The σ-algebra on $(0,\infty) \times \Omega$ which is generated by all sets of the form $(s,t] \times A$ where $0 \le s \le t < \infty$ and $A \in \mathcal{F}_s$ is called the *predictable* σ-algebra. A stochastic process is called predictable if it is measurable with respect to the σ-algebra of predictable sets.

Proposition 2.7. *If for $1 \le i \le k$ we have $M^i \in \mathcal{H}^0$ and f is predictable and in each of the spaces $\bar{\mathcal{L}}(M^i)$, then there exists a sequence of simple processes f^n such that $\|f - f^n\|_{M^i} \to 0$ for all i.*

Proof. Define

$$\mathcal{R} := \{f \text{ bounded predictable: there exist } f^n \text{ as claimed}\}.$$

Clearly, $\mathcal{R}$ is a linear space which contains the constant process $\mathbb{1}$. The family of indicators of $(s,t] \times A$ where $0 \le s \le t < \infty$ and $A \in \mathcal{F}_s$ belongs to $\mathcal{R}$, is closed under multiplication, and generates the predictable σ-algebra. The monotone class theorem (see appendix) states that $\mathcal{R}$ is equal to the set of all bounded predictable functions provided we can show that $\mathcal{R}$ is closed with respect to limits of increasing, nonnegative and uniformly bounded elements of $\mathcal{R}$. To show this, take a sequence of nonegative $g^n \in \mathcal{R}$ such that $g^n \uparrow g$ and g is bounded (in (t,ω)). Being the pointwise limit of predictable processes g is also predictable. Further, we have $\|g^n - g\|_{M^i} \to 0$ by dominated convergence and it follows from the usual "triangle trick" that $g \in \mathcal{R}$.

To complete the proof, let f be as in the proposition and define $f^{(N)} := f\mathbb{1}_{|f|<N}$. Then $f^{(N)}$ is bounded and predictable and hence in $\mathcal{R}$ and $\|f - f^{(N)}\|_{M^i} \to 0$ for each i by dominated convergence. Again the triangle trick shows that the claim in the proposition holds for f. $\square$

Before proceeding to the next step of the construction of the Itô integral with respect to a continuous semimartingale we state a few properties of the integral defined so far. First note that for general $M, N \in \mathcal{M}^0_{loc}$, we have for each $s < t$ almost surely

$$|\langle M,N\rangle_t - \langle M,N\rangle_s| \le \sqrt{\langle M\rangle_t - \langle M\rangle_s}\sqrt{\langle N\rangle_t - \langle N\rangle_s} \tag{2.1.2}$$

which is just the usual Cauchy-Schwarz inequality which holds true because $\langle .,. \rangle$ is (almost surely) bilinear, symmetric and positive semi-definite. This implies for $M, N \in \mathcal{H}^0$ (using part 4. of Theorem 1.66).

$$\mathbb{E} \sup_{0 \le t < \infty} |\langle M, N \rangle_t| \le \|M\|_{\mathcal{H}} \|N\|_{\mathcal{H}}. \tag{2.1.3}$$

Note that item b) in the following proposition follows directly from the isometry property of the stochastic integral in case $M = N$ (no matter if f and g agree or not) since an isometry between Hilbert spaces automatically preserves the inner product.

Proposition 2.8 (Kunita-Watanabe identity and inequality). *Let $M, N \in \mathcal{H}^0$ and $f \in L_*(M)$, $g \in L_*(N)$. Then*

a) $\big|((fg) \cdot \langle M, N \rangle)_t\big|^2 \le \big(f^2 \cdot \langle M \rangle\big)_t \big(g^2 \cdot \langle N \rangle\big)_t$ *a.s.*

b) $\langle f \cdot M, g \cdot N \rangle = (fg) \cdot \langle M, N \rangle$

c) $\mathbb{E}\big|\langle f \cdot M, g \cdot N \rangle_t\big| \le \sqrt{\mathbb{E}\big(f^2 \cdot \langle M \rangle\big)_t} \sqrt{\mathbb{E}\big(g^2 \cdot \langle N \rangle\big)_t}$.

Proof. a) This is easy to show for $f, g \in \mathcal{L}_s$ (see e.g. [Ku90]) and then for general f, g by approximation (see e.g. [KS91], Proposition 3.2.14).

b) For $f, g \in L_s$ the claim in part b) follows just like part ii) of Proposition 2.2. For general $f \in L_*(M)$, $g \in L_*(N)$ approximate by f^n, $g^n \in L_s$ in $L_*(M)$ resp. $L_*(N)$ and use (2.1.3) and part a) (exercise).

c) This is just a combination of parts a) and b) using the Cauchy Schwarz inequality.

□

We are now ready for the final step in the construction. For $M \in \mathcal{M}^0_{loc}$ let

$$\mathcal{L}_{loc}(M) := \{f \text{ progressive} : \int_0^t f_s^2 \, \mathrm{d}\langle M \rangle_s < \infty \text{ for all } t \text{ a.s.}\}.$$

Definition 2.9. For $M \in \mathcal{M}^0_{loc}$ and $f \in \mathcal{L}_{loc}(M)$ let S_n be a sequence of stopping times such that $S_n \to \infty$ and $t \mapsto M_{t \wedge S_n}$ is in $\mathcal{H}^0$. Further define $R_n := n \wedge \inf\{t \ge 0 : \int_0^t f_s^2 \, \mathrm{d}\langle M \rangle_s \ge n\}$. Define $T_n := R_n \wedge S_n$. Then $T_n \to \infty$ almost surely and the T_n are stopping times such that $M^n_t := M_{t \wedge T_n}$ is in $\mathcal{H}^0$ and $f^n_t := f_t 1\!\!1_{T_n \ge t}$ is in $\mathcal{L}_*(M^n)$. We define

$$(f \cdot M)_t := \int_0^t f_s \, \mathrm{d}M_s := (f^n \cdot M^n)_t \text{ on } \{T_n(\omega) \ge t\}.$$

We have to make sure that the integral is well-defined. Note that once this has been clarified the stochastic integral process is again in $\mathcal{M}^0_{loc}$. The fact that the stochastic integral is well-defined follows from the following proposition (see [KS91], Corollary 3.2.21 for the (easy) proof).

Proposition 2.10. *Let $M, N \in \mathcal{H}^0$ and $f \in L_*(M)$, $g \in L_*(N)$ and assume that there exists a stopping time T such that almost surely*

$$f_{t \wedge T(\omega)}(\omega) = g_{t \wedge T(\omega)}(\omega); \; M_{t \wedge T(\omega)}(\omega) = N_{t \wedge T(\omega)}(\omega) \text{ for all } t \ge 0.$$

Then, for almost all $\omega \in \Omega$

$$(f \cdot M)_{t \wedge T(\omega)} = (g \cdot N)_{t \wedge T(\omega)} \text{ for all } t \ge 0.$$

Roughly speaking, all properties derived for stochastic integrals in case $M \in \mathcal{H}^0$ and $f \in L_*(M)$ which do not involve expected values carry over to the general case (and are easily proved). Those properties which do involve expected values will generally not carry over.

Our next aim is to formulate and prove Itô's formula, the chain rule of stochastic calculus. To do this, we require a few further properties. The following is a kind of Riemann approximation result.

Theorem 2.11. *For each $n \in \mathbb{N}$, let $\Delta_n \in \mathcal{D}$ such that $|\Delta^n| \to 0$. Let $M \in \mathcal{M}^0_{loc}$ and f adapted and continuous. Then*

$$\lim_{n\to\infty} \sum_{i=0}^{\infty} f_{t_i^n}\left(M_{t_{i+1}^n \wedge t} - M_{t_i^n \wedge t}\right) = \int_0^t f_s \, \mathrm{d}M_s \text{ ucp.}$$

Proof. We point out that a continuous adapted process is automatically predictable because it can be pointwise approximated by simple processes.

First assume that $M \in \mathcal{H}^0$ and f is in addition bounded. For a given $\Delta^n \in \mathcal{D}$ we define the approximation f^n by $f^n_t := f_{t_i}$ in case $t \in (t_i, t_{i+1}]$ (and say $f^n_0 := f_0$). Then $f^n \in \mathcal{L}_s$ and $\|f - f^n\|_M \to 0$ and therefore the corresponding stochastic integrals converge, so the result follows in this case. The general case follows by stopping. $\square$

We start by proving the following *integration by parts* formula (or *product rule*). So far we have defined the bracket $\langle M, N\rangle$ only for $M, N \in \mathcal{M}^0_{loc}$ but of course it also makes sense for $M, N \in \mathcal{S}^0$ (and coincides with the bracket of the local martingale parts). Note that $[M, N] = 0$ in case either M or N is in $\mathcal{A}$. If $M, N \in \mathcal{S}$ but not in $\mathcal{S}^0$, then we define both the bracket and the stochastic integral as the bracket and stochastic integral of $M - M_0$ resp. $N - N_0$.

Proposition 2.12. *Let $X, Y \in \mathcal{S}$. Then*

$$X_t Y_t - X_0 Y_0 = \int_0^t X_s \, \mathrm{d}Y_s + \int_0^t Y_s \, \mathrm{d}X_s + \langle X, Y\rangle_t.$$

Proof. For $\Delta \in \mathcal{D}$, we have

$$\begin{aligned} X_t Y_t - X_0 Y_0 = &\sum_{k=0}^{\infty} \Big(X_{t_{k+1}^n \wedge t} - X_{t_k^n \wedge t}\Big)\Big(Y_{t_{k+1}^n \wedge t} - Y_{t_k^n \wedge t}\Big) \\ &+ \sum_{k=0}^{\infty} Y_{t_k^n \wedge t}\Big(X_{t_{k+1}^n \wedge t} - X_{t_k^n \wedge t}\Big) + \sum_{k=0}^{\infty} X_{t_k^n \wedge t}\Big(Y_{t_{k+1}^n \wedge t} - Y_{t_k^n \wedge t}\Big). \end{aligned}$$

Theorem 2.11 shows that the second and last sums converge to the corresponding stochastic integrals and by results in the first chapter the first sum converges ucp to the bracket. $\square$

2.2 Itô's formula

The following Itô's formula is the chain rule of stochastic integration.

Theorem 2.13 (Itô's formula)**.** *Let $F \in C^2(\mathbb{R}^d)$ and $X = (X^1, ..., X^d)$ such that $X^i \in \mathcal{S}$, $i = 1, ..., d$. Then $F(X^1, ..., X^d) \in \mathcal{S}$ and*

$$F(X_t) = F(X_0) + \sum_{k=1}^{d} \int_0^t \frac{\partial F}{\partial x_k}(X_s) \, \mathrm{d}X_s^k + \frac{1}{2} \sum_{k,l=1}^{d} \int_0^t \frac{\partial^2 F}{\partial x_k \partial x_l}(X_s) \, \mathrm{d}\langle X^k, X^l\rangle_s.$$

In the proof of Itô's formula we will need two properties of stochastic integrals which are also of interest otherwise.

Lemma 2.14. *Let $M \in \mathcal{H}^0$, $g \in \mathcal{L}_*(M)$ and $f \in \mathcal{L}_*(g \cdot M)$. Then $fg \in \mathcal{L}_*(M)$ and*

$$(fg) \cdot M = f \cdot (g \cdot M).$$

Proof. Let $\tilde{M} := g \cdot M$. Then $\tilde{M} \in \mathcal{H}^0$ and $\langle \tilde{M} \rangle = g^2 \cdot \langle M \rangle$ by Proposition 2.8, part ii). Hence,

$$\mathbb{E} \int_0^\infty (f_s g_s)^2 \, \mathrm{d}\langle M \rangle_s = \mathbb{E} \int_0^\infty f_s^2 \, \mathrm{d}\langle \tilde{M} \rangle_s < \infty,$$

so $fg \in \mathcal{L}_*(M)$. Further, for $N \in \mathcal{H}^0$, we have (using Proposition 2.8, part ii) once more),

$$\begin{aligned} \langle (fg) \cdot M, N \rangle &= \langle (fg) \cdot M, 1 \cdot N \rangle = (fg) \cdot \langle M, N \rangle = f \cdot (g \cdot \langle M, N \rangle) \\ &= f \cdot \langle g \cdot M, N \rangle = \langle f \cdot (g \cdot M), N \rangle, \end{aligned}$$

which implies

$$\langle (fg) \cdot M - f \cdot (g \cdot M), N \rangle = 0.$$

Since this holds for each $N \in \mathcal{H}^0$ we can choose $N = (fg) \cdot M - f \cdot (g \cdot M)$ which implies

$$\langle (fg) \cdot M - f \cdot (g \cdot M) \rangle_\infty = 0.$$

Now Theorem 1.66, part 4. shows that $((fg) \cdot M)_t - (f \cdot (g \cdot M))_t = 0$ almost surely for all t so the claim follows. □

Remark 2.15. The previous proposition extends to the case of general $M \in \mathcal{M}_{loc}^0$ and even $M \in \mathcal{S}^0$ and integrands f, g for which the stochastic integrals are defined by applying suitable stopping times.

The following continuity property is good to know.

Lemma 2.16. *Let $M \in \mathcal{M}_{loc}^0$ and f, f^n, $n \in \mathbb{N}$ be in $\mathcal{L}_{loc}(M)$ such that $f^n \to f$ ucp. Then, $f^n \cdot M \to f \cdot M$ ucp.*

Proof. Without loss of generality we assume that $f = 0$. Define

$$T_m := \inf\{s \geq 0 : |M_s| \geq m\}; \quad M_s^m := M_{s \wedge T_m}; \quad \tau_{m,n} := \inf\{s \geq 0 : \int_0^s |f_u^n|^2 \, \mathrm{d}\langle M^m \rangle_u \geq \frac{1}{m}\}.$$

Then, for fixed $t, \delta > 0$,

$$\mathbb{P}\Big(\sup_{0 \leq s \leq t} \Big| \int_0^s f_u^n \, \mathrm{d}M_u \Big| \geq \delta \Big) \leq \mathbb{P}\big(\tau_{m,n} \wedge T_m \leq t\big) + \mathbb{P}\Big(\sup_{0 \leq s \leq t} \Big| \int_0^{s \wedge \tau_{m,n} \wedge T_m} f_u^n \, \mathrm{d}M_u \Big| \geq \delta \Big).$$

Using – in this order – Chebychev's inequality, Doob's L^2-inequality and Proposition 2.2 iii), the last summand can be estimated from above by

$$\mathbb{P}\Big(\sup_{0 \leq s \leq t} \Big| \int_0^s f_u^n \, 1\!\!1_{[0,\tau_{m,n}]}(u) \mathrm{d}M_u^m \Big| \geq \delta \Big) \leq \frac{4}{\delta^2} \mathbb{E}\Big(\int_0^t (f_u^n \, 1\!\!1_{[0,\tau_{m,n}]}(u))^2 \, \mathrm{d}\langle M^m \rangle_u \Big) \leq \frac{4}{\delta^2 m}.$$

For given $\varepsilon, \delta > 0$, we first choose m such that $\mathbb{P}(T_m \leq t) \leq \varepsilon/3$ and $\frac{4}{\delta^2 m} \leq \varepsilon/3$. Then choose n_0 such that for all $n \geq n_0$, $\mathbb{P}(\tau_{m,n} \leq t) \leq \varepsilon/3$. Then we obtain

$$\mathbb{P}\Big(\sup_{0 \leq s \leq t} \Big| \int_0^s f_u^n \, \mathrm{d}M_u \Big| \geq \delta \Big) \leq \varepsilon.$$

The result follows since $\delta, \varepsilon, t > 0$ were arbitrary. □

Proof of Theorem 2.13. Obviously, Itô's formula holds for constant F as well as for $F(x) = x_k$ for some $k \in \{1, ..., d\}$. Let us next consider the case in which $F(x) = x_k x_l$ for some $k, l \in \{1, ..., d\}$. Then it is straightforward to check that Itô's formula for F is the same as the product rule which we established in Proposition 2.12. Further, if Itô's formula holds for F and G (both in $C^2(\mathbb{R}^d)$), then it clearly holds for each linear combination of F and G, so we already know that the formula is true for all polynomials of order at most 2. More generally, the following is true:

suppose that the formula holds true for $F \in C^2(\mathbb{R}^d)$ and let $G(x) := x_i F(x)$ for some $i \in \{1, ..., d\}$. Applying the integration by parts formula to G shows (after some computation, see e.g. [Re13] or [St13] using Lemma 2.14) that the formula also holds for G. By induction, we see that the formula holds for all polynomials. To finish the proof, observe that any $F \in C^2(\mathbb{R}^d)$ can be approximated by polynomials P_n in such a way that for given $k \in \mathbb{N}$, P_n and all its first and second partial derivatives converge to F and its corresponding derivatives uniformly on $[-k, k]^d$ as $n \to \infty$ (this is Weierstraß' approximation theorem). Using Lemma 2.16 it follows that Itô's fomula hold for $F \in C^2(\mathbb{R}^d)$ for all X which almost surely never leave $[-k, k]^d$. For general X we stop the process X when it threatens to leave $[-k, k]^d$ (i.e. when it hits the boundary of the cube), then for given $F \in C^2(\mathbb{R}^d)$ the formula holds for the stopped process. Letting $k \to \infty$, the assertion follows. □

Remark 2.17. Itô's formula can be generalized in various directions. If, for example F is only defined and C^2 on an open subset G of $\mathbb{R}^d$ and X remains in G almost surely for all t, then the formula remains true as stated (we will need this extension in an example below). If F also depends on t, then we can regard t as an additional semimartingale and Itô's formula holds provided that F is C^2 in all variables including t. One can show however that it is sufficient that F is only C^1 with respect to t (observe that all brackets of t and any X^i vanish). One can ask for the most general condition on F that will guarantee that $F(X)$ is a continuous semimartingale. In case $d = 1$ and X Brownian motion it is known that $F(X) \in \mathcal{S}$ iff F is the difference of two convex functions (which is true for any C^2 function but not for every C^1 function), i.e. (roughly speaking) the generalized first derivative of F is of locally finite variation. In this case Itô's formula has to be modified appropriately since second derivatives do not necessarily exist. This requires the concept of *local time* which we will not develop in this course, see e.g. [KS91], p. 214.

Remark 2.18. If B and $\tilde{B}$ are independent $\mathbb{F}$-Brownian motions, then $\langle B, \tilde{B} \rangle = 0$. To see this, we show that $B\tilde{B}$ is a martingale. Let $0 \le s < t < \infty$. Then

$$\begin{aligned}\mathbb{E}(B_t\tilde{B}_t - B_s\tilde{B}_s | \mathcal{F}_s) &= \mathbb{E}((B_t - B_s)(\tilde{B}_t - \tilde{B}_s) + B_s(\tilde{B}_t - \tilde{B}_s) | \mathcal{F}_s) \\ &= \mathbb{E}\big((B_t - B_s)(\tilde{B}_t - \tilde{B}_s)\big) = 0.\end{aligned}$$

Our first application of Itô's formula is a version of the famous Burkholder-Davis-Gundy inequality.

Theorem 2.19. *For $p \ge 2$ there exists a constant C_p such that for any $M \in \mathcal{M}^0_{loc}$ and any $t \ge 0$ we have*

$$\mathbb{E} \sup_{0 \le s \le t} |M_s|^p \le C_p \mathbb{E} \langle M \rangle_t^{p/2}.$$

Proof. It suffices to show the statement for bounded M. Since the function $x \mapsto |x|^p$ is in $C^2(\mathbb{R})$ we can apply Itô's formula and get

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

Consequently,

$$\begin{aligned}\mathbb{E}|M_t|^p &= \frac{1}{2}p(p-1)\mathbb{E}\int_0^t |M_s|^{p-2}\,\mathrm{d}\langle M\rangle_s\\ &\le \frac{1}{2}p(p-1)\mathbb{E}\big(\sup_{0\le s\le t}|M_s|^{p-2}\langle M\rangle_t\big)\\ &\le \frac{1}{2}p(p-1)\Big(\mathbb{E}\sup_{0\le s\le t}|M_s|^p\Big)^{(p-2)/p}\Big(\mathbb{E}\langle M\rangle_t^{p/2}\Big)^{2/p},\end{aligned}$$

where we used Hölder's inequality in the last step. Using Doob's L^p-inequality and rearranging terms implies the result. □

Remark 2.20. The inequality also holds for general $p > 0$ and in the reverse direction for appropriate constants but the proofs are more involved (see e.g. [RY94]).

Example 2.21. Assume that $B = (B^1, ..., B^d)$ is d-dimensional Brownian motion, i.e. all components are independent $\mathbb{F}$-Brownian motions. Let $F \in C^2(\mathbb{R}^d, \mathbb{R})$. Then – using Remark 2.18 – Itô's formula implies

$$\begin{aligned}F(B_t) &= F(B_0) + \sum_{i=1}^d \int_0^t \frac{\partial F}{\partial x_i}(B_s)\,\mathrm{d}B_s^i + \frac{1}{2}\sum_{i=1}^d \int_0^t \frac{\partial^2 F}{\partial x_i^2}(B_s)\,\mathrm{d}s\\ &= F(B_0) + \sum_{i=1}^d \int_0^t \frac{\partial F}{\partial x_i}(B_s)\,\mathrm{d}B_s^i + \frac{1}{2}\int_0^t \Delta F(B_s)\,\mathrm{d}s.\end{aligned}$$

Note that when $\Delta F = 0$ (in which case F is called *harmonic*), then $t \mapsto F(B_t)$ is a local martingale. Let us now look at a particular case. Fix $z \in \mathbb{R}^d\backslash\{0\}$, $d \ge 2$ and define $F(x) := \frac{1}{|z+x|}$ for $x \ne -z$ ($|.|$ denoting the Euclidean norm). It is straightforward to check that $\Delta F = 0$ on $\mathbb{R}^d\backslash\{-z\}$. Since Brownian motion B is known not to hit $\{z\}$ almost surely for any fixed $z \ne 0$ in dimension $d \ge 2$, we see that $M_t := F(B_t)$ is a continuous local martingale in case $d \ge 2$. In case $d \ge 3$, M is however not a martingale (which provides us with the first concrete example of a local martingale which is not a martingale). This follows from the fact that $\mathbb{E}M_t \to 0$ (which is easily checked by explicit computation) and the fact that a martingale has a constant expected value.

2.3 Representation of local martingales

In this section we collect a few useful results without proof. The proofs can be found in any textbook on stochastic analysis, e.g. [KS91]. We start with Lévy's characterization of Brownian motion.

Theorem 2.22. *If $M \in \mathcal{M}_{loc}^0$ is such that $\langle M\rangle_t = t$ for all $t \ge 0$, then M is an $\mathbb{F}$-Brownian motion.*

The following result is due to Dambis, Dubins and Schwarz and can also be found in any textbook on stochastic analysis, e.g. [KS91], p. 174.

Theorem 2.23. *Let $M \in \mathcal{M}_{loc}^0$ satisfy $\langle M \rangle_\infty = \infty$ almost surely. Define for each $s \in [0, \infty)$ the stopping time*

$$T(s) := \inf\{t \geq 0 : \langle M \rangle_t > s\}.$$

Then the process $B_s := M_{T(s)}$ is a standard Brownian motion and $M_t = B_{\langle M \rangle_t}$.

Idea of the proof. The idea is to check that the process B satisfies the assumption of Theorem 2.22 with respect to the filtration $\mathcal{G}_s := \mathcal{F}_{T(s)}$. The local martingale property of B can be proved via optional sampling and the fact that $\langle B \rangle_t = t$ is also easy to see. It then remains to show that B has almost surely continuous paths in order to be able to apply Theorem 2.22. □

Remark 2.24. If $M \in \mathcal{M}_{loc}^0$ does not satisfy $\langle M \rangle_\infty = \infty$ almost surely then a representation for M as in Theorem 2.23 may not be possible on the given probability space. If, for example, Ω consists of a single point and $M = 0$, then no Brownian motion can be defined on Ω. If $M \in \mathcal{M}_{loc}^0$ does not satisfy $\langle M \rangle_\infty = \infty$ almost surely, then one can however find a different FPS on which a continuous local martingale with the same law as M and a Brownian motion B can be defined such that $M_t = B_{\langle M \rangle_t}$ holds for all $t \geq 0$.

2.4 Stochastic differential equations

In this section we consider stochastic differential equations of the following kind

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sum_{k=1}^m \sigma_k(X_t)\,\mathrm{d}W_t^k, \; X_0 = x, \tag{2.4.1}$$

where W^k, $k = 1, ..., m$ are independent (given) $\mathbb{F}$-Brownian motions on a FPS $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ satisfying the usual conditions and $x \in \mathbb{R}^d$. We will specify assumptions on the coefficients b and σ_k (which map $\mathbb{R}^d$ to $\mathbb{R}^d$) below. We will denote the Euclidean norm on $\mathbb{R}^d$ by $|.|$ and the standard inner product on $\mathbb{R}^d$ by $\langle ., . \rangle$ (hoping that the reader will not confuse this with the bracket of semimartingales).

Definition 2.25. An $\mathbb{R}^d$-valued stochastic process X is called a *solution* of (2.4.1) if X is adapted with continuous paths and satisfies the corresponding integral equation

$$X_t = x + \int_0^t b(X_s)\,\mathrm{d}s + \sum_{k=1}^m \int_0^t \sigma_k(X_s)\,\mathrm{d}W_s^k,$$

which means in particular that the corresponding integrals are defined. We say that the stochastic differential equation (sde) has a unique solution if any two solutions (with the same initial value x) are indistinguishable.

In order to have a chance that an sde has a unique strong solution we need both local regularity properties as well as growth conditions for the coefficients. For given functions $\sigma_k : \mathbb{R}^d \to \mathbb{R}^d$, $k = 1, ..., m$ we define $\mathcal{A} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^{d \times d}$ by

$$\mathcal{A}_{i,j}(x, y) := \sum_{k=1}^m \Big(\big(\sigma_k^i(x) - \sigma_k^i(y)\big)\big(\sigma_k^j(x) - \sigma_k^j(y)\big) \Big)$$

Note that $\mathcal{A}(x,y)$ is a nonnegative definite $d \times d$-matrix with trace

$$\operatorname{tr} \mathcal{A}(x,y) = \sum_{i=1}^{d} \sum_{k=1}^{m} \big(\sigma_k^i(x) - \sigma_k^i(y)\big)^2 = \sum_{k=1}^{m} |\sigma_k(x) - \sigma_k(y)|^2.$$

Further, we define

$$a_{ij}(x,y) := \sum_{k=1}^{m} \sigma_k^i(x)\sigma_k^j(y), \quad i,j \in \{1,...,d\}.$$

The matrix $a(x,y)$ is generally *not* nonnegative definite, but $a(x,x)$ is and $\operatorname{tr} a(x,x) = \sum_{k=1}^m |\sigma_k(x)|^2$.

In our main result below we assume the following:

Assumption (H)

i) b and σ_k, $k = 1,...,m$ are continuous.

ii) For each $R > 0$ there exists K_R such that $2\langle b(x) - b(y), x - y\rangle + \operatorname{tr} \mathcal{A}(x,y) \le K_R|x-y|^2$ for all $|x|, |y| \le R$.

iii) There exists $\bar{K}$ such that $2\langle b(x), x\rangle + \operatorname{tr} a(x,x) \le \bar{K}\big(|x|^2 + 1\big)$ for all $x \in \mathbb{R}^d$.

In recent years it has become fashionable to replace the classical Lipschitz hypotheses by the above *monotonicity* or *one-sided Lipschitz* conditions (H). These assumptions are weaker than the usual Lipschitz assumptions:

Proposition 2.26. *Assume that b and σ_k, $k = 1,...,m$ are Lipschitz with constant L. Then (H) holds.*

Proof. Clearly, part i) of (H) is satisfied. Further, using the Cauchy-Schwarz inequality,

$$2\langle b(x) - b(y), x - y\rangle + \operatorname{tr} \mathcal{A}(x,y) \le 2L|x-y|^2 + L^2 m|x-y|^2,$$

so ii) holds. Finally

$$\begin{aligned} 2\langle b(x), x\rangle + \operatorname{tr} a(x,x) &= 2\langle b(x) - b(0), x - 0\rangle + 2\langle b(0), x\rangle + \sum_{k=1}^{m} |\sigma_k(x) - \sigma_k(0) + \sigma_k(0)|^2 \\ &\le 2L|x|^2 + 2|b(0)||x| + 2mL^2|x|^2 + 2\sum_{k=1}^{m} |\sigma_k(0)|^2. \end{aligned}$$

Estimating $2|b(0)||x| \le |b(0)|^2 + |x|^2$ we see that iii) follows. □

To see that (H) is strictly weaker than Lipschitz continuity consider the following example: for $d = m = 1$ take $b(x) := -x^3$ and $\sigma_1 = 0$. b is certainly not Lipschitz continuous but clearly satisfies (H) (in fact with $\bar{K} = K_R = 0$). An example for which b is not even locally Lipschitz continuous but still satisfies (H) is $b(x) = -x^{1/2} \cdot \mathbb{1}_{x \ge 0}$.

Theorem 2.27. *Assume (H). Then the sde* (2.4.1) *has a unique solution for each $x \in \mathbb{R}^d$.*

Before proving the theorem, we formulate both the classical Gronwall inequality as well as a (not so classical) stochastic Gronwall inequality.

Proposition 2.28 (Gronwall inequality)**.** *Let $L \geq 0$ and let $g : [0,T] \to [0,\infty)$ be continuous and $h : [0,T] \to \mathbb{R}$ be integrable with respect to Lebesgue measure on $[0,T]$ and satisfy*

$$g(t) \leq h(t) + L \int_0^t g(s)\,\mathrm{d}s \tag{2.4.2}$$

for all $0 \leq t \leq T$. Then

$$g(t) \leq h(t) + L \int_0^t h(s)\mathrm{e}^{L(t-s)}\,\mathrm{d}s$$

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

Proof.

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(\mathrm{e}^{-Lt}\int_0^t g(s)\,\mathrm{d}s\Big) = \mathrm{e}^{-Lt}\Big(g(t) - L\int_0^t g(s)\,\mathrm{d}s\Big) \leq \mathrm{e}^{-Lt}h(t),$$

so

$$\int_0^t g(s)\,\mathrm{d}s \leq \mathrm{e}^{Lt}\int_0^t h(s)\mathrm{e}^{-Ls}\,\mathrm{d}s$$

and therefore – using (2.4.2) –

$$g(t) \leq h(t) + L\int_0^t g(s)\,\mathrm{d}s \leq h(t) + L\int_0^t h(s)\mathrm{e}^{L(t-s)}\,\mathrm{d}s.$$

□

For the proof of the following stochastic Gronwall lemma we require the following martingale inequality (which is due to Burkholder).

Proposition 2.29. *For each $p \in (0,1)$ and $M \in \mathcal{M}^0_{loc}$, we have*

$$\mathbb{E}\big(\sup_{t\geq 0} M^p(t)\big) \leq c_p \mathbb{E}\big((-\inf_{t\geq 0} M(t))^p\big), \tag{2.4.3}$$

where $c_p := \Big(4 \wedge \frac{1}{p}\Big)\frac{\pi p}{\sin(\pi p)}$.

Remark 2.30. It is clear that the previous proposition does not extend to $p \geq 1$: consider the continuous martingale $M(t) := W(\tau_{-1} \wedge t)$ where W is standard Brownian motion and $\tau_x := \inf\{s \geq 0 : W(s) = x\}$. Then the left hand side of (2.4.3) is infinite for each $p \geq 1$ while the right hand side is finite. This example also shows that even though the constant c_p is certainly not optimal, it is at most off from the optimal constant by the factor $4\wedge(1/p)$ (which converges to one as p approaches one). It is also clear that the proposition does not extend to right-continuous martingales: consider a martingale which is constant except for a single jump at time 1 of height 1 with probability δ and height $-\frac{\delta}{1-\delta}$ with probability $1-\delta$ where $\delta \in (0,1)$. It is straightforward to check that for an inequality of type $(\mathbb{E}\sup_{t\geq 0} M^p(t))^{1/p} \leq c_{p,q}(\mathbb{E}(-\inf_{t\geq 0} M(t))^q)^{1/q}$ to hold for this class of examples for some finite $c_{p,q}$, we require that $q \geq 1$ irrespective of the value of $p \in (0,1)$.

Proof of Proposition 2.29. Since M is a continuous local martingale starting at 0 it can be represented as a time-changed Brownian motion W (on a suitable probability space, by Theorem 2.23 and Remark 2.24). We can and will assume that M converges almost surely (otherwise there is nothing to prove), so there exists an almost surely finite stopping time T for W such that $A := \sup_{0\le t\le T} W(t) = \sup_{0\le t} M(t)$ and $B := -\inf_{0\le t\le T} W(t) = -\inf_{0\le t} M(t)$. Let $0 = a_0 < a_1 < \ldots$ be a sequence which converges to ∞ and define

$$\tau_i := \inf\{t \ge 0 : W(t) = -a_i\},\ Y_i := \sup_{\tau_{i-1}\le t\le\tau_i} W(t),\ i \in \mathbb{N}, \quad N := \inf\{i \in \mathbb{N} : \tau_i \ge T\}.$$

The Y_i are independent by the strong Markov property of W and for $p \in (0,1)$ and $i \in \mathbb{N}$ we have

$$\Gamma_i := \mathbb{E}(Y_i \vee 0)^p = \frac{a_i - a_{i-1}}{a_i^{1-p}} \int_0^\infty \frac{1}{1+y^{1/p}}\,\mathrm{d}y = \frac{a_i - a_{i-1}}{a_i^{1-p}} \frac{\pi p}{\sin(\pi p)}.$$

Therefore,

$$\begin{aligned}
\mathbb{E}A^p &\le \sum_{n=1}^\infty \mathbb{E}\Big(\sup\{Y_1,\ldots,Y_n\}^p \mathbb{1}_{N=n}\Big) \le \sum_{n=1}^\infty \sum_{i=1}^n \mathbb{E}\Big((Y_i \vee 0)^p \mathbb{1}_{N=n}\Big)\\
&= \sum_{i=1}^\infty \mathbb{E}\Big((Y_i \vee 0)^p \mathbb{1}_{N\ge i}\Big) = \sum_{i=1}^\infty \Gamma_i \mathbb{P}\{N \ge i\},
\end{aligned}$$

where the last equality again follows from the strong Markov property. Inserting the formula for Γ_i, choosing the particular values $a_i = c\gamma^i$ for some $c > 0$ and $\gamma > 1$, and observing that $\mathbb{P}\{N \ge i\} \le \mathbb{P}\{B \ge a_{i-1}\}$, we get

$$\begin{aligned}
\mathbb{E}A^p &\le \frac{\pi p}{\sin(\pi p)} c^p \Big(\gamma^p + \Big(1 - \frac{1}{\gamma}\Big) \sum_{i=2}^\infty \gamma^{ip} \mathbb{P}\{B \ge c\gamma^{i-1}\}\Big)\\
&= \frac{\pi p}{\sin(\pi p)} c^p \Big(\gamma^p + \Big(1 - \frac{1}{\gamma}\Big) \sum_{j=2}^\infty \mathbb{P}\{B \in [c\gamma^{j-1}, c\gamma^j)\} \Big(\frac{\gamma^{p(j+1)} - 1}{\gamma^p - 1} - 1 - \gamma^p\Big)\Big)\\
&\le \frac{\pi p}{\sin(\pi p)} \Big(c^p\gamma^p + \Big(1 - \frac{1}{\gamma}\Big) \frac{\gamma^{2p}}{\gamma^p - 1} \mathbb{E}B^p - c^p \Big(1 - \frac{1}{\gamma}\Big)\Big(\frac{1}{\gamma^p - 1} + 1 + \gamma^p\Big) \mathbb{P}\{B \ge c\gamma\}\Big).
\end{aligned}$$

Dropping the last (negative) term, letting $c \to 0$ and observing that the function of γ in front of $\mathbb{E}B^p$ converges to $1/p$ as $\gamma \to 1$ and that $\inf_{\gamma>1} \gamma^{2p}/(\gamma^p - 1) = 4$ we obtain the assertion. □

Next, we apply the martingale inequality to prove a stochastic Gronwall lemma. For a real-valued process denote $Y_t^* := \sup_{0\le s\le t} Y_s$.

Proposition 2.31. *Let c_p be as in Proposition 2.29. Let Z and H be nonnegative, adapted processes with continuous paths. Let $M \in \mathcal{M}_{loc}^0$ and $L \ge 0$. If*

$$Z_t \le L \int_0^t Z_s\,\mathrm{d}s + M_t + H_t \tag{2.4.4}$$

holds for all $t \ge 0$, then for $p \in (0,1)$, we have

$$\mathbb{E} \sup_{0\le s\le t} Z_s^p \le (c_p + 1) \exp\{pLt\} \big(\mathbb{E}(H_t^*)^p\big), \tag{2.4.5}$$

and

$$\mathbb{E}Z_t \le \exp\{Lt\}\mathbb{E}H_t^*. \tag{2.4.6}$$

Proof. Let $N_t := \int_0^t \exp\{-Ls\}\,\mathrm{d}M_s$. Applying the usual Gronwall inequality (Proposition 2.28) for each fixed $\omega \in \Omega$ to Z and integrating by parts, we obtain

$$Z_t \le \exp\{Lt\}(N_t + H_t^*). \tag{2.4.7}$$

Since Z is nonnegative, we have $-N_t \le H_t^*$ for all $t \ge 0$. Therefore, using Proposition 2.29 and the inequality $(a+b)^p \le a^p + b^p$ for $a, b \ge 0$ and $p \in [0,1]$, we get

$$\begin{aligned}\mathbb{E}(Z_t^*)^p &\le \exp\{pLt\}\big(\mathbb{E}(N_t^*)^p + \mathbb{E}(H_t^*)^p\big)\\ &\le \exp\{pLt\}(c_p+1)\big(\mathbb{E}(H_t^*)^p\big),\end{aligned}$$

which is (2.4.5). The final statement follows by applying (2.4.7) to $\tau_n \wedge t$ for a sequence of localizing stopping times τ_n for N and applying Fatou's Lemma. □

Proof of Theorem 2.27. The basic idea of the proof is taken from [PR07] which in turn is based on a proof by Krylov. Increasing the number K_R if necessary, we can and will assume that $\sup_{|x|\le R} |b(x)| \le K_R$ for each $R > 0$. To prove existence of a solution for fixed $x \in \mathbb{R}^d$, we employ an Euler scheme. For $n \in \mathbb{N}$, we define the process $(\phi_t^n)_{t\in[0,\infty)}$ by $\phi_0^n := x \in \mathbb{R}^d$ and – for $l \in \mathbb{N}_0$ and $t \in (\frac{l}{n}, \frac{l+1}{n}]$ – by

$$\phi_t^n := \phi_{\frac{l}{n}}^n + \int_{\frac{l}{n}}^t b(\phi_{\frac{l}{n}}^n)\,\mathrm{d}s + \sum_{k=1}^m \int_{\frac{l}{n}}^t \sigma_k(\phi_{\frac{l}{n}}^n)\,\mathrm{d}W_s^k, \tag{2.4.8}$$

which is equivalent to

$$\phi_t^n = x + \int_0^t b(\bar\phi_s^{(n)})\,\mathrm{d}s + \sum_{k=1}^m \int_0^t \sigma_k(\bar\phi_s^n)\,\mathrm{d}W_s^k, \tag{2.4.9}$$

for $t \in [0,\infty)$, where $\bar\phi_s^n := \phi^n_{\frac{\lfloor ns\rfloor}{n}}$. Defining $p_t^n := \bar\phi_t^n - \phi_t^n$, we obtain

$$\phi_t^n = x + \int_0^t b(\phi_s^n + p_s^n)\,\mathrm{d}s + \sum_{k=1}^m \int_0^t \sigma_k(\phi_s^n + p_s^n)\,\mathrm{d}W_s^k \tag{2.4.10}$$

for $t \in [0,\infty)$. Observe that $t \mapsto \phi_t^n$ is adapted and continuous. Using Itô's formula, we obtain for $t \ge 0$

$$\begin{aligned}\left|\phi_t^n - \phi_t^{n'}\right|^2 &= \int_0^t 2\left\langle \phi_s^n - \phi_s^{n'}, b(\bar\phi_s^n) - b(\bar\phi_s^{n'})\right\rangle \mathrm{d}s\\ &\quad + \sum_{k=1}^m \int_0^t 2\left\langle \phi_s^n - \phi_s^{n'}, \sigma_k(\bar\phi_s^n) - \sigma_k(\bar\phi_s^{n'})\right\rangle \mathrm{d}W_s^k + \int_0^t \mathrm{tr}(\mathcal{A}(\bar\phi_s^n, \bar\phi_s^{n'}))\,\mathrm{d}s\\ &= \int_0^t 2\left\langle \bar\phi_s^n - \bar\phi_s^{n'}, b(\bar\phi_s^n) - b(\bar\phi_s^{n'})\right\rangle \mathrm{d}s + \int_0^t \mathrm{tr}(\mathcal{A}(\bar\phi_s^n, \bar\phi_s^{n'}))\,\mathrm{d}s\\ &\quad + M_t^{n,n'} + \int_0^t 2\left\langle p_s^{n'} - p_s^n, b(\bar\phi_s^n) - b(\bar\phi_s^{n'})\right\rangle \mathrm{d}s,\end{aligned}$$

where

$$M_t^{n,n'} := 2\sum_{k=1}^{m}\int_0^t \langle \phi_s^n - \phi_s^{n'}, \sigma_k(\bar\phi_s^n) - \sigma_k(\bar\phi_s^{n'})\rangle \,\mathrm{d}W_s^k$$

is in $\mathcal{M}_{loc}^0$. Let $R \in \mathbb{N}$ be such that $R > 3|x|$ and define the following stopping times

$$\tau_n(R) := \inf\left\{t \geq 0 : |\phi_t^n| \geq \frac{R}{3}\right\}.$$

Then

$$|p_t^n| \leq \frac{2R}{3} \quad \text{and} \quad |\phi_t^n| \leq \frac{R}{3} \quad \text{for } t \in [0, \tau_n(R)] \cap [0,\infty). \tag{2.4.11}$$

For $0 \leq s \leq \tau_n(R) \wedge \tau_{n'}(R) =: \gamma^{n,n'}(R)$, we have (using $|b(y)| \leq K_R$ for $|y| \leq R$)

$$\langle p_s^{n'} - p_s^n, b(\bar\phi_s^n) - b(\bar\phi_s^{n'})\rangle \leq 2K_R|p_s^{n'} - p_s^n| \leq 2K_R\big(|p_s^{n'}| + |p_s^n|\big).$$

Therefore, for $t \leq \gamma^{n,n'}(R)$, we get

$$\begin{aligned}
&\left|\phi_t^n - \phi_t^{n'}\right|^2 \\
&\leq \int_0^t \left(K_R\left|\bar\phi_s^n - \bar\phi_s^{n'}\right|^2 + 2\left\langle p_s^{n'} - p_s^n, b(\bar\phi_s^n) - b(\bar\phi_s^{n'})\right\rangle\right)\mathrm{d}s + M_t^{n,n'} \\
&\leq \int_0^t 2K_R\left(\left|\phi_s^n - \phi_s^{n'}\right|^2 + \left|p_s^n - p_s^{n'}\right|^2\right) + 4K_R\left(\left|p_s^n\right| + \left|p_s^{n'}\right|\right)\mathrm{d}s + M_t^{n,n'} \\
&\leq \int_0^t 2K_R\left|\phi_s^n - \phi_s^{n'}\right|^2 \mathrm{d}s + 4\int_0^t K_R\left(\left|p_s^n\right| + \left|p_s^{n'}\right| + \left|p_s^n\right|^2 + \left|p_s^{n'}\right|^2\right)\mathrm{d}s + M_t^{n,n'}.
\end{aligned}$$

Now, we apply Proposition 2.31 to the process $Z_t := \left|\phi^n_{t\wedge\gamma^{n,n'}(R)} - \phi^{n'}_{t\wedge\gamma^{n,n'}(R)}\right|^2$. Note that the assumptions are satisfied with $L := 2K_R$, $M_t := M^{n,n'}_{t\wedge\gamma^{n,n'}(R)}$, and

$$H_t := H_t^{n,n'} := 4\int_0^{t\wedge\gamma^{n,n'}(R)} K_R(|p_s^n| + |p_s^{n'}| + |p_s^n|^2 + |p_s^{n'}|^2)\,\mathrm{d}s.$$

Therefore, for each $p \in (0,1)$ and c_p as in Proposition 2.31 we obtain for $T > 0$

$$\begin{aligned}
\mathbb{E}\left[\sup_{t\in[0,T]}\left|\phi^n_{t\wedge\gamma^{n,n'}(R)} - \phi^{n'}_{t\wedge\gamma^{n,n'}(R)}\right|^{2p}\right] &\leq (c_p+1)\mathrm{e}^{2pK_RT}\,\mathbb{E}\left[\sup_{t\in[0,T]} H^{n,n'}(t)^p\right] \\
&= (c_p+1)\mathrm{e}^{2pK_RT}\,\mathbb{E}\left[H^{n,n'}(T)^p\right].
\end{aligned} \tag{2.4.12}$$

Now we show the following:

1. For fixed R, $\sup_{s\geq 0}\mathbb{E}\big(|p^n_{s\wedge\tau_n(R)}| + |p^n_{s\wedge\tau_n(R)}|^2\big)$ converges to 0 as $n \to \infty$. Hence

$$\lim_{n,n'\to\infty}\mathbb{E}\left[\sup_{t\in[0,T]}\left|\phi^n_{t\wedge\gamma^{n,n'}(R)} - \phi^{n'}_{t\wedge\gamma^{n,n'}(R)}\right|^{2p}\right] = 0.$$

2. For every $T>0$, we have $\lim_{R\to\infty}\limsup_{n\to\infty}\mathbb{P}\big(\tau_n(R)\le T\big)=0$.
3. The processes ϕ^n converge ucp to a process ϕ.
4. ϕ solves the sde.
5. Any two solutions ϕ and $\tilde\phi$ of the sde are indistinguishable.

Step 1: For $l/n \le s < (l+1)/n$, we have

$$-p_s^n = -\bar\phi_s^n + \phi_s^n = b\big(\phi_{l/n}^n\big)\big(s-\frac{l}{n}\big) + \sum_{k=1}^m \Big(\sigma_k\big(\phi_{l/n}^n\big)\big(W_s^k - W_{l/n}^k\big)\Big).$$

Define $L_R := \max_k \sup_{|y|\le R}|\sigma_k(y)|$. Then, for suitable independent standard normally distributed random variables $N_1,...,N_m$, we have

$$|p_s^n| \le \frac{K_R}{n} + L_R\sum_{k=1}^m |N_k|/\sqrt{n},$$

in case $\frac{l}{n}\le\tau_n(R)$, so the first (and hence the second) part of the claim in Step 1 follows.

Step 2: Applying Itô's formula and abbreviating

$$M_t^n := 2\sum_{k=1}^m\int_0^t \langle\phi_s^n,\sigma_k^n(\bar\phi_s^n)\rangle\,\mathrm{d}W_s^k,$$

we obtain (as above) for $t\le\tau_n(R)$, using part iii) of Assumption (H),

$$\begin{aligned}
|\phi_t^n|^2 &= |x|^2 + 2\int_0^t\langle\phi_s^n, b(\bar\phi_s^n)\rangle\,\mathrm{d}s + \int_0^t \operatorname{tr}(a(\bar\phi_s^n,\bar\phi_s^n))\,\mathrm{d}s + M_t^n\\
&= |x|^2 + 2\int_0^t\langle\bar\phi_s^n, b(\bar\phi_s^n)\rangle\,\mathrm{d}s + \int_0^t \operatorname{tr}(a(\bar\phi_s^n,\bar\phi_s^n))\,\mathrm{d}s + M_t^n - 2\int_0^t\langle p_s^n, b(\bar\phi_s^n)\rangle\,\mathrm{d}s\\
&\le |x|^2 + \bar K\int_0^t\big(|\bar\phi_s^n|^2+1\big)\,\mathrm{d}s + M_t^n + 2K_R\int_0^t|p_s^n|\,\mathrm{d}s\\
&\le |x|^2 + 2\bar K\int_0^t|\phi_s^n|^2\,\mathrm{d}s + M_t^n + 2K_R\int_0^t|p_s^n|\,\mathrm{d}s + 2\bar K\int_0^t|p_s^n|^2\,\mathrm{d}s + \bar K t.
\end{aligned}$$

Applying the stochastic Gronwall Lemma 2.31 again we obtain, for $p\in(0,1)$,

$$\mathbb{E}\sup_{0\le t\le T\wedge\tau_n(R)}|\phi_t^n|^{2p} \le (c_p+1)\mathrm{e}^{2\bar K pT}\mathbb{E}\Big(|x|^2+KT+2\int_0^T \mathbb{1}_{[0,\tau_n(R)]}(s)\big(K_R|p_s^n|+\bar K|p_s^n|^2\big)\,\mathrm{d}s\Big)^p. \tag{2.4.13}$$

The left hand side of (2.4.13) can be estimated from below by

$$\mathbb{E}\sup_{0\le t\le T\wedge\tau_n(R)}|\phi_t^n|^{2p} \ge \Big(\frac{R}{3}\Big)^{2p}\mathbb{P}\big(\tau_n(R)\le T\big).$$

Using the result in Step 1, we therefore get

$$\limsup_{n\to\infty}\mathbb{P}\big(\tau_n(R)\le T\big) \le \Big(\frac{3}{R}\Big)^{2p}(c_p+1)\mathrm{e}^{2\bar K pT}\mathbb{E}\big(|x|^2+KT\big)^p$$

and the claim in Step 2 follows by letting R tend to infinity.

Step 3: Steps 1 and 2 taken together show that the sequence ϕ_t^n is Cauchy with respect to ucp convergence, i.e. for each $\varepsilon, T > 0$

$$\lim_{n,n'\to\infty} \mathbb{P}\big(\sup_{0\le t\le T} |\phi_t^n - \phi_t^{n'}| \ge \varepsilon\big) = 0.$$

Since uniform convergence in probability on $[0,T]$ is metrizable with a complete metric, there exists some adapted and continuous process ϕ such that ϕ is the ucp limit of ϕ^n.

Step 4: Recall that ϕ^n satisfies the equation

$$\phi_t^n = x + \int_0^t b(\bar{\phi}_s^n)\,\mathrm{d}s + \sum_{k=1}^m \int_0^t \sigma_k(\bar{\phi}_s^n)\,\mathrm{d}W_s^k.$$

The left hand side of the equation converges to ϕ_t ucp by Step 3 and the integral over $b(\bar{\phi}_s^n)$ converges ucp to the integral of $b(\phi_s)$ since b is continuous and since $\bar{\phi}_s^n$ converges to ϕ_s ucp. It remains to show that the stochastic integrals also converge ucp to the correct limit, but

$$\int_0^t \sigma_k(\bar{\phi}_s^n) - \sigma_k(\phi_s)\,\mathrm{d}W_s^k \to 0 \quad \text{ucp}$$

by Lemma 2.16, so ϕ indeed solves our sde.

Step 5: Assume that ϕ and ψ both solve the sde with initial condition x. Then, by Itô's formula,

$$|\phi_t - \psi_t|^2 = \int_0^t 2\langle \phi_s - \psi_s, b(\phi_s) - b(\psi_s)\rangle + \mathrm{tr}(\mathcal{A}(\phi_s,\psi_s))\,\mathrm{d}s + M_t \le K_R \int_0^t |\phi_s - \psi_s|^2\,\mathrm{d}s + M_t,$$

for $t \le \tilde{\tau}(R) := \inf\{s \ge 0 : |\phi_s| \vee |\psi_s| \ge R\}$ for some $M \in \mathcal{M}_{loc}^0$. Applying the stochastic Gronwall lemma shows that ϕ and ψ agree up to time $\tilde{\tau}(R)$. Since $\lim_{R\to\infty} \tilde{\tau}(R) = \infty$ almost surely, uniqueness follows and the proof of the theorem is complete. $\square$

Only few stochastic differential equations can be solved in closed form. One class of such examples are linear (or affine) equations. We only consider the case $d = 1$ (more on this can be found in Section 5.6 of [KS91]).

Theorem 2.32. *The sde*

$$\mathrm{d}X_t = (aX_t + b)\,\mathrm{d}t + \sum_{k=1}^m (c_k X_t + d_k)\,\mathrm{d}W_t^k; \quad X_0 = x \in \mathbb{R}$$

has the unique solution

$$X_t = Z_t\Big(x + \big(b - \sum_{k=1}^m c_k d_k\big)\int_0^t \frac{1}{Z_u}\,\mathrm{d}u + \sum_{k=1}^m d_k \int_0^t \frac{1}{Z_u}\,\mathrm{d}W_u^k\Big),$$

where

$$Z_t := \exp\Big\{\Big(a - \frac{1}{2}\sum_{k=1}^m c_k^2\Big)t + \sum_{k=1}^m c_k\,W_t^k\Big\}.$$

Proof. Uniqueness follows from Theorem 2.27. The fact that the solution formula is correct is easy to verify using Itô's formula. □

Remark 2.33. We emphasize two interesting special cases in the previous theorem. One is the case $m = 1$ and $b = d_1 = 0$. Then

$$X_t = x \exp\Big\{\big(a - \frac{1}{2}c_1^2\big)t + c_1 W_t^1\Big\}$$

which is a so-called *geometric Brownian motion* which is used as a simple model for the evolution of the value of an asset in mathematical finance.

Another special case of interest is $m = 1$ and $b = c_1 = 0$. Then

$$X_t = \exp\{at\}\Big(x + d_1 \int_0^t \exp\{-as\} \mathrm{d}W_s^1\Big).$$

This process is called *Ornstein-Uhlenbeck process.* Note that the Ornstein-Uhlenbeck process is a *Gaussian process* (as opposed to geometric Brownian motion).

It is natural to ask if the solution of an sde depends continuously upon the initial condition $x \in \mathbb{R}^d$. Since solutions are only uniquely defined up to sets of measure zero and these sets of measure zero may well depend on x the answer is in general *no.* We can however change the question slightly by asking if there exists a continuous modification of the process which associates to x the solution of the sde with initial condition x. Then the answer is *yes* at least under slightly stronger assumptions on the coefficients. This will follow easily from the following lemma. We denote by $\|.\|$ the norm on the space of real $d \times d$ matrices which is associated to the Euclidean norm on $\mathbb{R}^d$ (and which equals the largest eigenvalue in case the matrix is non-negative definite).

Lemma 2.34. *Let $p \geq 2$ and $K \geq 0$. Assume that parts i) and iii) of Assumption (H) hold and that instead of part ii) we even have*

$$2\langle b(x) - b(y), x - y\rangle + \mathrm{tr}(\mathcal{A}(x, y)) + (p - 2)\|\mathcal{A}(x, y)\| \leq K|x - y|^2 \tag{2.4.14}$$

for all $x, y \in \mathbb{R}^d$. Let $t \mapsto \phi_t(x)$ be the unique solution of (2.4.1) *with initial condition $x \in \mathbb{R}^d$. Then, for every $T > 0$, $x, y \in \mathbb{R}^d$ and $q \in (0, p)$, we have*

$$\mathbb{E} \sup_{0\leq t\leq T} |\phi_t(x) - \phi_t(y)|^q \leq (c_{q/p} + 1)|x - y|^q \exp\{KqT/2\},$$

where c_r is defined as in Proposition 2.29.

Proof. Fix $x \neq y \in \mathbb{R}^d$ and define $D_t := \phi_t(x) - \phi_t(y)$ and $Z_t := |D_t|^p$. Then, by Itô's formula,

$$\begin{aligned}\mathrm{d}Z_t = {} & p|D_t|^{p-2}\langle b(\phi_t(x)) - b(\phi_t(y)), D_t\rangle\,\mathrm{d}t + p|D_t|^{p-2}\sum_{k=1}^m \langle D_t, \sigma_k(\phi_t(x)) - \sigma_k(\phi_t(y))\rangle\,\mathrm{d}W_s^k \\ & + \frac{1}{2}p|D_t|^{p-2}\mathrm{tr}(\mathcal{A}(\phi_t(x), \phi_t(y))\,\mathrm{d}t + \frac{1}{2}p(p-2)|D_t|^{p-4}\langle D_t, \mathcal{A}(\phi_t(x), \phi_t(y))D_t\rangle\,\mathrm{d}t,\end{aligned}$$

where the last term should be interpreted as zero when $D_t = 0$ even if $p < 4$. Using the assumption in the lemma, we obtain

$$Z_t \leq |x - y|^p + \frac{p}{2}K \int_0^t Z_s\,\mathrm{d}s + N_t,$$

for some $N \in \mathcal{M}^0_{loc}$. Applying Proposition 2.31 we obtain for $r \in (0,1)$

$$\mathbb{E} \sup_{0 \le t \le T} Z_t^r \le |x-y|^{pr}(c_r+1)\exp\Big\{\frac{1}{2}KprT\Big\},$$

so the assertion follows by choosing $r = q/p$. □

Proposition 2.35. *Let all assumptions of Lemma 2.34 hold for some $p > d$. Then there exists a modification φ of the solution ϕ which depends continuously upon the initial condition.*

Proof. Let $q \in (d,p)$. Then Lemma 2.34 and Kolmogorov's continuity theorem (which is well-known from WT2! – if not: look at Theorem 5.2 in the appendix) show that ϕ has a modification φ (with respect to the spatial variable x) such that $x \mapsto \varphi(x)$ is continuous from $\mathbb{R}^d$ to $C([0,T],\mathbb{R}^d)$ with respect to the supremum norm for each $T > 0$ and each $\omega \in \Omega$. □

Remark 2.36. In fact one can even obtain Hölder regularity of φ using Lemma 2.34 and Kolmogorov's continuity theorem 5.2.

One may ask whether the solutions of an sde even generate a *stochastic flow*, i.e. if there exists a modification φ of the solution which is a stochastic flow in the following sense. Note that a stochastic flow has two time indices: the first stands for the initial time (which we have assumed to be 0 up to now) and the second one for the final time, so $\varphi_{s,t}(x)(\omega)$ denotes the value of the solution at time t which starts at location x at time s. Note that we allow t to be smaller than s in which case $\varphi_{s,t} = \varphi_{t,s}^{-1}$ and s and/or t may also be negative.

Definition 2.37. A measurable map $\varphi : \mathbb{R}^2 \times \mathbb{R}^d \times \Omega \to \mathbb{R}^d$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *stochastic flow (of homeomorphisms)*, if

- $\varphi_{s,u}(\omega) = \varphi_{t,u}(\omega) \circ \varphi_{s,t}(\omega)$ for all $s,t,u \in \mathbb{R}$ and all $\omega \in \Omega$.
- $\varphi_{s,s}(\omega) = \mathrm{id}|_{\mathbb{R}^d}$ for all $s \in \mathbb{R}$ and all $\omega \in \Omega$.
- $(s,t,x) \mapsto \varphi_{s,t}(x)(\omega)$ is continuous for every $\omega \in \Omega$.

Note that for a stochastic flow φ, the map $x \mapsto \varphi_{s,t}(x)(\omega)$ is a homeomorphism of $\mathbb{R}^d$ for each s,t,ω.

We state the following theorem without proof (for the proof of an even more general result, see [Ku90], p.155ff).

Theorem 2.38. *Assume that b and $\sigma_1,...,\sigma_m$ are maps from $\mathbb{R}^d$ to $\mathbb{R}^d$ satisfying a global Lipschitz condition. Then the sde* (2.4.1) *generates a stochastic flow, i.e. there exists a modification φ of the solution which can be extended to a stochastic flow*

Remark 2.39. If part iii) of Assumption (H) is not satisfied (but i) and ii) are), then it is not hard to show that a unique *local* solution of (2.4.1) still exists. Roughly speaking this means that for each initial condition there is a unique solution up to a (possibly finite) stopping time at which the solution process *explodes*. A corresponding statement also holds for the flow property: if we only require that for each $R > 0$ there exists some K_R such that (2.4.14) with K replaced by K_R holds for all $|x|,|y| \le R$, then we still obtain a modification which depends continuously upon the initial condition up to explosion. We point out that there exist examples without drift (i.e. $b = 0$) and bounded and C^∞ (but not globally Lipschitz) coefficients $\sigma_1,...,\sigma_m$ for which no global flow exists: the assumptions of Theorem 2.27 are satisfied, so a unique (almost surely non-exploding) solution exists for each initial condition and the solutions are locally continuous with respect to the initial condition but there exist initial conditions which explode in finite time (and these may even be dense in $\mathbb{R}^d$). Feel free to ask me for a reference if you are interested.

Theorem 2.40. *Assume that X is a solution of the sde*

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sum_{k=1}^m \sigma_k(X_t)\,\mathrm{d}W_t^k,\ X_0 = x,$$

on some FPS where $W = (W^1, ..., W^m)$ is d-dimensional $\mathbb{F}$-standard Brownian motion and $b, \sigma_1, ..., \sigma_m : \mathbb{R}^d \to \mathbb{R}^d$. Define the operator $\mathcal{L}$ from $C^2(\mathbb{R}^d, \mathbb{R})$ to the space of real-valued functions on $\mathbb{R}^d$ by

$$(\mathcal{L}f)(x) = \langle b(x), \nabla f(x)\rangle + \frac{1}{2}\sum_{i,j=1}^d \sum_{k=1}^m \sigma_k^i(x)\sigma_k^j(x)\frac{\partial^2 f}{\partial x_i \partial x_j}(x).$$

Then, for any $f \in C^2(\mathbb{R}^d, \mathbb{R})$,

$$t \mapsto f(X_t) - \int_0^t \big(\mathcal{L}f\big)(X_s)\,\mathrm{d}s \tag{2.4.15}$$

is a continuous local martingale.

Proof. This follows easily by applying Ito's formula to $f(X_t)$. □

Remark 2.41. The previous theorem says that X solves the *martingale problem* associated to $\mathcal{L}$ which by definition means that for any $f \in C^2(\mathbb{R}^d, \mathbb{R})$, the process defined in (2.4.15) is a (continuous) local martingale. We will see later that the concept of a martingale problem makes sense in a much wider set-up.

Chapter 3

Lévy Processes

In this short chapter, we provide an introduction to Lévy processes. As before, we fix a FPS $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$.

Definition 3.1. An adapted $\mathbb{R}^d$-valued stochastic process X (with index set $[0, \infty)$) is called an $\mathbb{F}$-Lévy process, if

(i) X has *independent increments*, i.e. for each $0 \le s \le t$, $X_t - X_s$ is independent of $\mathcal{F}_s$.

(ii) For every $0 \le s \le t < \infty$, $\mathcal{L}(X_t - X_s) = \mathcal{L}(X_{t-s})$.

(iii) X has *càdlàg* paths, i.e. for all $\omega \in \Omega$, the map $t \mapsto X_t(\omega)$ is right continuous and has left limits.

Remark 3.2. The acronym *càdlàg* is derived from the French *continue à droite, limites à gauche*. Note that (ii) implies that a Lévy process satisfies $\mathbb{P}(X_0 = 0) = 1$. Property (ii) (together with (i)) says that a Lévy process has *stationary increments*, i.e. the joint law of a finite number of increments of X remains unchanged if all increments are shifted by the same number h.

Lévy processes have many nice properties. They are semimartingales (this is not obvious) and Markov processes (which is even less obvious since so far we have not even introduced Markov processes).

We will now restrict to the case $d = 1$. Our aim is to classify all Lévy processes and to establish the semimartingale property. Before doing so, we point out that some authors (e.g. [Pr92]) replace property (iii) in the definition by the formally weaker requirement that the process is *continuous in probability* (i.e. $\lim_{h \to 0} X_{t+h} = X_t$ in probability for all t) and then prove that there exists a càdlàg modification.

Clearly, an $\mathbb{F}$-Brownian motion is a Lévy process and so is a Poisson process. A rather large class of Lévy processes is known from the Cramér-Lundberg model in *ruin theory*: if an insurance company has to settle i.i.d. claims of sizes $U_1, U_2, ..$ which arrive according to a Poisson process N_t with intensity λ which is independent of the claim sizes and if the continuous premium rate is $c > 0$, then the net income of the company up to time t is $X_t := ct - \sum_{i=1}^{N_t} U_i$ which is a Lévy process.

An even more general class of Lévy processes is the class of all processes of the form

$$X_t = bt + \sigma W_t + \sum_{i=1}^{N_t} U_i, \ t \ge 0, \tag{3.0.1}$$

where N is a Poisson process of (some) intensity $\lambda > 0$, $b \in \mathbb{R}$ and $U_1, U_2, \ldots$ are i.i.d. and independent of N. Are all Lévy processes of this form? The answer is no! The process X in (3.0.1) jumps only finitely many times in each bounded interval but a general Lévy process can have infinitely many jumps in a finite interval. Since X has to be càdlàg most of the jumps have to be very small however.

We will see that there is a very close link between Lévy processes and *infinitely divisible distributions*, so we continue by introducing the latter (without providing full proofs; a good reference is [Kl06]).

Definition 3.3. A probability measure $\mu \in \mathcal{M}_1(\mathbb{R})$ is called *infinitely divisible* if for each $n \in \mathbb{N}$ there exists some $\mu_n \in \mathcal{M}_1(\mathbb{R})$ such that $\mu = \mu_n^{*n}$, where ν^{*n} is the n-fold convolution of ν. We say that a real-valued random variable is infinitely divisible if its law $\mathcal{L}(X)$ is.

Remark 3.4. Recall that the *characteristic function* ϕ of $\mu \in \mathcal{M}_1(\mathbb{R})$ is defined as

$$\phi(u) = \int \mathrm{e}^{iux} \, \mathrm{d}\mu(x), \; u \in \mathbb{R}$$

and that the characteristic function of a convolution is the product of the characteristic functions. Therefore μ is infinitely divisible iff for each $n \in \mathbb{N}$ there exists a characteristic function whose n-th power is equal to ϕ.

Example 3.5. The following probability measures are infinitely divisible.

(i) $\mathcal{N}(m, \sigma^2)$

(ii) The Cauchy distribution Cau_a with density $f_a(x) = (a\pi)^{-1}(1 + (x/a)^2)^{-1}$, where $a > 0$ is a parameter. Its characteristic function is $\phi_a(u) = \mathrm{e}^{-a|u|}$ which equals the n-th power of $\phi_{a/n}(u)$.

(iii) The Poisson distribution.

The link between Lévy processes and infinitely divisible distributions is the following. Let X be a Lévy process and let $t > 0$ and $n \in \mathbb{N}$. Then $X_t = \sum_{i=1}^n (X_{it/n} - X_{(i-1)t/n})$, so X_t is the sum of n i.i.d. random variables and therefore (in particular) $\mathcal{L}(X_1)$ is infinitely divisible. One can show that conversely, for any infinitely divisible distribution μ, there exists a Lévy process X such that $\mathcal{L}(X_1) = \mu$ and that the law of X is uniquely determined by μ.

Now we show how to generate the most general Lévy process. A Lévy process is characterized by a so-called *Lévy triplet* (b, σ, ν).

Definition 3.6. A *Lévy triplet* (b, σ, ν) consists of

- $b \in \mathbb{R}$
- $\sigma \geq 0$ and
- A measure ν on $\mathbb{R}\backslash\{0\}$ such that $\int_{\mathbb{R}\backslash\{0\}} \left(1 \wedge x^2\right) \mathrm{d}\nu(x) < \infty$.

The Lévy process associated to a Lévy triplet is of the form $X_t = bt + \sigma W_t + J_t$, where W is standard one-dimensional Brownian motion and J is a process which is independent of W and which is defined via the measure ν. To explain how J is defined, we need the definition of a Poisson random measure associated to ν.

Definition 3.7. Let ν be as above. A random measure N on $[0,\infty)\times(\mathbb{R}\backslash\{0\})$ is called *Poisson measure with intensity* ν if

- $N([s,t]\times A)$ is $\mathrm{Poi}((t-s)\nu(A))$-distibuted for each $A\in\mathcal{B}(\mathbb{R}\backslash\{0\})$ such that $\nu(A)<\infty$.
- For disjoint measurable subsets $B_1,....B_m$ of $[0,\infty)\times(\mathbb{R}\backslash\{0\})$, the random variables $N(B_1),...,N(B_m)$ are independent.

Given the Poisson measure N associated to ν, we define processes P^k, $k\in\mathbb{N}_0$ as follows:

$$P_t^0 := \int_{[0,t]\times((-\infty,-1]\cup[1,\infty))} x\,N(\mathrm{d}s,\mathrm{d}x)$$
$$P_t^k := \int_{[0,t]\times((-\frac{1}{k},-\frac{1}{k+1}]\cup[\frac{1}{k+1},\frac{1}{k}))} x\,N(\mathrm{d}s,\mathrm{d}x), \quad k\in\mathbb{N}.$$

Now we define

$$X_t := bt+\sigma W_t + J_t := bt+\sigma W_t + P_t^0 + M_t := bt+\sigma W_t + P_t^0 + \sum_{k=1}^{\infty} Q_t^k, \quad t\geq 0$$

where

$$Q_t^k := P_t^k - \mathbb{E}P_t^k\big(= P_t^k - t\mathbb{E}P_1^k\big).$$

We claim that the infinite sum converges uniformly on compact time intervals in L^2 and that M is a right continuous martingale (with finite second moment). Once we have shown the latter, we know that X is a semimartingale since P^0 is a process of locally finite variation. Clearly, each of the processes Q^k is a martingale and the variance of Q_t^k is easily seen to be $t\int_{A_k} x^2\,\mathrm{d}\nu(x)$ where $A_k := (-\frac{1}{k},-\frac{1}{k+1}]\cup[\frac{1}{k+1},\frac{1}{k})$. Note that the martingales Q^k, $k\in\mathbb{N}$ are independent and their variances are summable due to the integrability condition on ν. Using Doob's L^2-inequality, we see that $\sum Q_t^k$ converges uniformly on compact intervals in L^2 to a process M which is therefore both a martingale and has right continuous paths.

One can show (which we will not do) that in this way one obtains *every* real-valued Lévy process.

We briefly introduce two important sub-classes of real-valued Lévy processes: subordinators and stable Lévy processes.

Definition 3.8. A Lévy process with almost surely non-decreasing paths is called *subordinator*.

The standard Poisson process is the most prominent example of a subordinator.

Remark 3.9. It is not hard to see (and very easy to believe!) that a Lévy process with triplet (b,σ,ν) is a subordinator if and only if each of the following are true:

- $\sigma=0$
- $\nu(-\infty,0)=0$
- $b\geq\int_{(0,1)} x\,\mathrm{d}\nu(x)$.

Definition 3.10. A real-valued Lévy process is called *stable process* of index $\alpha\in(0,2]$ if, for any $c>0$, the processes $(X_{ct})_{t\geq0}$ and $(c^{1/\alpha}X_t)_{t\geq0}$ have the same law.

It can be shown that for $\alpha>2$ there cannot exist a stable process of index α. Brownian motion is a stable process of index 2. For a complete characterization of all stable Lévy processes we refer the reader to the literature, e.g. [RW94], p77.

Chapter 4

Markov Processes

4.1 Markov transition functions and Markov processes

Definition 4.1. Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be measurable spaces. $K : E \times \mathcal{F} \to [0, \infty]$ is called a *kernel* (from E to F) if

(i) $x \mapsto K(x, B)$ is $\mathcal{E}$-measurable for every $B \in \mathcal{F}$ and

(ii) $B \mapsto K(x, B)$ is a measure on $(F, \mathcal{F})$ for every $x \in E$.

K is called *sub-Markov kernel* resp. *Markov kernel*, if in addition $K(x, F) \leq 1$ resp. $K(x, F) = 1$ for all $x \in E$.

In the following, we only consider (sub-)Markov kernels. The following lemma is very useful.

Lemma 4.2. *Let K be a sub-Markov kernel from $(E, \mathcal{E})$ to $(F, \mathcal{F})$ and assume that $H : (E \times F, \mathcal{E} \otimes \mathcal{F}) \to \mathbb{R}$ is measurable and either bounded or non-negative. Then*

$$\varphi(x) := \int_F H(x, y) K(x, \mathrm{d}y)$$

is $\mathcal{E}$-measurable.

Proof. We only show the claim for bounded H. Define

$$\mathcal{H} := \{H : E \times F \to \mathbb{R} |\, H \text{ measurable, bounded s.t. } \varphi \text{ measurable}\}.$$

Clearly, $\mathbb{1} \in \mathcal{H}$ (since K is a kernel) and $\mathcal{H}$ is a linear space. Let

$$K := \{\mathbb{1}_{A \times B} : A \in \mathcal{E},\, B \in \mathcal{F}\}.$$

Then K is multiplicative, $\sigma(K) = \mathcal{E} \otimes \mathcal{F}$ and $K \subset \mathcal{H}$ since

$$x \mapsto \int_F \mathbb{1}_{A \times B}(x, y)\, K(x, \mathrm{d}y) = \mathbb{1}_A(x) K(x, B)$$

is measurable for each $A \in \mathcal{E}$, $B \in \mathcal{F}$. Now, let $f : E \times F \to \mathbb{R}$ be bounded and measurable and assume that f_n is a sequence of nonnegative functions in $\mathcal{H}$ such that $f_n \uparrow f$. Then $\varphi_n(x) := \int_F f_n(x, y) K(x, \mathrm{d}y)$ converges pointwise to $\varphi(x) := \int_F f(x, y) K(x, \mathrm{d}y)$ and therefore φ is also measurable. We have thus verified all assumptions in the monotone class theorem (in the appendix) and therefore $\mathcal{H}$ contains every bounded and measurable function, so the claim follows. $\square$

In the following we will denote the space of real-valued bounded and measurable functions on $(E,\mathcal{E})$ by b$\mathcal{E}$. To each sub-Markov kernel K from $(E,\mathcal{E})$ to $(F,\mathcal{F})$ we can associate a linear operator T from b$\mathcal{F}$ to b$\mathcal{E}$ by

$$(Tf)(x) := \int_F f(y)\, K(x,\mathrm{d}y). \tag{4.1.1}$$

Note that the measurability of Tf follows from Lemma 4.2. The operator T has the following properties.

- T is linear.
- $T\mathbb{1} \le \mathbb{1}$ and $T\mathbb{1} = \mathbb{1}$ if K is a Markov kernel.
- $Tf \ge 0$ whenever $f \ge 0$.
- If (f_n) is a non-decreasing sequence of non-negative functions in b$\mathcal{F}$ which converges pointwise to $f \in$ b$\mathcal{F}$, then $Tf_n \to Tf$ pointwise (i.e. T is σ-continuous).

An operator $T :$ b$\mathcal{F} \to$ b$\mathcal{E}$ is called a *(sub-)Markov operator* if it satisfies the properties above. Conversely, given a (sub-)Markov operator T, we can define a (sub-)Markov kernel K as follows: $K(x,B) := (T\mathbb{1}_B)(x)$. Due to this one-to-one correspondence we will often use the same symbol for a (sub-)Markov kernel and its associated (sub-)Markov operator.

Remark 4.3 (Composition of kernels)**.** If K is a sub-Markov kernel from $(E,\mathcal{E})$ to $(F,\mathcal{F})$ and L is a sub-Markov kernel from $(F,\mathcal{F})$ to $(G,\mathcal{G})$, then

$$L \circ K(x,B) := \int_F L(y,B)\, K(x,\mathrm{d}y),\ x \in E,\ B \in \mathcal{G}$$

defines a sub-Markov kernel from $(E,\mathcal{E})$ to $(G,\mathcal{G})$. Note that the composition of kernels corresponds to the composition of the associated operators.

Let $(E,\mathcal{E})$ be a measurable space. As before, we denote the set of probability measures on $(E,\mathcal{E})$ by $\mathcal{M}_1(E)$.

Definition 4.4. A family $(P_t(x,\Gamma),\ t \ge 0,\ x \in E,\ \Gamma \in \mathcal{E})$ is called a *(Markov) transition function* on $(E,\mathcal{E})$ if each P_t is a Markov kernel on $(E,\mathcal{E})$ (i.e. from $(E,\mathcal{E})$ to $(E,\mathcal{E})$) and for each $s,t \ge 0$

$$P_{t+s} = P_t \circ P_s.$$

(P_t) is called *normal* if, in addition, for each $x \in E$ we have $\{x\} \in \mathcal{E}$ and $P_0(x,\{x\}) = 1$.

Note that the associated family of Markov operators forms a *semigroup* (P_t) of operators on b$\mathcal{E}$.

Example 4.5. For $d \in \mathbb{N}$ define

$$p(t,x,y) := (2\pi t)^{-d/2} \exp\Big\{-\frac{|y-x|^2}{2t}\Big\},\ t > 0,\ x,y \in \mathbb{R}^d$$

and

$$P_t(x,B) := \int_B p(t,x,y)\,\mathrm{d}y,\ t > 0,\ x \in \mathbb{R}^d,\ B \in \mathcal{B}(\mathbb{R}^d),$$

and $P_0(x,B) := 1$ if $x \in B$ and $P_0(x,B) := 0$ otherwise. Then (P_t) is a normal Markov transition function known as the *Brownian transition function*. Note that $P_t(x,.) = \mathcal{N}(x,t)$ which is the law of $x + W_t$ where W is Brownian motion.

Remark 4.6. Just as in the previous example one can define a Markov transition function associated to a Lévy process X. Just define $P_t(x, B) := \mathbb{P}(X_t \in B - x)$.

Definition 4.7. Let (P_t) be a Markov transition function on $(E, \mathcal{E})$ and let $H \subseteq \mathrm{b}\mathcal{E}$. (P_t) is called

a) *strongly continuous* on H, if for each $f \in H$ we have

$$\lim_{t \downarrow 0} \|P_t f - P_0 f\|_\infty = 0,$$

b) *weakly continuous* on H, if for each $f \in H$ and $x \in E$ we have

$$\lim_{t \downarrow 0} (P_t f)(x) = (P_0 f)(x).$$

Example 4.8. The Brownian semigroup is strongly continuous on the space $\{f \in C(\mathbb{R}^d) : \lim_{|x| \to \infty} f(x) = 0\}$. The Brownian semigroup is weakly but not strongly continuous on $C_b(\mathbb{R}^d)$, the space of bounded continuous functions from $\mathbb{R}^d$ to $\mathbb{R}$ but not even weakly continuous on $\mathrm{b}\mathbb{R}^d$ (just take $f = \mathbb{1}_{\{0\}}$ and $x = 0$ to see this).

Remark 4.9. If K is a sub-Markov kernel from $(E, \mathcal{E})$ to $(F, \mathcal{F})$, then one can adjoin an element ∂ to F and define $\tilde{F} := F \cup \{\partial\}$ and $\tilde{\mathcal{F}} := \sigma(F, \{\partial\})$ and define a unique Markov kernel $\tilde{K}$ from $(E, \mathcal{E})$ to $(\tilde{F}, \tilde{\mathcal{F}})$ satisfying $\tilde{K}(x, B) = K(x, B)$ for $B \subseteq F$ (and hence $\tilde{K}(x, \{\partial\}) = 1 - K(x, F)$). If K is a sub-Markov kernel from $(F, \mathcal{F})$ to $(F, \mathcal{F})$ then one can extend K to a unique Markov kernel $\bar{K}$ from $(\tilde{F}, \tilde{\mathcal{F}})$ to itself by defining $\bar{K}(x, B) := \tilde{K}(x, B)$ whenever $x \in F$ (and $B \in \tilde{\mathcal{F}}$) and $\bar{K}(\partial, B) = 1$ if $\partial \in B$ and 0 otherwise. In the same way, a sub-Markov transition function on $(E, \mathcal{E})$ (which we did not define but it should be clear how it should be defined) can be extended to a Markov transition function on $(\tilde{E}, \tilde{\mathcal{E}})$. The adjoined state ∂ is often called the *coffin* state (for obvious reasons).

Definition 4.10. A tuple

$$\mathbb{X} = (\Omega, \mathcal{F}, \mathbb{F}, (X_t)_{t \geq 0}, E, \mathcal{E}, (\theta_t)_{t \geq 0}, (\mathbb{P}_x)_{x \in E})$$

is called a *Markov process* or a *Markov family* if

(i) $(\Omega, \mathcal{F}, \mathbb{F})$ is a FMS,

(ii) $(E, \mathcal{E})$ is a measurable space,

(iii) (X_t) is adapted,

(iv) $\theta_t : \Omega \to \Omega$ and $\theta_t \circ \theta_s = \theta_{t+s}$ for all $s, t \geq 0$,

(v) $X_{t+h}(\omega) = X_t(\theta_h \omega)$ for all $t, h \geq 0$, $\omega \in \Omega$,

(vi) For each $x \in E$, $\mathbb{P}_x$ is a probability measure on $(\Omega, \mathcal{F})$,

(vii) $x \mapsto \mathbb{P}_x\big(X_t \in B\big)$ is $\mathcal{E}$-measurable for each $B \in \mathcal{E}$,

(viii) $\mathbb{P}_x\big(X_{t+h} \in B | \mathcal{F}_t\big) = \mathbb{P}_{X_t}\big(X_h \in B\big)$, $\mathbb{P}_x$-a.s. for all $x \in E$, $h \geq 0$, $t \geq 0$, $B \in \mathcal{E}$.

$\mathbb{X}$ is called *normal* if, in addition, $\{x\} \in \mathcal{E}$ for every $x \in E$ and $\mathbb{P}_x(X_0 = x) = 1$ for every $x \in E$.

In the following, we denote expected values with respect to $\mathbb{P}_x$ by $\mathbb{E}_x$.

Remark 4.11. Several authors do not include (iv) and (v) in the definition of a Markov process (and drop θ from the tupel $\mathbb{X}$). This is really a matter of taste. We will formulate our theorems such that they do not explicitly use θ but when we construct Markov processes explicitly, then we will check all conditions in the definition including (iv) and (v).

Proposition 4.12. *Let $\mathbb{X}$ be a Markov process. Then $P_t(x, B) := \mathbb{P}_x(X_t \in B)$, $x \in E$, $B \in \mathcal{E}$, is a Markov transition function. $\mathbb{X}$ is normal iff (P_t) is normal.*

Proof. It is clear that P_t is a Markov kernel on E for each $t \geq 0$. The final claim is also clear. It remains to show the semigroup property of (P_t). Let $s, t \geq 0$, $x \in E$ and $B \in \mathcal{E}$. Then

$$\begin{aligned} P_{s+t}(x, B) &= \mathbb{P}_x\big(X_{s+t} \in B\big) = \mathbb{E}_x \mathbb{1}_{\{X_{s+t} \in B\}} = \mathbb{E}_x\big(\mathbb{E}_x\big(\mathbb{1}_{\{X_{s+t} \in B\}} | \mathcal{F}_s\big)\big) \\ &= \mathbb{E}_x\big(\mathbb{P}_x\big(X_{s+t} \in B | \mathcal{F}_s\big)\big) = \mathbb{E}_x\big(\mathbb{P}_{X_s}\big(X_t \in B\big)\big) = \mathbb{E}_x\big(P_t(X_s, B)\big) \\ &= \int_\Omega P_t(X_s, B)\, \mathrm{d}\mathbb{P}_x = \int_E P_t(y, B) P_s(x, \mathrm{d}y) = P_t \circ P_s(x, B), \end{aligned}$$

where we used the Markov property and the transformation formula for integrals. □

Remark 4.13. If $\mathbb{X}$ is a Markov process with associated transition function (P_t) and $f \in \mathrm{b}\mathcal{E}$, then $\mathbb{E}_x f(X_t) = (P_t f)(x)$.

Conversely, given a Markov transition function, we can construct an associated Markov process (under weak assumptions on the state space E). Before doing so, we show that, using the monotone class theorem, we can formulate the Markov property much more generally.

Lemma 4.14. *Let $\mathbb{X}$ be a Markov process. Then, for each $f \in \mathrm{b}\mathcal{E}$, $x \in E$, $t, h \geq 0$ we have*

$$\mathbb{E}_x\big(f(X_{t+h})\big|\mathcal{F}_t\big) = \mathbb{E}_{X_t}(f(X_h)), \ \mathbb{P}_x\text{-a.s.} \tag{4.1.2}$$

Proof. Fix $t, h \geq 0$ and $x \in E$ and denote by $\mathcal{H}$ the set of all f for which (4.1.2) holds. Then $\mathcal{H}$ is a linear space which contains all indicators $\mathbb{1}_B$ for $B \in \mathcal{E}$ by the Markov property. Since $\mathcal{H}$ is closed with respect to nonnegative increasing uniformly bounded sequences, the monotone class theorem shows that $\mathcal{H}$ contains all bounded and measurable functions and the proof is complete. □

Lemma 4.15. *Let $\mathbb{X}$ be a Markov process with associated transition semigroup (P_t). Then, for $n \geq 2$, $x \in E$, $0 \leq t_1 < t_2 < ... < t_n$ and $B_1, ..., B_n \in \mathcal{E}$ we have*

$$\begin{aligned} &\mathbb{P}_x\big(X_{t_1} \in B_1, ..., X_{t_n} \in B_n\big) \\ &\qquad = \int_{B_1} ... \int_{B_{n-2}} \int_{B_{n-1}} P_{t_n - t_{n-1}}(x_{n-1}, B_n)\, P_{t_{n-1}-t_{n-2}}(x_{n-2}, \mathrm{d}x_{n-1}) ... P_{t_1}(x, \mathrm{d}x_1) \end{aligned}$$

This function is measurable with respect to x.

Proof. The measurability of the right hand side follows from Lemma 4.2. Let $A_k := \{X_{t_1} \in B_1, ..., X_{t_k} \in B_k\}$. Using Lemma 4.14, we get

$$\begin{aligned}
\mathbb{P}_x\big(A_n\big) &= \int_{A_{n-1}} \mathbb{1}_{B_n}(X_{t_n})\,\mathrm{d}\mathbb{P}_x \\
&= \int_{A_{n-1}} \mathbb{E}_x\big(\mathbb{1}_{B_n}(X_{t_n})|\mathcal{F}_{t_{n-1}}\big)\,\mathrm{d}\mathbb{P}_x \\
&= \int_{A_{n-1}} \mathbb{E}_{X_{t_{n-1}}}\big(\mathbb{1}_{B_n}(X_{t_n - t_{n-1}})\big)\,\mathrm{d}\mathbb{P}_x \\
&= \int_{A_{n-2}} \mathbb{1}_{B_{n-1}}(X_{t_{n-1}})\mathbb{E}_{X_{t_{n-1}}}\big(\mathbb{1}_{B_n}(X_{t_n - t_{n-1}})\big)\,\mathrm{d}\mathbb{P}_x \\
&= \int_{A_{n-2}} \mathbb{E}_x\big(\mathbb{1}_{B_{n-1}}(X_{t_{n-1}})\mathbb{E}_{X_{t_{n-1}}}\big(\mathbb{1}_{B_n}(X_{t_n - t_{n-1}})\big)|\mathcal{F}_{t_{n-2}}\big)\,\mathrm{d}\mathbb{P}_x \\
&= \int_{A_{n-2}} \mathbb{E}_{X_{t_{n-2}}}\Big(\mathbb{1}_{B_{n-1}}(X_{t_{n-1}-t_{n-2}})\mathbb{E}_{X_{t_{n-1}}}\big(\mathbb{1}_{B_n}(X_{t_n - t_{n-1}})\big)\Big)\,\mathrm{d}\mathbb{P}_x \\
&= ... \\
&= \int_{\Omega} \mathbb{1}_{B_1}(X_{t_1})\mathbb{E}_{X_{t_1}}\big(\mathbb{1}_{B_2}(X_{t_2-t_1})\mathbb{E}_{X_{t_2}}\big(\mathbb{1}_{B_3}(X_{t_3-t_2})...\big)\big)\,\mathrm{d}\mathbb{P}_x \\
&= \mathbb{E}_x\big(\mathbb{1}_{B_1}(X_{t_1})\mathbb{E}_{X_{t_1}}\big(\mathbb{1}_{B_2}(X_{t_2-t_1})\mathbb{E}_{X_{t_2}}\big(\mathbb{1}_{B_3}(X_{t_3-t_2})...\big)\big)\big) \\
&= P_{t_1}\big(\mathbb{1}_{B_1}P_{t_2-t_1}\big(\mathbb{1}_{B_2}...\big)\big)(x),
\end{aligned}$$

which is the right hand side of the formula. □

Now we generalize the last assertion in the previous Lemma.

Lemma 4.16. *Let $\mathbb{X}$ be a Markov process and $\mathcal{M} := \sigma\big(X_t, t \geq 0\big)$. Then for any $Y \in \mathrm{b}\mathcal{M}$, the map $x \mapsto \mathbb{E}_x(Y)$ is $\mathcal{E}$-measurable.*

Proof. Define

$$\mathcal{H} := \{Y \in \mathrm{b}\mathcal{M} : x \mapsto \mathbb{E}_x(Y) \text{ is measurable}\}$$

and

$$K := \big\{\mathbb{1}_{B_1 \times ... \times B_n}(X_{t_1}, ..., X_{t_n}) : n \in \mathbb{N},\, 0 \leq t_1 < ... < t_n,\, B_1, ..., B_n \in \mathcal{E}\big\}.$$

The previous lemma shows that $K \subset \mathcal{H}$. The remaining assumptions in the monotone class theorem are clearly satisfied, so the claim follows. □

Remark 4.17. It is generally *not* true that $x \mapsto \mathbb{E}_x Y$ is $\mathcal{E}$-measurable when Y is only assumed to be in $\mathrm{b}\mathcal{F}$.

From now on, we will assume that all transition functions and Markov processes are normal.

Proposition 4.18. *Let (P_t) be a Markov transition function of a Polish space $(E, \mathcal{E})$ (i.e. (E, d) is a separable and complete metric space and $\mathcal{E}$ is its Borel-σ-algebra). Then there exists a Markov process $\mathbb{X}$ whose associated Markov transition function is (P_t).*

Proof. The proof is adapted from [He79]. We use the *canonical* construction: let $\Omega := E^{[0,\infty)}$, $\mathcal{F} = \mathcal{E}^{[0,\infty)}$, $X_t(\omega) := \omega_t$ for $t \geq 0$, $(\theta_t\omega)_s := \omega_{t+s}$, $\mathcal{F}_t = \sigma(X_s,\, s \leq t)$. For given $x \in E$, we define $\mathbb{P}_x$ on $(\Omega, \mathcal{F})$ by requiring that

$$\mathbb{P}_x\big(X_{t_1} \in B_1,..., X_{t_n} \in B_n\big) \tag{4.1.3}$$
$$= \int_{B_1} ... \int_{B_{n-2}} \int_{B_{n-1}} P_{t_n - t_{n-1}}(x_{n-1}, B_n)\, P_{t_{n-1}-t_{n-2}}(x_{n-2}, \mathrm{d}x_{n-1}) ... P_{t_1}(x, \mathrm{d}x_1)$$

for all $n \geq 2$, $0 \leq t_1 < t_2 < ... < t_n$ and $B_1, ..., B_n \in \mathcal{E}$ (see Lemma 4.15). Since these sets form a $\cap$-stable generator of $\mathcal{F}$ there can be at most one such probability measure $\mathbb{P}_x$ on $(\Omega, \mathcal{F})$. In order to see that there exists at least one such $\mathbb{P}_x$, one has to check that (4.1.3) defines a consistent family of finite dimensional distributions (we will not do this formally). Due to our assumption that $(E, \mathcal{E})$ is Polish it follows from Kolmogorov's consistency theorem that there exists a probability measure $\mathbb{P}_x$ on $(\Omega, \mathcal{F})$ which satisfies (4.1.3). We have now defined all ingredients of a Markov family and need to check that they satisfy all items in the definition. This is clear for (i)-(vii). Note that $\mathbb{P}_x\big(X_t \in B\big) = P_t(x, B)$ by construction, so once we have verified the Markov property (viii) in Definition 4.10, then we have constructed a Markov process $\mathbb{X}$ with associated transition function (P_t).

Fix $x \in E$, $s, h \geq 0$ and $B \in \mathcal{E}$. We have to show that

$$P_h(X_s, B) = \mathbb{P}_x\big(X_{s+h} \in B | \mathcal{F}_s\big), \quad \mathbb{P}_x\text{-a.s.}$$

Clearly, the left hand side is $\mathcal{F}_s$-measurable, so it remains to show that for each $Q \in \mathcal{F}_s$, we have

$$\int_Q P_h(X_s, B)\, \mathrm{d}\mathbb{P}_x = \int_Q \mathbb{1}_B(X_{s+h})\, \mathrm{d}\mathbb{P}_x.$$

Both sides of the equality define a finite measure on $(\Omega, \mathcal{F}_s)$. To see that they are equal, it suffices to show that they agree on a $\cap$-stable generator of $\mathcal{F}_s$ and on the set Ω. As a $\cap$-stable generator of $\mathcal{F}_s$ we take the family of sets of the form $Q := \{X_{s_1} \in B_1, ..., X_{s_n} \in B_n, X_s \in B_{n+1}\}$, where $n \in \mathbb{N}_0$, $0 \leq s_1 < ... < s_n < s$, and $B_1, ..., B_{n+1} \in \mathcal{E}$. For such a set Q and $s_{n+1} := s$, we get

$$\begin{aligned}
\int_Q P_h(X_s, B)\, \mathrm{d}\mathbb{P}_x &= \int_\Omega \prod_{i=1}^{n+1} \mathbb{1}_{B_i}(X_{s_i}) P_h(X_{s_{n+1}}, B)\, \mathrm{d}\mathbb{P}_x \\
&= \int_{E^{n+1}} \Big(\prod_{i=1}^{n+1} \mathbb{1}_{B_i}(y_i)\Big) P_h(y_{n+1}, B) \mu_{s_1,...,s_{n+1}}(\mathrm{d}y) \\
&= \mathbb{P}_x\big(X_{s_1} \in B_1, ..., X_{s_{n+1}} \in B_{n+1}, X_{s+h} \in B\big) \\
&= \mathbb{P}_x(Q \cap \{X_{s+h} \in B\}) = \int_Q \mathbb{1}_B(X_{s+h})\, \mathrm{d}\mathbb{P}_x,
\end{aligned}$$

where we denoted the joint law of $(X_{s_1}, ..., X_{s_{n+1}})$ under $\mathbb{P}_x$ by $\mu_{s_1,...,s_{n+1}}$ (which is the image of $\mathbb{P}_x$ under the map $\omega \mapsto (X_{s_1}(\omega), ..., X_{s_{n+1}}(\omega))$) and we used the explicit formula (4.1.3) at the beginning of the third line. □

In view of Lemma 4.15 it is natural to ask if there is a corresponding formula for the joint law conditioned on $\mathcal{F}_t$. Indeed there is.

Proposition 4.19. *Let $\mathbb{X}$ be a Markov process with associated Markov transition semigroup (P_t). Then, for $n \geq 2$, $x \in E$, $0 \leq t_1 < t_2 < ... < t_n$, $t \geq 0$, and $B_1, ..., B_n \in \mathcal{E}$ we have*

$$\mathbb{P}_x\big(X_{t+t_1} \in B_1, ..., X_{t+t_n} \in B_n | \mathcal{F}_t\big) = \mathbb{P}_{X_t}\big(X_{t_1} \in B_1, ..., X_{t_n} \in B_n\big).$$

Proof. Obviously, the right hand side of the formula in the proposition is $\mathcal{F}_t$-measurable, so it is enough to prove that for each $A \in \mathcal{F}_t$ we have

$$\int_A \mathbb{P}_{X_t}\big(X_{t_1} \in B_1, ..., X_{t_n} \in B_n\big)\, \mathrm{d}\mathbb{P}_x = \int_A \mathbb{1}_{B_1}\big(X_{t+t_1}\big) ... \mathbb{1}_{B_n}\big(X_{t+t_n}\big)\, \mathrm{d}\mathbb{P}_x. \tag{4.1.4}$$

By Lemma 4.15, the left hand side of (4.1.4) equals

$$\mathbb{E}_x\Big(\mathbb{1}_A\big(P_{t_1}\big(\mathbb{1}_{B_1} P_{t_2 - t_1}\big(...\big)\big)\big)\big(X_t\big)\Big). \tag{4.1.5}$$

We can compute the right hand side of (4.1.4) as in the proof of Lemma 4.15 – the only difference (apart from the shifted times) being the restriction to the set $A \in \mathcal{F}_t$. Therefore, the right hand side of (4.1.4) equals

$$\begin{aligned}
&\int_A \mathbb{1}_{B_1}\big(X_{t+t_1}\big) \mathbb{E}_{X_{t+t_1}}\big(\mathbb{1}_{B_2} ...\big)\, \mathrm{d}\mathbb{P}_x \\
&\qquad = \int_A \mathbb{E}_x\Big(\mathbb{1}_{B_1}\big(X_{t+t_1}\big) \mathbb{E}_{X_{t+t_1}}\big(\mathbb{1}_{B_2} ...\big) \big| \mathcal{F}_t\Big)\, \mathrm{d}\mathbb{P}_x \\
&\qquad = \int_A \mathbb{E}_{X_t}\Big(\mathbb{1}_{B_1}\big(X_{t_1}\big) \mathbb{E}_{X_{t_1}}\big(\mathbb{1}_{B_2} ...\big)\Big)\, \mathrm{d}\mathbb{P}_x,
\end{aligned}$$

which equals (4.1.5). □

The following proposition generalizes the previous one.

Proposition 4.20. *Let $\mathbb{X}$ be a Markov process with associated Markov transition semigroup (P_t). Then, for $x \in E$, $t \geq 0$, and $F \in \mathrm{b}\mathcal{E}^{[0,\infty)}$ we have*

$$\mathbb{P}_x\big(F(X_{t+\cdot}) | \mathcal{F}_t\big) = \mathbb{P}_{X_t}\big(F(X_\cdot)\big).$$

Proof. This is a straightforward application of the monotone class theorem together with Proposition 4.19. □

4.2 Strong Markov property

Definition 4.21. Let $\mathbb{X}$ be a normal Markov process. $\mathbb{X}$ is called a *strong Markov process* if for each optional time τ the following is true:

(i) X_τ is $\mathcal{F}_\tau^+$-measurable, i.e. $\big\{X_\tau \in B\big\} \cap \{\tau < \infty\} \in \mathcal{F}_\tau^+$ for all $B \in \mathcal{E}$,

(ii) $\mathbb{P}_x\big(X_{\tau+t} \in B | \mathcal{F}_\tau^+\big) = \mathbb{P}_{X_\tau}\big(X_t \in B\big)$ on $\{\tau < \infty\}$, $\mathbb{P}_x$-a.s. for all $x \in E$, $t \geq 0$, $B \in \mathcal{E}$.

Lemma 4.22. *Let $\mathbb{X}$ be strong Markov. Then, for $f \in \mathrm{b}\mathcal{E}$, $x \in E$, $t \geq 0$, and optional time τ, we have*

$$\mathbb{E}_x\big(f(X_{\tau+t}) | \mathcal{F}_\tau^+\big) = \mathbb{E}_{X_\tau}\big(f(X_t)\big) \text{ on } \{\tau < \infty\}, \mathbb{P}_x\text{-a.s.}$$

Proof. This follows from a straightforward application of the monotone class theorem as in Lemma 4.14. $\square$

More generally, the following lemma holds.

Lemma 4.23. *Let $\mathbb{X}$ be strong Markov. Then, for $F \in b\mathcal{E}^{[0,\infty)}$, $x \in E$, and optional time τ, we have*

$$\mathbb{E}_x\big(F(X_{\tau+\cdot})|\mathcal{F}_\tau^+\big) = \mathbb{E}_{X_\tau}\big(F(X_\cdot)\big) \text{ on } \{\tau < \infty\}, \mathbb{P}_x\text{-a.s.}$$

Proof. This follows from a straightforward application of the monotone class theorem as in Proposition 4.20. $\square$

Theorem 4.24 (Blumenthal-0-1 law)**.** *Let $\mathbb{X}$ be a (normal) strong Markov process. Let $\mathcal{M}_t := \sigma(X_s, s \le t)$. Then $\mathcal{M}_0^+$ is trivial for each $\mathbb{P}_x$, i.e. for every $A \in \mathcal{M}_0^+$ and every $x \in E$, we have $\mathbb{P}_x(A) \in \{0,1\}$.*

Proof. Apply Lemma 4.23 to the optional time $\tau = 0$ and $F \in b\mathcal{E}^{[0,\infty)}$. Then, for each $x \in E$,

$$\mathbb{E}_x\big(F(X_\cdot)|\mathcal{F}_0^+\big) = \mathbb{E}_x\big(F(X_\cdot)\big),\ \mathbb{P}_x\text{-a.s.}$$

and hence

$$\mathbb{E}_x\big(F(X_\cdot)|\mathcal{M}_0^+\big) = \mathbb{E}_x\big(F(X_\cdot)\big),\ \mathbb{P}_x\text{-a.s.}$$

Let $\mathcal{M} := \sigma(X_\cdot)$. By the factorization theorem any real-valued $(\Omega, \mathcal{M})$-measurable map g can be factorized as $g = F(X_\cdot)$ for some real-valued $\mathcal{E}^{[0,\infty)}$-measurable map F. In particular, this holds true for $g = 1\!\!1_A$ when $A \in \mathcal{M}_0^+$ and then F can also chosen to be bounded. Therefore, for fixed $x \in E$,

$$g = 1\!\!1_A = \mathbb{E}_x\big(1\!\!1_A|\mathcal{M}_0^+\big) = \mathbb{E}_x\big(F(X_\cdot)|\mathcal{M}_0^+\big) = \mathbb{E}_x\big(F(X_\cdot)\big),\ \mathbb{P}_x\text{-a.s.},$$

so $g = 1\!\!1_A$ is $\mathbb{P}_x$-a.s. constant, so $\mathbb{P}_x(A) \in \{0,1\}$. $\square$

Theorem 4.25. *Let $\mathbb{X}$ be a normal Markov process with right continuous paths taking values in a metric space (E,d) (equipped with its Borel-σ algebra $\mathcal{E}$). Assume that its semigroup (P_t) maps $C_b(E)$ to itself. Then $\mathbb{X}$ is strong Markov.*

Proof. Fix an optional time τ. By Propositions 1.34 and 1.31 applied to the filtration $\mathbb{F}^+$, we know that X_τ is $\mathcal{F}_\tau^+$-measurable, so (i) in Definition 4.21 follows. It remains to show (ii) in Definition 4.21. For $n \in \mathbb{N}$ define

$$\tau_n := \begin{cases} \frac{k+1}{2^n}, & \tau \in \left[\frac{k}{2^n}, \frac{k+1}{2^n}\right), \qquad k = 0,1,... \\ \infty, & \tau = \infty. \end{cases}$$

Note that τ_n is a stopping time. Let $A \in \mathcal{F}_\tau^+$ and $x \in E$. We have to show that

$$\int_{A\cap\{\tau<\infty\}} f(X_{\tau+t})\,\mathrm{d}\mathbb{P}_x = \int_{A\cap\{\tau<\infty\}} \mathbb{E}_{X_\tau} f(X_t)\,\mathrm{d}\mathbb{P}_x \tag{4.2.1}$$

and that $\mathbb{E}_{X_\tau} f(X_t)$ is $\mathcal{F}_\tau^+$-measurable, both for every $f \in b\mathcal{E}$ (or every indicator function of a set in $\mathcal{E}$). The last property is true due to property (i) and Lemma 4.2, Remark 4.13 and the fact that a composition of measurable maps is measurable. Now we show (4.2.1) for $f \in C_b(E)$. Observe that for $A \in \mathcal{F}_\tau^+$ we have

$$A \cap \Big\{\tau_n = \frac{k}{2^n}\Big\} = \Big(A \cap \Big\{\tau < \frac{k}{2^n}\Big\}\Big) \setminus \Big(A \cap \Big\{\tau < \frac{k-1}{2^n}\Big\}\Big) \in \mathcal{F}_{k/2^n}.$$

On the one hand, by right continuity of the paths and using the fact that f is bounded and continuous, and using the dominated convergence theorem, we get

$$\lim_{n\to\infty} \int_{A\cap\{\tau<\infty\}} f(X_{\tau_n+t})\,\mathrm{d}\mathbb{P}_x = \int_{A\cap\{\tau<\infty\}} f(X_{\tau+t})\,\mathrm{d}\mathbb{P}_x.$$

Further,

$$\begin{aligned}
\int_{A\cap\{\tau<\infty\}} f(X_{\tau_n+t})\,\mathrm{d}\mathbb{P}_x &= \sum_{k=0}^{\infty} \int_{A\cap\{\tau_n=\frac{k}{2^n}\}} f(X_{\frac{k}{2^n}+t})\,\mathrm{d}\mathbb{P}_x \\
&= \sum_{k=0}^{\infty} \int_{A\cap\{\tau_n=\frac{k}{2^n}\}} \mathbb{E}_x\big(f(X_{\frac{k}{2^n}+t})|\mathcal{F}_{k/2^n}\big)\,\mathrm{d}\mathbb{P}_x \\
&= \sum_{k=0}^{\infty} \int_{A\cap\{\tau_n=\frac{k}{2^n}\}} \mathbb{E}_{X_{k/2^n}} f(X_t)\,\mathrm{d}\mathbb{P}_x \\
&= \int_{A\cap\{\tau<\infty\}} \mathbb{E}_{X_{\tau_n}} f(X_t)\,\mathrm{d}\mathbb{P}_x = \int_{A\cap\{\tau<\infty\}} (P_t f)(X_{\tau_n})\,\mathrm{d}\mathbb{P}_x.
\end{aligned}$$

Using the property $P_t(C_b) \subseteq C_b$ and right continuity of the paths, we obtain

$$\lim_{n\to\infty} \int_{A\cap\{\tau<\infty\}} (P_t f)(X_{\tau_n})\,\mathrm{d}\mathbb{P}_x = \int_{A\cap\{\tau<\infty\}} (P_t f)(X_\tau)\,\mathrm{d}\mathbb{P}_x = \int_{A\cap\{\tau<\infty\}} \mathbb{E}_{X_\tau} f(X_t)\,\mathrm{d}\mathbb{P}_x, \quad (4.2.2)$$

so we have shown (4.2.1) for all $f \in C_b(E)$. The general case now follows using the monotone class theorem. Define

$$\mathcal{H} := \Big\{f \in \mathrm{b}\mathcal{E} : \int_{A\cap\{\tau<\infty\}} f(X_{\tau+t})\,\mathrm{d}\mathbb{P}_x = \int_{A\cap\{\tau<\infty\}} \mathbb{E}_{X_\tau} f(X_t)\,\mathrm{d}\mathbb{P}_x\Big\}.$$

Clearly, $\mathcal{H}$ is a linear space which contains the constant function $\mathbb{1}$ and also the multiplicative class $K := C_b(E)$ (as we just showed). Further, $\mathcal{H}$ is closed with respect to monotone limits. It is well-known (e.g. from WT2 or Bauer [Ba68]) that for a metric space we always have $\sigma(C_b(E)) = \mathcal{B}(E)$ and therefore the monotone class theorem implies (4.2.1) and the proof is complete. □

Remark 4.26. Note that we used the fact that E is metric only at the end when we stated that $\sigma(C_b(E)) = \mathcal{B}(E)$ and when we applied Proposition 1.34. For a general topological space $\sigma(C_b(E)) = \mathcal{B}(E)$ is not always true. The σ-algebra $\sigma(C_b(E))$ which is the same as the σ-algebra generated by all continuous functions on a topological space is called *Baire σ-algebra*. It is easy to see that on each topological space the Baire-σ-algebra is contained in the Borel-σ-algebra but there are cases where they disagree. The simplest example is the space $E = \{a, b\}$ in which $\{a\}$ is open but $\{b\}$ is not. Then the Borel-σ-algebra is the same as 2^E whereas the Baire-σ-algebra consists of the set $\emptyset$ and E only (since every real-valued continuous function on E is necessarily constant). Note that the previous theorem remains true if we allow E to be a general topological space provided we replace the Borel-σ-algebra by the Baire σ-algebra (and allow $\mathbb{X}$ to take values in an arbitrary topological space). It is not hard to show that Proposition 1.34 (and Lemma 1.32) remain true for arbitrary topological spaces E if $\mathcal{B}(E)$ is replaced by the Baire σ-algebra. The interested reader can find more on Baire σ-algebras in the monograph of Bauer [Ba68].

We now state examples of Markov processes which are not strong Markov.

Example 4.27. Let (P_t) be the Brownian semigroup for $d = 1$ as defined in Example 4.5 and $\mathbb{X}$ the corresponding canonical Markov process as constructed in the proof of Proposition 4.18. Define $\tilde{\mathbb{P}}_x := \mathbb{P}_x$ for $x \in \mathbb{R}\backslash\{0\}$ and let $\tilde{\mathbb{P}}_0$ be the point mass on the function which is identically 0. It is easy to check that $\tilde{\mathbb{X}}$ defined like $\mathbb{X}$ with the only exception that $E^{[0,\infty)}$ is replaced by $\Omega = C([0,\infty))$ and that $\tilde{\mathbb{P}}_0$ is defined as mentioned is a Markov process: the only item requiring a proof is the Markov property (viii) which is clear when $x = 0$, but (viii) is also true when $x \neq 0$ because for each $t \geq 0$ we have $\mathbb{P}_x\big(X_t = 0\big) = 0$. To see that $\tilde{\mathbb{X}}$ is not strong Markov, define $\tau := \inf\{s \geq 0 : X_s = 0\}$. This is a stopping time which is almost surely finite (for each $\tilde{\mathbb{P}}_x$) but for which part (ii) of the definition of a strong Markov process obviously fails (take any $x \neq 0$, any $t > 0$ and, for example, $B = \{0\}$). In this example X has right continuous paths and is therefore progressive (so (i) of the definition of a strong Markov process holds) but the associated transition semigroup does not map $C_b(\mathbb{R})$ to itself.

Example 4.28. A slightly simpler example of a real-valued Markov process with continuous paths which is not strong Markov is the following one: let Ω be the set of functions X of the form $= \{X_t = a + bt : a, b \in \mathbb{R}\}$ where $a, b \in \mathbb{R}$ and $\mathcal{F}_t := \sigma(X_s, s \leq t)$. For $x > 0$ let $\mathbb{P}_x$ be concentrated on the function $x + t$, for $x < 0$ let $\mathbb{P}_x$ be concentrated on the function $x - t$, and let $\mathbb{P}_0$ be the probability measure on Ω which puts measure 1/2 on each of the functions t and $-t$. We can define a family of shifts θ as in the canonical construction. It is very easy to see that the tupel $\mathbb{X}$ is a Markov process. It is however not strong Markov. One can check this explicitly but it also follows from the Blumenthal-0-1 law: just observe that $\mathcal{M}_0^+ = \mathcal{F}_0^+$ is not trivial under $\mathbb{P}_0$, since by waiting an infinitesimal amount of time we can see whether the process starting at 0 moves up or down (both happen with probability 1/2).

It is not hard to construct an example of a real-valued Markov process for which the associated semigroup maps $C_b(\mathbb{R})$ to itself but for which the paths are not right continuous and the strong Markov property does not hold.

Remark 4.29. It is certainly of interest to formulate criteria solely in terms of the Markov transition function which guarantee that there exists a modification of the corresponding Markov process which is strong Markov or which has right continuous paths. Such criteria exist (e.g. [RW94] or [He79] but are usually formulated under the assumption that the state space is locally compact (meaning that each point has a compact neighborhood) with countable base (LCCB) which is a somewhat restrictive assumption. In addition, it is often not hard to show the validity of the assumptions of Theorem 4.25 directly. Certainly, $\mathbb{R}^d$ is LCCB but no infinite dimensional Banach space is LCCB. In comparison, the assumption that a space is Polish is considerably more general. Every separable Banach space is Polish.

Proposition 4.30. *Under the assumptions of Theorem 2.27, the solution process is an $\mathbb{R}^d$-valued Markov process which satisfies the assumptions of Theorem 4.25 and which is therefore strong Markov.*

Proof. We do not provide a proof of the Markov property (see e.g. [RW94]). Intuitively it should be obvious anyway. By definition, the paths of a solution are continuous. We only need to check that the associated semigroup maps $C_b(\mathbb{R}^d)$ to itself. This is true in general but we only show it under the assumption of Lemma 2.34 with $p = 2$. Fix $t > 0$ and $x \in \mathbb{R}^d$. For any sequence $x_n \to x$ it follows from Lemma 2.34 that the solution $\phi_t(x_n)$ converges to $\phi_t(x)$ in L^1 and hence the laws converge weakly meaning that $\mathbb{E}f(\phi_t(x_n))$ converges to $\mathbb{E}f(\phi_t(x))$ for every $f \in C_b(\mathbb{R}^d)$ which is the assertion. □

4.3 Hille-Yosida Theory

In this section we introduce resolvents and infinitesimal generators of Markov transition semigroups. The goal is to characterize a Markov semigroup by a single operator, the infinitesimal generator. We will not provide all proofs. Good references which also include the proofs are [Dy65], [EK86], and [RW94].

Let (P_t) be a normal sub-Markov transition function on a measurable space $(E, \mathcal{E})$. In addition, we assume that for every $B \in \mathcal{E}$, the map $(x,t) \mapsto P_t(x,B)$ is $\mathcal{E} \otimes \mathcal{B}[0,\infty)$-measurable (this is not automatically true).

Definition 4.31. For $\lambda > 0$, the map $R_\lambda : \mathrm{b}\mathcal{E} \to \mathrm{b}\mathcal{E}$ defined as

$$\big(R_\lambda f\big)(x) := \int_0^\infty \mathrm{e}^{-\lambda t}\big(P_t f\big)(x)\,\mathrm{d}t \tag{4.3.1}$$

is called the λ-*resolvent* associated to (P_t).

Remark 4.32. R_λ is well-defined and measurable by Lemma 4.2 and the fact that $(t,x) \mapsto \big(P_t f\big)(x)$ is measurable for each $f \in \mathrm{b}\mathcal{E}$ by our measurability assumption on (P_t).

Theorem 4.33. *Let R_λ be the λ-resolvent associated to (P_t) and $f, f_n \in \mathrm{b}\mathcal{E}$. Then*

i) $0 \le f \le \mathbb{1}$ *implies* $\lambda R_\lambda f \le \mathbb{1}$ *and* $|f| \le \mathbb{1}$ *implies* $|\lambda R_\lambda f| \le \mathbb{1}$.

ii) $R_\lambda - R_\mu + (\lambda - \mu) R_\lambda R_\mu = 0$ *for any* $\lambda, \mu > 0$ *(resolvent identity).*

iii) $f_n \downarrow 0$ *implies* $R_\lambda f_n \downarrow 0$ *(both pointwise).*

iv) $P_t \mathbb{1} = \mathbb{1}$ *for all* $t \ge 0$ *implies* $\lambda R_\lambda \mathbb{1} = \mathbb{1}$.

Proof. i), iii) and iv) are clear. To see ii), assume without loss of generality that $\mu \ne \lambda$ and compute $\big(R_\lambda R_\mu f\big)(x)$ using the definition of R_λ and R_μ and the semigroup property of (P_t). □

Next, we introduce an axiomatic approach to resolvents and semigroups and investigate their relation.

Definition 4.34. Let $(B, \|.\|)$ be a real Banach space. A family $(R_\lambda)_{\lambda > 0}$ of bounded linear operators on B is called *contraction resolvent* (CR) if

i) $\|\lambda R_\lambda\| \le 1$ for all $\lambda > 0$ and

ii) $R_\lambda - R_\mu + (\lambda - \mu) R_\lambda R_\mu = 0$ for any $\lambda, \mu > 0$ (resolvent identity).

A CR $(R_\lambda)_{\lambda>0}$ on B is called *strongly continuous contraction resolvent* (SCCR), if it also satisfies

iii) $\lim_{\lambda\to\infty} \|\lambda R_\lambda f - f\| = 0$ for all $f \in B$.

Remark 4.35. The family (R_λ) from Definition 4.31 is a contraction resolvent on the Banach space $B = (\mathrm{b}\mathcal{E}, \|.\|_\infty)$.

Definition 4.36. Let $(B, \|.\|)$ be a real Banach space. A family $(P_t)_{t\geq 0}$ of bounded linear operators on B is called *contraction semigroup* (CSG) if

i) $\|P_t\| \leq 1$ for $t \geq 0$ (contraction property)

ii) $P_s P_t = P_{s+t}$ for $s, t \geq 0$ (semigroup property).

A CSG $(P_t)_{t\geq 0}$ on B is called *strongly continuous contraction semigroup* (SCCSG), if it also satisfies

iii) $\lim_{t\to 0} \|P_t f - f\| = 0$ for all $f \in B$.

Proposition 4.37. *For a SCCSG (P_t), the map $t \mapsto P_t f$ is uniformly continuous for each $f \in B$.*

Proof. For $t, h \geq 0$ and $f \in B$, we have

$$\|P_{t+h} f - P_t f\| = \|P_t(P_h f - f)\| \leq \|P_t\|\|P_h f - f\| \leq \|P_h f - f\| \to 0 \text{ as } h \downarrow 0.$$

□

We want to define the infinitesimal generator of a CSG.

Definition 4.38. A linear operator $G : \mathcal{D}(G) \to B$ is called *infinitesimal generator* of the CSG (P_t) on B if

$$Gf = \lim_{t\downarrow 0} \frac{P_t f - f}{t}$$

and $\mathcal{D}(G)$ consists of all $f \in B$ for which the limit exists (with respect to the norm on B).

We will actually construct G differently (and call it A until we know that the two objects are actually identical).

Fix the Banach space B. Just like at the beginning of the section, we can associate to a SCCSG (P_t) a SCCR (R_λ) by

$$R_\lambda f := \int_0^\infty \mathrm{e}^{-\lambda t} P_t f \,\mathrm{d}t. \tag{4.3.2}$$

This definition requires some explanation. The integrand is a uniformly continuous B-valued (not real-valued) function, so we need to say what this means. One can define this integral like an improper Riemann integral due to the fact that the integrand is continuous and its norm decays exponentially fast. Further, one can show that

$$\|R_\lambda f\| = \|\int_0^\infty \mathrm{e}^{-\lambda t} P_t f \,\mathrm{d}t\| \leq \int_0^\infty \mathrm{e}^{-\lambda t} \|P_t f\| \,\mathrm{d}t.$$

(in general, the norm of an integral can be estimated from above by the integral of the norm.) In particular, $\|\lambda R_\lambda\| \leq 1$, so i) of Definition 4.34 holds. One can check as in the proof of Theorem

4.33 that ii) of Definition 4.34 holds as well. To see that (R_λ) is strongly continuous, we argue as follows:

$$\lambda R_\lambda f - f = \int_0^\infty \mathrm{e}^{-t}\big(P_{t/\lambda} f - f\big)\,\mathrm{d}t.$$

Taking norms on both sides, and using the facts that the norm of an integral can be estimated from above by the integral of the norm and that (P_t) is strongly continuous, it follows by dominated convergence that (R_λ) is a SCCR.

It is natural to ask if every SCCR can be represented in a form like (4.3.2). This is the content of the Hille-Yosida theorem which we will formulate below. Before doing this, we start with a CR on a Banach space B which is not necessarily strongly continuous and try to find a large subspace of B on which it is a SCCR.

Remark 4.39. If (R_λ) is a contraction resolvent, then R_λ and R_μ commute due to the resolvent identity. Further, the resolvent identity implies

$$R_\lambda(B) + (\lambda-\mu)R_\lambda R_\mu(B) = R_\mu(B),$$

and therefore $R_\mu(B) \subseteq R_\lambda(B)$. By symmetry and the fact that R_λ and R_μ commute, we also have $R_\lambda(B) \subseteq R_\mu(B)$, so $R_\lambda(B) = R_\mu(B)$. Therefore

$$\mathcal{R} := R_\lambda(B)$$

does not depend on $\lambda > 0$. Further, we define

$$B_0 := \overline{\mathcal{R}},$$

the closure of $\mathcal{R}$ in the Banach space B. Note that B_0 is itself a Banach space.

Proposition 4.40. $B_0 = \{f \in B : \lim_{\lambda\to\infty} \lambda R_\lambda f = f\}$.

Proof. First assume that for $f \in B$ we have $\lim_{\lambda\to\infty} \lambda R_\lambda f = f$. Then, for given $\varepsilon > 0$, we find some $\lambda_0 > 0$ such that for all $\lambda > \lambda_0$, we have $\|f - R_\lambda(\lambda f)\| < \varepsilon$. Therefore, $f \in \overline{\mathcal{R}} = B_0$.

Conversely, assume that $f \in B_0$ and let $\varepsilon > 0$. There exist $g, h \in B$ such that $\|h\| < \varepsilon/3$ and $f = R_1 g + h$ and for $\lambda > 1$ we have

$$\begin{aligned}
\|\lambda R_\lambda f - f\| &= \|\lambda R_\lambda(R_1 g + h) - (R_1 g + h)\| \le \|\lambda R_\lambda h\| + \|h\| + \|\lambda R_\lambda R_1 g - R_1 g\| \\
&\le \|\lambda R_\lambda\|\|h\| + \|h\| + \left\|\lambda \frac{R_1 g - R_\lambda g}{\lambda - 1} - \frac{(\lambda-1)R_1 g}{\lambda - 1}\right\| \\
&\le \frac{2}{3}\varepsilon + \frac{1}{\lambda-1}\big(\|\lambda R_\lambda g\| + \|R_1 g\|\big) \\
&\le \frac{2}{3}\varepsilon + \frac{1}{\lambda-1}\big(\|g\| + \|R_1 g\|\big) \to \frac{2}{3}\varepsilon,
\end{aligned}$$

so the assertion follows. □

Proposition 4.40 shows that (R_λ) is a SCCR on B_0. Now, we forget about the original Banach space B and start with a SCCR (R_λ) on the Banach space B_0 which we henceforth denote by B. Let

$$\mathcal{D} := R_\lambda(B)$$

be the range of R_λ. By Remark 4.39 $\mathcal{D}$ does not depend on λ and $\mathcal{D}$ is dense in B. Further, $R_\mu : B \to \mathcal{D}$ is bijective, since it is onto by definition of $\mathcal{D}$ and if $R_\mu f = 0$, then the resolvent identity implies that $R_\lambda f = 0$ for every $\lambda > 0$, whence $\lambda R_\lambda f = 0$ for all $\lambda > 0$ which, by strong continuity, shows that $f = 0$, so R_μ is one-to-one.

Next, we define an operator A (generally unbounded) from $\mathcal{D}$ to B which will turn out to be equal to the infinitesimal generator G defined above. Note however that so far we do not even know whether there exists a (SC)CSG associated to (R_λ).

Define the operator $\mathcal{A}_\lambda := \lambda - R_\lambda^{-1} := \lambda\,\mathrm{id}_B - R_\lambda^{-1}$ with $\mathcal{D}(A_\lambda) = \mathcal{D}$, so

$$\mathcal{A}_\lambda\big(R_\lambda f\big) = \lambda R_\lambda f - f.$$

Then, for $\mu > 0$, using the resolvent identity,

$$\begin{aligned}\mathcal{A}_\lambda(R_\mu f) &= \mathcal{A}_\lambda\big(R_\lambda f + (\lambda-\mu)R_\lambda R_\mu f\big) = \lambda R_\lambda f - f + (\lambda-\mu)\big(\lambda R_\lambda R_\mu f - R_\mu f\big)\\ &= \lambda\big(R_\lambda f + (\lambda-\mu)R_\lambda R_\mu f\big) - f - (\lambda-\mu)R_\mu f\\ &= \lambda R_\mu f - f - (\lambda-\mu)R_\mu f = \mu R_\mu f - f,\end{aligned}$$

showing that $\mathcal{A}_\lambda$ does not depend on the choice of λ. Therefore we define $A := \mathcal{A}_\lambda$ with $\mathcal{D}(A) = \mathcal{D}$ and we have the formula

$$R_\lambda = (\lambda - A)^{-1}.$$

Proposition 4.41. *The operator A is closed, i.e. if f_n is a sequence in $\mathcal{D}$ which converges (with respect to the norm on B) to $f \in B$ and for which Af_n converges to some $g \in B$, then $f \in \mathcal{D}$ and $Af = g$.*

Proof.

$$R_\lambda^{-1} f_n = (\lambda - A) f_n \to \lambda f - g.$$

Since R_λ is continuous, we see that f_n converges to $R_\lambda(\lambda f - g)$ which must therefore be equal to f. In particular, $f \in \mathcal{D}$ and

$$Af = A\big(R_\lambda(\lambda f - g)\big) = \lambda R_\lambda(\lambda f - g) - (\lambda f - g) = g.$$

□

Theorem 4.42 (Hille-Yosida). *Let (R_λ) be a SCCR on B. There exists a unique SCCSG (P_t) on B such that*

$$R_\lambda f = \int_0^\infty \mathrm{e}^{-\lambda t} P_t f \,\mathrm{d}t$$

for all $f \in B$ and $\lambda > 0$.

(P_t) is defined as follows: let $A_\lambda := \lambda(\lambda R_\lambda - I)$ and

$$P_{t,\lambda} := \exp\{tA_\lambda\} := \mathrm{e}^{-\lambda t}\sum_{n=1}^{\infty}(\lambda t)^n(\lambda R_\lambda)^n/n!.$$

Then

$$P_t f = \lim_{\lambda\to\infty} P_{t,\lambda} f, \quad f \in B, t \geq 0.$$

Let us now state a few results without proof. Observe that iii) states that A is the same as the infinitesimal generator G defined before.

Proposition 4.43. *In the set-up above (with A_λ as in the previous theorem), we have*

i) $f \in \mathcal{D}$ iff $g = \lim_{\lambda\to\infty} A_\lambda f$ exists. In this case $Af = g$.

ii) $P_t f - f = \int_0^t (P_s A) f \,\mathrm{d}s$ for all $f \in \mathcal{D}$.

iii) Assume $f \in B$. Then $f \in \mathcal{D}$ iff $g = \lim_{t\downarrow 0} \frac{P_t f - f}{t}$ exists. In this case $Af = g$.

iv) (P_t) is uniquely determined by A, i.e. any two SCCSGs with generator $(A, \mathcal{D}(A))$ coincide.

We finish this part by stating a necessary and sufficient criterion for an operator to be the infinitesimal generator of a SCCSG. The proof can be found in [Dy65].

Theorem 4.44. *Let $(A, \mathcal{D}(A))$ be a linear operator on a Banach space B. A is the infinitesimal generator of a SCCSG (P_t) (i.e.*

$$Af = \lim_{t\downarrow 0} \frac{P_t f - f}{t} \text{ for all } f \in \mathcal{D}(A)$$

and the existence of

$$\lim_{t\downarrow 0} \frac{P_t f - f}{t}$$

implies that $f \in \mathcal{D}(A)$) if and only if all of the following conditions hold.

i) $\mathcal{D}(A)$ is a dense subspace of B.

ii) $\mathcal{R}(\lambda - A) = B$ for some $\lambda > 0$.

iii) $\|\lambda f - Af\| \geq \lambda \|f\|$ for all $f \in \mathcal{D}(A)$, $\lambda > 0$ ("A is dissipative").

Example 4.45. Let (P_t) be the Brownian semigroup defined in Example 4.5 for $d = 1$. Let $B \subset \mathrm{b}\mathcal{B}(\mathbb{R})$ be some closed linear subspace, so B is a Banach space. (P_t) may or may not be strongly continuous on B. We want to identify the generator A (including its domain). The associated λ-resolvent is

$$(R_\lambda f)(x) = \int_0^\infty \mathrm{e}^{-\lambda t} \int_{\mathbb{R}} f(y) \frac{1}{\sqrt{2\pi t}} \exp\Big\{\frac{-(x-y)^2}{2t}\Big\} \,\mathrm{d}y \,\mathrm{d}t = \int_{\mathbb{R}} f(y) \frac{1}{\gamma} \mathrm{e}^{-\gamma|y-x|} \,\mathrm{d}y,$$

where $\gamma := \sqrt{2\lambda}$ (a proof of the last identity can be found in [RW94], p.234). Let us now assume that $B = C_0 := C_0(\mathbb{R})$ (the space of continuous functions on $\mathbb{R}$ which are 0 at infinity; see Example 4.8) and let $h \in \mathcal{D}(A)$ be in the domain of the generator. Since $\mathcal{D}(A) = \mathcal{D}$ is the range of R_λ, there exists some $f \in B$ such that

$$h(x) = \big(R_\lambda f\big)(x) = \int_{\mathbb{R}} f(y) \frac{1}{\gamma} \mathrm{e}^{-\gamma|y-x|} \,\mathrm{d}y.$$

Further, for $\varepsilon \neq 0$,

$$\frac{h(x+\varepsilon) - h(x)}{\varepsilon} = \int_{\mathbb{R}} f(y) \frac{1}{\gamma} \Big(\frac{\mathrm{e}^{-\gamma|y-x-\varepsilon|} - \mathrm{e}^{-\gamma|y-x|}}{\varepsilon}\Big) \,\mathrm{d}y \to \int_{\mathbb{R}} f(y) \mathrm{e}^{-\gamma|y-x|} \mathrm{sgn}(y-x) \,\mathrm{d}y = h'(x),$$

showing that $h \in C^1(\mathbb{R}) \cap C_0$. It is a nice exercise to prove that h is even C^2 and to show that

$$h''(x) = \int_{\mathbb{R}} f(y)\gamma \mathrm{e}^{-\gamma|y-x|}\,\mathrm{d}y - 2f(x) = \gamma^2 h(x) - 2f(x) = 2\lambda h(x) - 2f(x). \tag{4.3.3}$$

(To compute $h''(x)$ as the derivative of $h'(x)$ split the integral into two parts: a neighborhood of x and the rest.) From (4.3.3) we see that $h \in C^2 \cap C_0$ and $h'' \in C_0$. Therefore,

$$\mathcal{D}(A) \subseteq \{h \in C^2 \cap C_0 \text{ and } h'' \in C_0\}.$$

Further, for $h \in \mathcal{D}(A)$,

$$-\frac{1}{2}h''(x) + \lambda h(x) = f(x) = \big(R_\lambda^{-1}h\big)(x) = (\lambda - A)h(x),$$

so

$$Ah = \frac{1}{2}h''.$$

On the other hand, for $h \in C^2 \cap C_0$ and $h'' \in C_0$, define $\tilde{h} := R_\lambda\big(-\frac{1}{2}h'' + \lambda h\big)$, so $\tilde{h} \in \mathcal{D}$ and

$$-\frac{1}{2}\tilde{h}'' + \lambda\tilde{h} = -\frac{1}{2}h'' + \lambda h,$$

but the equation $-\frac{1}{2}\bar{h}'' + \lambda\bar{h} = 0$ has only one solution in the class $C^2 \cap C_0$, namely 0, so $\tilde{h} = h$, so $h \in \mathcal{D}$ and we have shown that

$$\mathcal{D}(A) = \{h \in C^2 \cap C_0 \text{ and } h'' \in C_0\} \text{ and } Ah = \frac{1}{2}h'' \text{ for } h \in \mathcal{D}(A).$$

Remark 4.46. One can do a similar but somewhat more complicated computation for the d-dimensional Brownian semigroup. It turns out that for $d \geq 2$ the domain is strictly larger than the one corresponding to the one-dimensional case (see e.g. [Dy65]).

Remark 4.47. Instead of starting with the Brownian semigroup restricted to C_0, we could have started with $B = \mathrm{b}\mathcal{B}(\mathbb{R})$. Then the space B_0 on which the resolvent is strongly continuous is the space B_0 of uniformly continuous functions which is strictly larger than C_0 and $\mathcal{D}$ also increases. For details, see [Dy65].

Remark 4.48. If we start with a normal Markov semigroup (P_t) on $\mathrm{b}\mathcal{E}$ which is not strongly continuous (but at least measurable in (t, x) so that the resolvent can be defined), then it may happen that the space B_0 on which the restriction of the resolvent (and of the semigroup) is strongly continuous is so much smaller than $\mathrm{b}\mathcal{E}$ that it is not possible to reconstruct (P_t) on $\mathrm{b}\mathcal{E}$ uniquely from its restriction (or, equivalently, from the generator of the restriction). Fortunately there are many cases in which this problem does not occur. The following definition provides such a case (see e.g [RW94], p241).

Definition 4.49. Let E be a locally compact space with countable base (LCCB) (i.e. the space has a countable base and each $x \in E$ has a compact neighborhood). Let $C_0(E)$ be the space of continuous real-valued functions which vanish at infinity (if the space is compact then we define $C_0(E) := C_b(E) = C(E)$). A strongly continuous sub-Markov semigroup of linear operators (P_t) on $C_0(E)$ is called a *Feller-Dynkin semigroup.*

Note that this way we avoid the problems above by assuming that the semigroup is strongly continuous on a sufficiently large subspace of $b\mathcal{E}$ which allows to reconstruct the original semigroup of sub-Markov kernels on E. Note that this was our approach for the Brownian semigroup. Let us discuss the example of the Poisson process.

Example 4.50. Let $E = \mathbb{N}_0$ (with the metric induced by the usual distance on $\mathbb{R}$). E is certainly locally compact. The Poisson transition function with rate $\lambda > 0$ is given by

$$P_t(x, \{y\}) = \mathrm{e}^{-\lambda t}(\lambda t)^{y-x}/(y-x)!, \quad t \geq 0, y \geq x \geq 0.$$

Clearly, the semigroup is Feller-Dynkin. What is its generator A (and what is $\mathcal{D}(A)$)? We will discuss this in class.

Remark 4.51. We will probably look at the following generalization of the previous example in class. The Poisson process with rate λ is a so-called *pure birth process* on $\mathbb{N}_0$. The birth rates are all equal to λ i.e. the parameter of the exponential waiting time to move from state i to state $i+1$ is always λ. Instead we can assume that this rate is λ_i. As long as the λ_i are uniformly bounded or at least their inverses are summable nothing spectacular happens. The corresponding semigroup is still Feller-Dynkin (exercise: prove this in case you missed the proof in class!). Dramatic things happen however if this summability condition fails. We will see.

The following theorem (and/or some of its corollaries) is often called *Dynkin formula*.

Theorem 4.52. *Let $\mathbb{X}$ be an $(E, \mathcal{E})$-valued progressive Markov process with generator $(A, \mathcal{D}(A))$, i.e. $(A, \mathcal{D}(A))$ is the generator of the restriction of the associated Markov semigroup (P_t) to a closed subspace B of $b\mathcal{E}$ on which it is strongly continuous. Then, for each $f \in \mathcal{D}(A)$ and $x \in E$, the process*

$$M_t := f(X_t) - \int_0^t (Af)(X_s)\,\mathrm{d}s \tag{4.3.4}$$

is a martingale with respect to $\mathbb{P}_x$.

Proof. Since $Af \in b\mathcal{E}$ and X is progressive, the integral in (4.3.4) is well-defined, continuous, adapted and uniformly bounded on compact time intervals. Further, $t \mapsto f(X_t)$ is bounded and progressive. In particular, M is adapted and integrable. It remains to show the martingale property. Let $t, u \geq 0$. Then, using the Markov property,

$$\begin{aligned}
\mathbb{E}_x\big(M_{t+u}\big|\mathcal{F}_t\big) &= \mathbb{E}_x\big(f(X_{t+u})\big|\mathcal{F}_t\big) - \mathbb{E}_x\Big(\int_0^{t+u}(Af)(X_s)\,\mathrm{d}s\big|\mathcal{F}_t\Big)\\
&= \mathbb{E}_{X_t}\big(f(X_u)\big) - \mathbb{E}_x\Big(\int_t^{t+u}(Af)(X_s)\,\mathrm{d}s\big|\mathcal{F}_t\Big) - \int_0^t (Af)(X_s)\,\mathrm{d}s\\
&= \big(P_u f\big)(X_t) - \mathbb{E}_x\Big(\int_t^{t+u}(Af)(X_s)\,\mathrm{d}s\big|\mathcal{F}_t\Big) - \int_0^t (Af)(X_s)\,\mathrm{d}s\\
&= f(X_t) + \int_0^u \big(P_s(Af)\big)(X_t)\,\mathrm{d}s - \mathbb{E}_x\Big(\int_t^{t+u}(Af)(X_s)\,\mathrm{d}s\big|\mathcal{F}_t\Big) - \int_0^t (Af)(X_s)\,\mathrm{d}s,
\end{aligned}$$

where we used Proposition 4.43 ii) in the last step. Using the Markov property again, we get

$$\begin{aligned}
\int_0^u \big(P_s(Af)\big)(X_t)\,\mathrm{d}s &= \int_0^u \mathbb{E}_{X_t}\big((Af)(X_s)\big)\,\mathrm{d}s = \int_0^u \mathbb{E}_x\big((Af)(X_{t+s})|\mathcal{F}_t\big)\,\mathrm{d}s\\
&= \int_t^{t+u} \mathbb{E}_x\big((Af)(X_s)|\mathcal{F}_t\big)\,\mathrm{d}s = \mathbb{E}_x\Big(\int_t^{t+u}(Af)(X_s)\,\mathrm{d}s\big|\mathcal{F}_t\Big).
\end{aligned}$$

(Exercise: prove that the last step is legal!) This finishes the proof of the martingale property. □

4.4 Invariant measures: existence

Definition 4.53. Let $(P_t)_{t\geq 0}$ be a normal Markov transition function on $(E,\mathcal{E})$. A probability measure μ on $(E,\mathcal{E})$ is called *invariant probability measure* of $(P_t)_{t\geq 0}$ if $P_t^*\mu=\mu$ for all $t\geq 0$, where $P_t^*:\mathcal{M}_1(E)\to\mathcal{M}_1(E)$ is defined as $(P_t^*\nu)(A)=\int_E P_t(y,A)\,\mathrm{d}\nu(y)=\int_E \big(P_t \mathbb{1}_A\big)(y)\,\mathrm{d}\nu(y)$.

A Markov transition function may or may not admit an invariant probability measure μ and if it does, then μ may or may not be unique. If there exists a unique invariant probability measure, then it may or may not be true that for each initial probability measure ν we have that $\lim_{t\to\infty}\mathbb{P}_t^*\nu=\mu$ in some sense, e.g. in total variation or weakly (the latter requires that we specify a topology on E).

In the rest of this section we will assume that the state space $(E,\mathcal{E})$ is a Polish space with complete metric ρ and that $\mathcal{E}$ is its Borel-σ-algebra. Denote by $B(x,\delta)$ the open δ-ball around x.

Definition 4.54. A normal Markov semigroup on E is called *stochastically continuous* if

$$\lim_{t\downarrow 0} P_t(x,B(x,\delta))=1,\ \text{for all } x\in E,\ \delta>0.$$

In the following, we denote the space of uniformly continuous and bounded real-valued functions on E by $UC_b(E)$ and the space of bounded real-valued Lipschitz continuous functions on E by $\mathrm{Lip}_b(E)$.

Proposition 4.55. *The following are equivalent for a normal Markovian semigroup (P_t).*

i) (P_t) *is stochastically continuous.*

ii) $\lim_{t\downarrow 0} P_t f(x)=f(x)$*, for all* $f\in C_b(E), x\in E$.

iii) $\lim_{t\downarrow 0} P_t f(x)=f(x)$*, for all* $f\in UC_b(E), x\in E$.

iv) $\lim_{t\downarrow 0} P_t f(x)=f(x)$*, for all* $f\in \mathrm{Lip}_b(E), x\in E$.

Proof. The proof is adapted from [DZ96], p13. Clearly, ii) implies iii) and iii) implies iv). To see that i) implies ii), take $f\in C_b(E)$, $x\in E$ and $\delta>0$. Then

$$\begin{aligned}|P_t f(x)-f(x)| &= \Big|\int_{B(x,\delta)}\big(f(y)-f(x)\big)\,P_t(x,\mathrm{d}y)+\int_{B(x,\delta)^c}\big(f(y)-f(x)\big)\,P_t(x,\mathrm{d}y)\Big|\\ &\leq \sup_{y\in B(x,\delta)}|f(y)-f(x)|+2\|f\|_\infty(1-P_t(x,B(x,\delta))),\end{aligned}$$

so ii) follows by the definition of stochastic continuity and the continuity of f.

Next, we show that iv) implies i). For given $x_0\in E$ and $\delta>0$ define

$$f(x):=\frac{1}{\delta}\big((\delta-\rho(x,x_0)\big)^+,\ x\in E.$$

Then $f \in \mathrm{Lip}_b(E)$ and

$$\begin{aligned} f(x_0) - P_t f(x_0) &= 1 - \int_E f(y)\, P_t(x_0, \mathrm{d}y) \\ &= 1 - \int_{B(x_0,\delta)} f(y)\, P_t(x_0, \mathrm{d}y) \\ &\geq 1 - P_t(x_0, B(x_0, \delta)), \end{aligned}$$

so i) follows. □

Proposition 4.56. *If (P_t) is stochastically continuous, then the map $t \mapsto \big(P_t f\big)(x)$ is right continuous for each $x \in E$ and $f \in C_b(E)$.*

Proof. For $t \geq 0$ and $h > 0$ we have

$$\big(P_{t+h} f\big)(x) - \big(P_t f\big)(x) = \big(P_t\big(P_h f - f\big)\big)(x) = \int_E \big(P_h f - f\big)(y) P_t(x, \mathrm{d}y) \to 0,$$

as $h \downarrow 0$, so the claim follows by dominated convergence. □

Proposition 4.57. *If (P_t) is stochastically continuous, then the map $(t, x) \mapsto \big(P_t f\big)(x)$ is measurable for each $f \in C_b(E)$.*

Proof. This follows from the fact that the function is right continuous in t by the previous lemma and measurable in x for fixed t (cf. Proposition 1.34). □

Definition 4.58. A stochastically continuous transition semigroup (P_t) is called *Feller semigroup* if $P_t(C_b) \subseteq C_b$ holds for every $t \geq 0$. A Markov process $\mathbb{X}$ is called *Feller process* if its associated semigroup is Feller.

For a Feller semigroup (P_t) we define

$$R_T(x, B) := \frac{1}{T} \int_0^T P_t(x, B)\, \mathrm{d}t,\ x \in E,\ T > 0,\ B \in \mathcal{E}.$$

Note that the integral exists and that R_T is a Markov kernel thanks to Proposition 4.57, so we will define $R_T^* \nu$ like in Definition 4.53. Observe however that (R_t) is not a Markov semigroup.

We will start by discussing methods to prove the existence of an invariant probability measure. The following theorem is often useful.

Theorem 4.59 (Krylov-Bogoliubov)**.** *Let (P_t) be a Feller semigroup. If for some $\nu \in \mathcal{M}_1(E)$ and some sequence $T_n \uparrow \infty$, $R_{T_n}^* \nu \to \mu$ weakly as $n \to \infty$, then μ is an invariant probability measure of (P_t).*

Proof. Fix $r > 0$ and $\varphi \in C_b(E)$. Then $P_r \varphi \in C_b(E)$ and – abbreviating $\langle \psi, \kappa \rangle := \int \psi\, \mathrm{d}\kappa$ for

$\psi \in C_b(E)$ and a finite measure κ on E –

$$\begin{aligned}
\langle \varphi, P_r^* \mu \rangle &= \langle P_r \varphi, \mu \rangle = \langle P_r \varphi, \lim_{n\to\infty} R_{T_n}^* \nu \rangle \\
&= \lim_{n\to\infty} \frac{1}{T_n} \langle P_r \varphi, \int_0^{T_n} P_s^* \nu \,\mathrm{d}s \rangle \\
&= \lim_{n\to\infty} \frac{1}{T_n} \langle \varphi, \int_r^{T_n + r} P_s^* \nu \,\mathrm{d}s \rangle \\
&= \lim_{n\to\infty} \Big(\frac{1}{T_n} \langle \varphi, \int_0^{T_n} P_s^* \nu \,\mathrm{d}s \rangle + \frac{1}{T_n} \langle \varphi, \int_{T_n}^{T_n + r} P_s^* \nu \,\mathrm{d}s \rangle - \frac{1}{T_n} \langle \varphi, \int_0^{r} P_s^* \nu \,\mathrm{d}s \rangle \Big) \\
&= \langle \varphi, \mu \rangle.
\end{aligned}$$

Since this holds for all $\varphi \in C_b(E)$ it follows that $P_r^* \mu = \mu$ and the theorem is proved. $\square$

Corollary 4.60. *If (P_t) is Feller and for some $\nu \in \mathcal{M}_1(E)$ and a some sequence $T_n \uparrow \infty$ the sequence $R_{T_n}^* \nu$, $n \in \mathbb{N}$ is tight, then (P_t) has at least one invariant probability measure.*

Proof. Prohorov's Theorem (which is well-known from WT2!) implies that any tight family of probability measures on a Polish space contains a weakly convergent subsequence. Therefore the claim follows from Theorem 4.59. $\square$

Note that the previous corollary implies that any Feller transition function on a *compact* space E admits at least one invariant probability measure (since in that case $\mathcal{M}_1(E)$ is tight). In the non-compact case however, an invariant probability measure μ need not exist.

We first apply the Krylov-Bogoliubov Theorem to Markov processes on $\mathbb{R}^d$ which are generated by stochastic differential equations. We will formulate sufficient conditions for the existence of an invariant probability measure in terms of a *Lyapunov function*. In the qualitative theory of ordinary differential equations, Lyapunov functions are nonnegative functions V on the state space such that $\lim_{|x|\to\infty} V(x) = \infty$ and such that V decreases along solution trajectories outside some bounded set K. Therefore solutions have to enter K eventually and therefore the ode has some kind of stability property. The situation is similar in the stochastic case. Due to the stochastic nature of the system we cannot expect the solutions to decrease almost surely but only on average, i.e. V applied to a solution is a supermartingale as long as it stays outside a compact set K. Let us formulate the precise assumptions.

Theorem 4.61. *Consider the stochastic differential equation*

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sum_{k=1}^m \sigma_k(X_t)\,\mathrm{d}W_t^k,$$

and assume that Assumption (H) from Chapter 2 is satisfied (so existence and uniqueness of a solution is ensured by Theorem 2.27). Let $\mathcal{L}$ be the operator defined in Theorem 2.40, i.e.

$$(\mathcal{L}f)(x) = \langle b(x), \nabla f(x) \rangle + \frac{1}{2} \sum_{i,j=1}^d \sum_{k=1}^m \sigma_k^i(x) \sigma_k^j(x) \frac{\partial^2 f}{\partial x_i \partial x_j}(x)$$

for $f \in C^2(\mathbb{R}^d)$. Let (P_t) be the Feller semigroup associated to the solution (see Proposition 4.30).

Assume that there exists a nonnegative function $V \in C^2(\mathbb{R}^d)$ such that

$$\sup_{|x|\geq R} \mathcal{L}V(x) =: -A_R \to -\infty \text{ as } R \to \infty.$$

Then (P_t) has at least one invariant probability measure.

Proof. The proof is adapted from [Kh12], p80. Take any $x \in \mathbb{R}^d$ and let $\phi_t(x)$ be the solution starting at x. Let $\tau_n := \inf\{t \geq 0 : |\phi_t(x)| \geq n\}$ and $\tau_n(t) := \tau_n \wedge t$. Then, by Theorem 2.40,

$$\mathbb{E}V\big(\phi_{\tau_n(t)}(x)\big) - V(x) = \mathbb{E}\int_0^{\tau_n(t)} \mathcal{L}V(\phi_u(x))\,\mathrm{d}u.$$

We estimate

$$\mathcal{L}V(\phi_u(x)) \leq -A_R \mathbb{1}_{[R,\infty)}(\phi_u(x)) + \sup_{y\in\mathbb{R}^d} \mathcal{L}V(y),$$

so

$$A_R\mathbb{E}\int_0^{\tau_n(t)} \mathbb{1}_{[R,\infty)}(\phi_u(x))\,\mathrm{d}u \leq c_1 t + c_2,$$

where $c_1 = \sup_{y\in\mathbb{R}^d} \mathcal{L}V(y)$ and $c_2 = V(x)$. Since $\lim_{n\to\infty} \tau_n(t) = t$, we get

$$\frac{1}{t}\int_0^t P_u\big(x, B^c(0,R)\big)\,\mathrm{d}u \leq \frac{c_1}{A_R} + \frac{c_2}{A_R t}. \tag{4.4.1}$$

Since $\lim_{R\to\infty} A_R = \infty$, the assumptions of the Krylov-Bogoliubov Theorem hold and hence there exists an invariant probability measure. □

Remark 4.62. Even though the previous theorem does not provide a formula for an invariant probability measure π, we can actually obtain some of its properties from the proof. If π is any limit point of the probability measures defined by the left hand side of (4.4.1) (which implies that π is invariant), then we get

$$\pi\big(\overline{B(x,R)}^c\big) \leq \frac{c_1}{A_R}.$$

If the A_R go to infinity sufficiently quickly, then we see that π has finite moments.

It is not always easy to find a Lyapunov function. In some cases, the particular Lyapunov function $V(x) = |x|^2$ works. Here is an example.

Example 4.63. In the set-up of the previous theorem, define $V(x) := |x|^2$ for $x \in \mathbb{R}^d$. Then $\mathcal{L}V(x) = 2\langle x, b(x)\rangle + \mathrm{tr}\big(a(x,x)\big)$, so if this function converges to $-\infty$ as $|x| \to \infty$, then an invariant probability measure exists. If, for example, all σ_k are bounded, then $\lim_{|x|\to\infty}\langle x, b(x)\rangle = -\infty$ is a sufficient condition for an invariant probability measure to exist.

4.5 Invariant measures: uniqueness

It turns out that *coupling* is a very efficient method to prove uniqueness of an invariant measure as well as convergence of the transition probabilities to the invariant probability measure. We saw that already in the case of Markov chains in WT1 or the course *Stochastische Modelle*.

Definition 4.64. Let μ and ν be probability measures on the measurable space $(\Omega, \mathcal{F})$. Any probability measure ξ on $(\Omega \times \Omega, \mathcal{F} \otimes \mathcal{F})$ for which the image of the projection map onto the first resp. second coordinate equals μ resp. ν is called a *coupling* of μ and ν.

As before, we denote by

$$\|\mu - \nu\|_{TV} := \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)|$$

the total variation distance of μ and ν. It is easy to check that this is a norm on the space of finite signed measures.

Proposition 4.65. *Let $(E, \mathcal{E})$ be a measurable space for which the diagonal $\Delta := \{(x,x) : x \in E\} \in \mathcal{E} \otimes \mathcal{E}$ (this is automatically true when E is a Polish space). Let μ and ν be probability measures on $(E, \mathcal{E})$. Then*

$$\|\mu - \nu\|_{TV} = \inf\{\xi(\Delta^c)\},$$

where the infimum is taken over all couplings ξ of μ and ν. There exists a coupling ξ_0 for which the infimum is attained (often called optimal coupling*).*

Proof. Let ξ be a coupling of μ and ν and let (X, Y) be an $E \times E$-valued random variable with law ξ defined on $(\Omega, \mathcal{F}, \mathbb{P})$. Then, for $A \in \mathcal{E}$,

$$\mu(A) - \nu(A) = \mathbb{P}(X \in A) - \mathbb{P}(Y \in A) \leq \mathbb{P}(X \in A, Y \notin A) \leq \mathbb{P}(X \neq Y) = \xi(\Delta^c)$$

and therefore $\|\mu - \nu\|_{TV} \leq \inf\{\xi(\Delta^c)\}$. For the proof that an optimal coupling exists (and its explicit construction), the reader is referred to the literature. □

Now we formulate a criterion in terms of couplings which is sufficient for uniqueness of an invariant probability measure and which, moreover, guarantees convergence of the transition probabilities in total variation.

Proposition 4.66. *Let $(E, \mathcal{E})$ be a measurable space for which the diagonal $\Delta := \{(x,x) : x \in E\} \in \mathcal{E} \otimes \mathcal{E}$ and let $\mathbb{X}$ be an E-valued Markov process. Assume that for each pair $x, y \in E$, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and processes (X_t^x) and (X_t^y) whose laws coincide with that of the Markov process started at x respectively y and which are coupled in such a way that $\lim_{t \to \infty} \mathbb{P}(X_t^x = X_t^y) = 1$. If there exists an invariant probability measure π, then it is unique and for each $x \in E$ we have*

$$\lim_{t \to \infty} \|P_t^* \delta_x - \pi\|_{TV} = 0. \tag{4.5.1}$$

Proof. Note that uniqueness follows from (4.5.1). If π is an invariant probability measure and

$x \in E$, then

$$\begin{aligned}
\|\pi - P_t^* \delta_x\|_{TV} &= \sup_{A \in \mathcal{E}} |\pi(A) - (P_t^* \delta_x)(A)| \\
&= \sup_{A \in \mathcal{E}} \big|(P_t^* \pi)(A) - (P_t^* \delta_x)(A)\big| \\
&= \sup_{A \in \mathcal{E}} \Big| \int_E (P_t^* \delta_y)(A) \, \pi(\mathrm{d}y) - (P_t^* \delta_x)(A)\Big| \\
&= \sup_{A \in \mathcal{E}} \Big| \int_E (P_t^* \delta_y)(A) - (P_t^* \delta_x)(A) \, \pi(\mathrm{d}y)\Big| \\
&\le \int_E \sup_{A \in \mathcal{E}} \big|(P_t^* \delta_y)(A) - (P_t^* \delta_x)(A)\big| \, \pi(\mathrm{d}y) \\
&\le \int_E \sup_{A \in \mathcal{E}} \big|\mathbb{P}\big(X_t^y \in A\big) - \mathbb{P}\big(X_t^x \in A\big)\big| \, \pi(\mathrm{d}y) \\
&\le \int_E \big|\mathbb{P}\big(X_t^y \ne X_t^x\big)\big| \, \pi(\mathrm{d}y)
\end{aligned}$$

so the claim follows using the assumption thanks to the dominated convergence theorem. □

We will now formulate a similar proposition which instead of total variation convergence yields only weak convergence. On the other hand, we do not require that the trajectories starting at different initial conditions couple in the sense that they become equal but we require only that they are close with high probability.

Proposition 4.67. *Let (E, ρ) be a Polish space and let $\mathbb{X}$ be an E-valued Markov process. Assume that for each pair $x, y \in E$, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and processes (X_t^x) and (X_t^y) whose laws coincide with that of the Markov process started at x respectively y and which are coupled in such a way that $\rho(X_t^x, X_t^y)$ converges to 0 in probability as $t \to \infty$. If there exists an invariant probability measure π, then it is unique and for each $x \in E$ we have*

$$\lim_{t \to \infty} P_t^* \delta_x = \pi,$$

where the limit is to be understood in the sense of weak convergence.

Proof. If π is an invariant probability measure, $x \in E$ and $f \in \mathrm{Lip}_b(E)$ with Lipschitz constant L, then

$$\begin{aligned}
\Big| \int_E f(y) \, \pi(\mathrm{d}y) - \int_E f(y) \, \big(P_t^* \delta_x\big)(\mathrm{d}y)\Big| &= \Big| \int_E \mathbb{E} f(X_t^z) \, \pi(\mathrm{d}z) - \int_E \mathbb{E} f(X_t^x) \, \pi(\mathrm{d}z)\Big| \\
&\le \int_E \mathbb{E}\big|f(X_t^z) - f(X_t^x)\big| \, \pi(\mathrm{d}z) \\
&\le \int_E \mathbb{E}\big((L\rho(X_t^z, X_t^x)) \wedge \big(2\|f\|_\infty\big)\big) \, \pi(\mathrm{d}z)
\end{aligned}$$

so the claim follows using the assumption thanks to the dominated convergence theorem. □

Let us apply the previous proposition to the solution of a stochastic differential equation.

Proposition 4.68. *Consider the stochastic differential equation*

$$\mathrm{d}X_t = b(X_t)\,\mathrm{d}t + \sum_{k=1}^{m} \sigma_k(X_t)\,\mathrm{d}W_t^k,$$

and assume that Assumption (H) from Chapter 2 is satisfied (so existence and uniqueness of a solution is ensured by Theorem 2.27). Assume further that there exists some $K < 0$ such that

$$2\langle b(x) - b(y), x - y\rangle + \sum_{k=1}^{m} |\sigma_k(x) - \sigma_k(y)|^2 \leq K|x - y|^2$$

for all $x, y \in \mathbb{R}^d$. Then the Markov semigroup associated to the sde has at most one invariant probability measure π and in that case all transition probabilities converge to π weakly.

Proof. Proceeding as in the proof of Lemma 2.34 with $p = 2$, denoting the solution starting in x by X_t^x and abbreviating $Z_t := |X_t^x - X_t^y|^2$ for given $x, y \in \mathbb{R}^d$, we get for $0 \leq s \leq t$:

$$Z_t \leq Z_s + K \int_s^t Z_u\,\mathrm{d}u + N_t - N_s,$$

where $N \in \mathcal{M}_{loc}^0$. Note that we cannot apply the (stochastic) Gronwall lemma since $K < 0$. Applying a straightforward stopping argument yields

$$\mathbb{E}Z_t \leq \mathbb{E}Z_s + K \int_s^t \mathbb{E}Z_u\,\mathrm{d}u,$$

from which we easily obtain

$$\mathbb{E}Z_t \leq Z_0 \mathrm{e}^{Kt} = |x - y|^2 \mathrm{e}^{Kt} \to 0,$$

since $K < 0$, so the assumptions of Proposition 4.67 are satisfied and so the claim follows. □

The following example shows that the existence of an asymptotic coupling alone does not guarantee *existence* of an invariant probability measure.

Example 4.69. Consider the one-dimensional ode $y' = b(y)$, where $b(y) = \mathrm{e}^{-y}$ for $y \geq 0$ and arbitrary otherwise (but bounded, strictly positive and $C^\infty(\mathbb{R})$). If $y_0 \geq 0$, then $y_t = \log(t + \mathrm{e}^{y_0})$, $t \geq 0$ is the unique solution of the equation, so all solutions (also those starting at negative values) converge to $+\infty$ as $t \to \infty$. In particular, there does not exist an invariant probability measure. On the other hand the difference of any pair of solutions with different initial conditions converges to 0 as $t \to \infty$.

Our next aim is to establish the classical criteria for the uniqueness of an invariant probability measure (i.e. those which do not involve coupling). We essentially follow Chapters 3 and 4 of [DZ96].

Definition 4.70. A Markovian semigroup (P_t) is called

i) t_0*-regular* if all probability measures $P_{t_0}(x, .)$, $x \in E$ are mutually equivalent (i.e. mutually absolutely continuous, i.e. have the same sets of measure 0),

ii) *strong Feller at* t_0 if $P_{t_0}(\mathrm{b}\mathcal{E}) \subseteq C_b(E)$,

iii) t_0-*irreducible* if we have $P_{t_0}(x, B) > 0$ for all $x \in E$ and all non-empty open sets B.

Note that ii) and iii) require that we specify a topology on E while i) does not.

Remark 4.71. It follows from the semigroup property that if (P_t) is t_0-regular resp. strong Feller at t_0 resp. t_0-irreducible, then the same holds for t_0 replaced by some $s > t_0$. Note that the irreducibility concept is rather strong. There exist (many) discrete time Markov chains with countable state space E which are irreducible (in the sense that for each pair of states $i, j \in E$ there exists some n such that the n-step transition probability from i to j is strictly positive) but which are not n-irreducible for any $n \in \mathbb{N}$, e.g. the simple symmetric random walk.

Example 4.72. The Brownian semigroup (P_t) is t_0-regular, strong Feller at t_0 and t_0-irreducible for any $t_0 > 0$.

Remark 4.73. We will show below that a stochastically continuous semigroup which is strong Feller and irreducible has at most one invariant probability measure. We point out, that none of these two properties (strong Feller and irreducible) alone suffices for this – even if it holds for every $t_0 > 0$. One example is $E = \{0, 1\}$ with the identity semigroup, i.e. $P_t(i, \{i\}) = 1$ for all $t \geq 0$ and $i \in \{0, 1\}$ which is strong Feller at t_0 for each $t_0 \geq 0$ (since every real-valued function on E is continuous) but not t_0-irreducible for any $t_0 > 0$ and which has more than one invariant probability measure (in fact *every* probability measure on E is invariant). It is slightly more difficult to construct an example of a stochastically continuous semigroup which is irreducible but not strong Feller and which has more than one invariant probability measure (I will show an example in class, don't miss it!).

The following proposition is due to Khasminskii. The proof is adapted from [DZ96].

Proposition 4.74. *If the Markov semigroup* (P_t) *is strong Feller at* $t_0 > 0$ *and* s_0*-irreducible, then it is* $t_0 + s_0$*-regular.*

Proof. Assume that for some $x_0 \in E$ and $B \in \mathcal{E}$, $P_{t_0+s_0}(x_0, B) > 0$. Since

$$P_{t_0+s_0}(x_0, B) = \int_E P_{t_0}(y, B)\, P_{s_0}(x_0, \mathrm{d}y),$$

it follows that there exists some $y_0 \in E$ for which $\big(P_{t_0} \mathbb{1}_B\big)(y_0) = P_{t_0}(y_0, B) > 0$. By the strong Feller property at t_0 there exists some $r_0 > 0$ such that $P_{t_0}(y, B) > 0$ for all $y \in B(y_0, r_0)$.

Consequently, for arbitrary $x \in E$,

$$P_{t_0+s_0}(x, B) = \int_E P_{t_0}(y, B)\, P_{s_0}(x, \mathrm{d}y) \geq \int_{B(y_0,r_0)} P_{t_0}(y, B)\, P_{s_0}(x, \mathrm{d}y) > 0,$$

since (P_t) is s_0-irreducible, so the claim follows. □

The following theorem (or a slight modifications of it) is known as *Doob's theorem.*

Theorem 4.75. *Let* (P_t) *be a stochastically continuous Markov semigroup with invariant probability measure* π*. If* (P_t) *is* t_0*-regular for some* $t_0 > 0$*, then* π *is the unique invariant probability measure and*

$$\lim_{t\to\infty} \|P_t^* \delta_x - \pi\|_{TV} = 0$$

for each $x \in E$.

Proof. For a (not very short) proof, see [DZ96], p.43ff and Remark 4.2.3 in the same reference. The proofs I have seen so far do not use coupling. On the other hand it should be possible to find a coupling proof and apply Proposition 4.66 (note added in September 2014: recently such a proof has been found in a paper by Kulik and Scheutzow). Note however that the coupling proof will only work under the condition that an invariant probability measure exists (which we do assume). It is easy to construct examples which are t_0-regular and for which the assumptions of Proposition 4.66 are not satisfied. □

Remark 4.76. The readers who are interested in learning more about existence and uniqueness of invariant measures and convergence of transition probabilities to the invariant measure are referred to papers and lecture notes of Martin Hairer (Warwick) or the classical monograph of Meyn and Tweedie.

4.6 Coupling, optimal transportation and Wasserstein distance

Let us add a few comments on the connection between coupling, *optimal transportation* and the *Wasserstein distance*.

The *optimal transportation problem* was first formulated by the French mathematician Gaspard Monge in 1781. In today's language it can be phrased as follows: given two probability measures μ and ν on a measurable space $(E, \mathcal{E})$ and a measurable *cost function* $c : E \times E \to [0, \infty)$, find a *transport map* $T : E \to E$ which minimizes the *transportation cost*

$$\int_E c(x, T(x))\,\mathrm{d}\mu(x)$$

subject to the constraint $\nu = \mu T^{-1}$, i.e. T is the cheapest way to transport goods which are distributed according to μ (e.g. in factories) to customers which are distributed according to ν.

It is easy to see that an optimal transport may not exist or – even worse – there may not be any admissible transport at all. A simple example is the case in which μ is a Dirac measure and ν is not. Therefore it is more natural to allow *transportation kernels* rather than transportation maps (i.e. allow that goods from the same factory can be shipped to different customers). So the modified problem (formulated by the Russian mathematician Kantorovich) is to find a kernel K on $E \times E$ which minimizes

$$\int_E \Big(\int_E c(x, y)\, K(x, \mathrm{d}y) \Big) \,\mathrm{d}\mu(x),$$

subject to the constraint that $\nu(A) = \int_E K(x, A)\,\mathrm{d}\mu(x)$ for each $A \in \mathcal{E}$ or to find a coupling ξ of μ and ν which minimizes

$$\int_{E\times E} c(x, y)\,\mathrm{d}\xi(x, y). \tag{4.6.1}$$

Note that at least on nice spaces these two formulations are equivalent: each kernel K as above defines a coupling $\xi := \mu \otimes K$ (this is always true) and conversely, if E is a Polish space, then each probablity measure ξ on $E \times E$ can be disintegrated in the form $\xi = \mu \otimes K$ (cf. WT2: section on regular conditional probabilities).

Note that if we choose $c(x, y) = 1$ if $x \neq y$ and $c(x, y) = 0$ otherwise, then the coupling ξ which minimizes the integral in (4.6.1) is precisely the one for which the expression equals the total variation distance of μ and ν. Other choices of cost functions c lead to other distance measures of probability spaces.

Definition 4.77. Let (E, ρ) be a separable metric space, $p \geq 1$ and define

$$\mathcal{M}_1^p(E) := \{\mu \in \mathcal{M}_1(E) : \int_E \rho(x, x_0)^p \, \mathrm{d}\mu(x) < \infty \text{ for some } x_0 \in E\}.$$

$$W_p(\mu, \nu) := \Big(\inf_{\xi} \int_{E \times E} \rho(x, y)^p \, \mathrm{d}\xi(x, y) \Big)^{1/p}$$

is called the p-th *Wasserstein distance* of μ and ν in $\mathcal{M}_1^p(E)$.

Note that the set $\mathcal{M}_1^p(E)$ is independent of the choice of x_0 (by the triangle inequality). The reason for requiring the space to be separable is only to guarantee that ρ is jointly measurable.

It is not hard to show that on a bounded separable metric space W_p is a metric which induces weak convergence. On a discrete space with $\rho(x, y) = 1$ if $x \neq y$ and $\rho(x, y) = 0$ otherwise, we see that the first Wasserstein distance equals the total variation distance.

Chapter 5

Appendix

The following *Monotone class theorem* (also called *Dynkin class theorem*) is important. A good reference is the monograph [Sh88].

Theorem 5.1 (Monotone class theorem)**.** *Let K be a* multiplicative *family of bounded real-valued functions on a nonempty set S, i.e. $f, g \in K$ implies $fg \in K$. Let $\mathcal{S} := \sigma(K)$. If V is a linear space of real-valued functions on S satisfying*

i) $\mathbb{1} \in V$,

ii) $(f_n) \subset V, 0 \leq f_1 \leq ... \uparrow f$, f *bounded, then* $f \in V$, *and*

iii) $K \subseteq V$,

then $V \supseteq \mathrm{b}\mathcal{S}$, where $\mathrm{b}\mathcal{S}$ denotes the set of all bounded and measurable function on S.

Proof. **Step 1.** (according to [Sh88]):

We show that V is automatically closed with respect to uniform convergence:

$$(f_n) \subset V,\ \lim_{n\to\infty} f_n = f \text{ unif. } \Rightarrow f \in V.$$

To see this, let $\|.\|$ be the supremum norm on the set of bounded real-valued functions on S. Switching to a sub-sequence if necessary we can and will assume that the sequence $\varepsilon_n := \|f_{n+1} - f_n\|$ is summable. Let $a_n := \varepsilon_n + \varepsilon_{n+1} + ...$ and $g_n := f_n - a_n + 2a_1$. Then $g_n \in V$ by i) and the fact that V is a linear space. The sequence (g_n) is uniformly bounded and $g_{n+1} - g_n = f_{n+1} - f_n + \varepsilon_n \geq 0$ and $g_1 = f_1 + a_1 \geq 0$. Now ii) implies that $g := \lim_{n\to\infty} g_n \in V$ and so is $f = g - 2a_1$.

Step 2. (according to [EK86]):

Let $f \in K$ and assume that $F : \mathbb{R} \to \mathbb{R}$ is continuous. On each bounded interval, F is the uniform limit of polynomials and hence $F(f) \in V$ due to Step 1. For $a \in \mathbb{R}$, $f \in K$ and $f^{(n)} := (1 \wedge (f - a) \vee 0)^{1/n}$, $n \in \mathbb{N}$. Then $f^{(n)} \in K$ and $0 \leq f^{(1)} \leq ... \leq \mathbb{1}$ and hence $1_{\{f>a\}} = \lim_{n\to\infty} f^{(n)} \in K$ by ii). In the same way, we obtain for each $m \in \mathbb{N}$, $f_1, ..., f_m \in K$ and $a_1, ..., a_m \in \mathbb{R}$ that $\mathbb{1}_{\{f_1>a_1,...,f_m>a_m\}} \in K$ (to see this let $f^{(n)} := \Pi_{i=1}^m (1 \wedge (f_i - a_i) \vee 0)^{1/n}$ and apply Weierstrass' approximation theorem to $F \in \mathrm{C}(\mathbb{R}^m, \mathbb{R})$). The set of all $A \subseteq S$ for which $\mathbb{1}_A \in K$ is a Dynkin class which thus contains the family $\mathcal{U}$ of all sets of the form

$\{f_1 > a_1, ..., f_m > a_m\}$ with $f_1, ..., f_m \in K$ which is a $\cap$-stable generator of $\mathcal{S}$. Therefore $\mathbb{1}_A \in K$ for every $A \in \mathcal{S}$. Since every bounded and $\mathcal{S}$-measurable function is the uniform limit of simple $\mathcal{S}$-measurable functions the claim follows again by Step 1. $\square$

Here is a version of Kolmogorov's continuity theorem. Here, $|.|_1$ and $|.|_\infty$ denote the l^1-norm resp. the maximum norm on $\mathbb{R}^d$.

Theorem 5.2. *Let $\Theta = [0,1]^d$ and let Z_x, $x \in \Theta$ be a stochastic process with values in a separable and complete metric space (E, ρ). Assume that there exist $a, b, c > 0$ such that for all $x, y \in [0,1]^d$, we have*

$$\mathbb{E}\left((\rho(Z_x, Z_y))^a\right) \le c|x-y|_1^{d+b}.$$

Then Z has a continuous modification $\bar{Z}$. For each $\kappa \in (0, b/a)$, there exists a random variable S such that $\mathbb{E}(S^a) \le \frac{cd2^{a\kappa-b}}{1-2^{a\kappa-b}}$ and

$$\sup\left\{\rho(\bar{Z}_x(\omega), \bar{Z}_y(\omega)) : x, y \in [0,1]^d, |x-y|_\infty \le r\right\} \le \frac{2d}{1-2^{-\kappa}} S(\omega) r^\kappa \tag{5.0.1}$$

for each $r \in [0,1]$. In particular, for all $u > 0$, we have

$$\mathbb{P}\left\{\sup_{x,y\in[0,1]^d} \rho(\bar{Z}_x, \bar{Z}_y) \ge u\right\} \le \left(\frac{2d}{1-2^{-\kappa}}\right)^a \frac{cd2^{a\kappa-b}}{1-2^{a\kappa-b}} u^{-a}. \tag{5.0.2}$$

Proof. For $n \in \mathbb{N}$ define

$$\begin{aligned} D_n &:= \{(k_1, \dots, k_d) \cdot 2^{-n};\ k_1, \dots k_d \in \{1, \dots, 2^n\}\} \\ \xi_n(\omega) &:= \max\{\rho(Z_x(\omega), Z_y(\omega)) : x, y \in D_n, |x-y| = 2^{-n}\}. \end{aligned}$$

The $\xi_n, n \in \mathbb{N}$ are measurable since (E, ρ) is separable. Further,

$$|\{x, y \in D_n : |x-y| = 2^{-n}\}| \le d \cdot 2^{dn}.$$

Hence, for $\kappa \in (0, \frac{b}{a})$,

$$\begin{aligned} \mathbb{E}\left(\sum_{n=1}^{\infty}(2^{\kappa n}\xi_n)^a\right) &= \sum_{n=1}^{\infty} 2^{\kappa n a}\mathbb{E}(\xi_n^a) \le \sum_{n=1}^{\infty} 2^{\kappa n a}\mathbb{E}\left(\sum_{(x,y)\in D_n^2, |x-y|=2^{-n}} (\rho(Z_x(\omega), Z_y(\omega))^a\right) \\ &\le \sum_{n=1}^{\infty} 2^{\kappa n a} \cdot d \cdot 2^{dn} \cdot c \cdot 2^{-n(d+b)} = cd \sum_{n=1}^{\infty} 2^{-n(b-a\kappa)} = \frac{cd2^{a\kappa-b}}{1-2^{a\kappa-b}} < \infty. \end{aligned}$$

Hence, there exists $\Omega_0 \in \mathbb{F}$, $\mathbb{P}(\Omega_0) = 1$ such that

$$S(\omega) := \sup_{n\ge 1}\left\{2^{\kappa n}\xi_n(\omega)\right\} < \infty \quad \text{for all } \omega \in \Omega_0.$$

Further,

$$\mathbb{E}(S^a) \le \mathbb{E}\left(\sum_{n=1}^{\infty}(2^{\kappa n}\xi_n)^a\right) \le \frac{cd2^{a\kappa-b}}{1-2^{a\kappa-b}}.$$

Let $x, y \in \bigcup_{n=1}^{\infty} D_n$ be such that $|x-y|_\infty \le r < 2^{-m}$, where $m \in \mathbb{N}_0$. There exists a sequence

$$x = x_1, x_2 \dots, x_l = y$$

in $\bigcup_{n=m+1}^{\infty} D_n$, such that for each $i = 1, \dots l-1$ there exists $n(i) \geq m+1$ which satisfies $x_i, x_{i+1} \in D_{n(i)}$ and $|x_i - x_{i+1}| = 2^{-n(i)}$ and

$$|\{i \in \{1, \dots, l-1\} : n(i) = k\}| \leq 2d \quad \text{for all } k \geq m+1.$$

For $\omega \in \Omega_0$ and $0 < r < 1$ with $2^{-m-1} \leq r < 2^{-m}$, we get

$$\begin{aligned} & \sup\{\rho(Z_x(\omega), Z_y(\omega)); x, y \in \bigcup_{n=1}^{\infty} D_n, |x-y|_\infty \leq r\} \\ \leq\ & 2d \sum_{n=m+1}^{\infty} \xi_n(\omega) \leq 2dS(\omega) \sum_{n=m+1}^{\infty} 2^{-\kappa n} \\ =\ & 2^{-\kappa(m+1)} \frac{2d}{1-2^{-\kappa}} S(\omega) \leq \frac{2d}{1-2^{-\kappa}} S(\omega) r^\kappa, \end{aligned}$$

showing that $x \mapsto Z_x(\omega)$ is uniformly continuous on $\cup_{n=1}^\infty D_n$ for each $\omega \in \Omega_0$ and can therefore be extended to a continuous map $\bar{Z}(\omega) : \Theta \to E$ since (E, ρ) is complete. Clearly, $\bar{Z}$ satisfies (5.0.1) and – using Chebychev's inequality – (5.0.2). It remains to show that $\bar{Z}$ is a modification of Z. To see this, fix $x \in \Theta$ and let x_m be a sequence in $\cup_{n=1}^\infty D_n$ which converges to x. Then Z_{x_m} converges to Z_x in probability by assumption and to $\bar{Z}_x$ almost surely (namely on Ω_0) and hence also in probability. Since limits in probability are unique it follows that $Z_x = \bar{Z}_x$ almost surely, so $\bar{Z}$ is a modification of Z and the theorem is proved. □

Remark 5.3. Observe that Kolmogorov's theorem contains three statements: the existence of a continuous modification, an upper bound for the modulus of continuity and an upper bound on the tails of the diameter of the range of the continuous modification. Note that the theorem states that the continuous modification is automatically almost surely Hölder continuous with exponent κ. If the index set in Theorem is $\mathbb{R}^d$ instead of $[0,1]^d$, then the statement about the existence of a κ-Hölder modification remains unchanged but the quantitative estimates do not necessarily hold.

Bibliography

[Ba68] H. Bauer, *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie*, de Gruyter, Berlin, 1968.

[DM78] C. Dellacherie and P. A. Meyer, *Probabilities and Potential*, North-Holland, Amsterdam, 1978.

[DZ96] G. Da Prato and J. Zabczyk, *Ergodicity for Infinite Dimensional Systems*, Cambridge Univ. Press, Cambridge, 1996.

[Dy65] E. Dynkin, *Markov Processes*, Vol. I, Springer, Berlin, 1965.

[EK86] S. N. Ethier and T. G. Kurtz, *Markov Processes: Characterization and Convergence*, Wiley, New York, 1986.

[He79] H. Heyer, *Einführung in die Theorie Markoffscher Prozesse*, B.I. Verlag, Mannheim, 1979.

[Kh12] R. Khasminskii, *Stochastic Stability of Differential Equations*, second edition, Springer, New York, 2012.

[Kl06] A. Klenke, *Wahrscheinlichkeitstheorie*, Springer, New York, 2006.

[KS91] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*, second edition, Springer, New York, 1991.

[Ku90] H. Kunita, *Stochastic Flows and Stochastic Differential Equations*, Cambridge University Press, Cambridge, 1990.

[PR07] C. Prévôt and M. Röckner, *A Concise Course on Stochastic Partial Differential Equations*, Springer, New York, 2007.

[Pr92] P. Protter, *Stochastic Integration and Stochastic Differential Equations*, second edition, Springer, New York, 1992.

[Re13] M. v.Renesse, *Stochastische Differentialgleichungen*, Course Notes, Leipzig, http://www.math.uni-leipzig.de/~renesse, 2013.

[RY94] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, second edition, Springer, 1994.

[RW94] L.C.G. Rogers and D. Williams, *Diffusions, Markov Processes and Martingales*, Volume I: Foundations, second edition, Wiley, 1994.

[Sh88] M. Sharpe, *General Theory of Markov Processes*, Academic Press, Boston, 1988.

[St13] T. Sturm, *Stochastische Analysis*, Course Notes, Bonn, http://www-wt.iam.uni-bonn.de/~sturm/skript/stochana.pdf