MARKOV PROCESSES: THEORY AND EXAMPLES

JAN SWART AND ANITA WINTER

Date: March 1, 2005.

Contents

1. Stochastic processes

In this section we recall some basic definitions and facts on topologies and stochastic processes (Subsections 1.1 and 1.2). Subsection 1.3 is devoted to the study of the space of paths which are continuous from the right and have limits from the left. Finally, for sake of completeness, we collect facts on compactifications in Subsection 1.4. These will only find applications in later sections.

1.1. **Random variables.** Probability theory is the theory of random variables, i.e., quantities whose value is determined by chance. Mathematically speaking, a *random variable* is a measurable map $X : \Omega \to E$, where $(\Omega, \mathcal{F}, \mathbf{P})$ is a probability space and $(E, \mathcal{E})$ is a measurable space. The probability measure

$$\mathbf{P}_X = \mathcal{L}(X) := \mathbf{P} \circ X^{-1}$$

on $(E, \mathcal{E})$ is called the *law* of X and usually the only object that we are really interested in.[1] If $(X_t)_{t \in T}$ is a family of random variables, taking values in measurable spaces $(E_t, \mathcal{E}_t)_{t \in T}$, then we can view $(X_t)_{t \in T}$ as a single random variable, taking values in the product space $\prod_{t \in T} E_t$ equipped with the product-σ-field $\prod_{t \in T} \mathcal{E}_t$.[2] The law $\mathbf{P}_{(X_t)_{t \in T}} = \mathcal{L}((X_t)_{t \in T})$ of this random variable is called the *joint law* of the random variables $(X_t)_{t \in T}$.

In practise, we usually need a bit more structure on the spaces that our random variables take values in. For our purposes, it will be sufficient to consider random variables taking values in Polish spaces.

Recall that a *topology* on a space E is a collection $\mathcal{O}$ of subsets of E, called *open sets*, such that:

(1) $E, \emptyset \in \mathcal{O}$.
(2) $O_t \in \mathcal{O}\ \forall t \in T$ implies $\bigcup_{t \in T} O_t \in \mathcal{O}$.
(3) $O_1, O_2 \in \mathcal{O}$ implies $O_1 \cap O_2 \in \mathcal{O}$.

A topology is *metrizable* if there exists a metric d on E such that the open sets in this topology are the sets O with the property that $\forall x \in O\ \exists \varepsilon > 0$ s.t. $B_\varepsilon(x) \subset O$, where $B_\varepsilon(x) := \{y \in E : d(x, y) < \varepsilon\}$ is the open ball around x with radius ε. Two metrics are called *equivalent* if they define the same topology. Concepts such as convergence, continuity, and compactness depend only on the topology but completeness depends on the choice of the

[1] At this point, one may wonder why probabilists speak of a random variables at all and do not immediately focus on the probability measures that are their laws, if that is what they are really after. The reason is mainly a matter of convenient notation. If $\mu = \mathcal{L}(X)$ is the law of a real-valued random variable X, then what is the law of X^2? In terms of random variables, this is simply $\mathcal{L}(X^2)$. In terms of probability measures, this is the image of the probability measure μ under the map $x \mapsto x^2$, i.e., the measure $\mu \circ f^{-1}$ where $f : \mathbb{R} \to \mathbb{R}$ is defined as $f(x) = x^2$ –an unpleasantly long mouthful.

[2] Recall that $\prod_{t \in T} E_t := \{(x_t)_{t \in T} : x_t \in E_t\ \forall t \in T\}$. The *coordinate projections* $\pi_t : \prod_{t \in T} E_t := \{(x_t)_{t \in T} : x_t \in E_t\} \to E_t$ are defined by $\pi_t((x_s)_{s \in T}) := x_t$, $t \in T$. By definition, the product-σ-field $\prod_{t \in T} \mathcal{E}_t$ is the σ-field on $\prod_{t \in T} E_t$ that is generated by the coordinate projections, i.e., $\prod_{t \in T} \mathcal{E}_t := \sigma(\pi_t : t \in T) = \sigma(\{\pi_t^{-1}(A) : A \in \mathcal{E}_t\})$.

metric.[3] A topological space E is called *separable* if there exists a countable set $D \subset E$ such that D is dense in E.

By definition, a topological space $(E, \mathcal{O})$ is *Polish* if E is separable and there exists a complete metric defining the topology on E. We always equip Polish spaces with the *Borel-σ-field* $\mathcal{B}(E)$, which is the σ-field generated by the open sets.

The reason why we are interested in Polish spaces is that for random variables taking values in Polish spaces, certain useful results are true that do not hold in general, since they make use of the fact.

Lemma 1.1 (Probability measures on Polish spaces are tight)**.** *Each probability measure* $\mathbf{P}$ *on a Polish space* $(E, \mathcal{O})$ *is* tight*, i.e., for all* $\varepsilon > 0$ *there is a compact set* $K \subseteq E$ *such that* $\mathbf{P}(K) \geq 1 - \varepsilon$*.*

Proof. Let $(x_k)_{k \in \mathbb{N}}$ be dense in $(E, \mathcal{O})$, and let $\mathbf{P}$ be a probability measure on $(E, \mathcal{O})$. Given $\varepsilon > 0$ and a metric d on $(E, \mathcal{O})$, we can choose $N_1, N_2, \ldots$ such that

$$\mathbf{P}\big(\cup_{k=1}^{N_n} \{x' : d(x', x_k) < \frac{1}{n}\}\big) \geq 1 - \frac{\varepsilon}{2^n}. \tag{1.1}$$

Let K be the closure of $\bigcap_{n \geq 1} \bigcup_{k=1}^{N_n} \{x' : d(x', x_k) < \frac{1}{n}\}$. Then K is totally bounded[4], and hence compact, and we have $\mathbf{P}(K) \geq 1 - \varepsilon \sum_{n=1}^{\infty} 2^{-n} = 1 - \varepsilon$. ∎

For example, the following result states that provided the state space $(E, \mathcal{O})$ is Polish, for each projective family of probability measures there exists a projective limit.

Theorem 1.2 (Percy J. Daniell [Dan19], Andrei N. Kolmogorov [Kol33])**.** *Let* $(E_t)_{t \in T}$ *be (a possibly uncountable) collection of Polish spaces and let* μ_S *(*$S \subset T$ *finite) be probability measures on* $(E_t)_{t \in S}$ *such that*

$$\mu_{S'} \circ (\pi_S)^{-1} = \mu_S, \qquad S \subset S' \subset T,\ S, S' \text{ finite}, \tag{1.2}$$

where π_S *denotes the projection on* $(E_t)_{t \in S}$*. Then there exists a unique probability measure* μ_T *on* $\prod_{t \in T} E_t$*, equipped with the product* σ*-field, such that*

$$\mu_T \circ \pi_S^{-1} = \mu_S, \qquad S \subset T,\ S \text{ finite}. \tag{1.3}$$

Proof. For $E_t \equiv \mathbb{R}$ see e.g. Theorem 2.2.2 in [KS88]. ∎

---

[3] More precisely: completeness depends on the uniform structure defined by the metric. For the theory of uniform spaces, see for example [Kel55].

[4] Recall that a set A is totally bounded if for each $\varepsilon > 0$, A possesses a finite ε-net, where an ε-net for A is a collection of points $\{x_n\}$ with the property that for each $x \in A$ there is an x_k such that $d(x, x_k) < \varepsilon$.

A consequence of Kolmogorov's extension theorem is that if $\{\mu_S : S \subset T \text{ finite}\}$ are probability measures satisfying the consistency relation (1.2), then there exist random variables $(X_t)_{t\in T}$ defined on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$ such that $\mathcal{L}((X_t)_{t\in S}) = \mu_S$ for each finite $S \subset T$. (The canonical choice is $\Omega = \prod_{t\in T} E_t$.)

Exercise 1.3. *For $n \in \mathbb{N}$, $k_i \in \{0,1\}$, $i = 0,\dots,n$, and $0 =: t_0 < t_1 < \cdots < t_n$ in $[0,\infty)$, let $\tau_n := \inf\{l \geq 0 : k_l = 1\} \wedge (1+n)$, and*

$$\begin{aligned} &\mu^{t_1,\dots,t_n}(k_1,\dots,k_n) \\ &:= 1\{0 \leq k_1 \leq \cdots \leq k_n \leq 1\} \begin{cases} e^{-t_{\tau_n - 1}} - e^{-t_{\tau_n}}, & \text{if } \tau_n \leq n, \\ e^{-t_n}, & \text{if } \tau_n = 1+n. \end{cases} \end{aligned} \tag{1.4}$$

(i) *Show that the collection $\{\mu^{t_1,\dots,t_n}; 0 =: t_0 < t_1 < \cdots < t_n\}$ of probability measures on $\{\underline{k} \in \{0,1\}^n : 0 \leq k_1 \leq \cdots \leq k_n \leq 1\}$ satisfies the consistence condition (1.2).*

(ii) *Can you find one (or even more than one) $\{0,1\}$-valued stochastic process X with*

$$\mathbf{P}\{X_{t_1} = k_1, \dots, X_{t_n} = k_n\} = \mu^{t_1,\dots,t_n}(k_1,\dots,k_n), \quad 0 \leq k_1 \leq \cdots \leq k_n \leq 1. \tag{1.5}$$

1.2. **Stochastic processes.** A *stochastic process* with *index set* T and *state space* E is a collection of random variables $X = (X_t)_{t\in T}$ (defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$) with values in E. We will usually be interested in the case that $T = [0,\infty)$ and E is a Polish space. We interpret $X = (X_t)_{t\in[0,\infty)}$ as a quantity the value of which is determined by chance and that develops in time.

A stochastic process is called *measurable* if the map $(t,\omega) \mapsto X_t(\omega)$ from $[0,\infty) \times \Omega$ into E is measurable. The functions $t \mapsto X_t(\omega)$ (with $\omega \in \Omega$) are called the *sample paths* of the process X.

Lemma 1.4 (Right continuous sample paths)**.** *If X has right continuous sample paths then X is measurable.*

Proof. Define processes $X^{(n)}$ by $X^{(n)}_t := X_{\lfloor nt+1 \rfloor / n}$. Then, for each measurable set $A \subseteq E$, $\{(t,\omega) : X^{(n)}_t(\omega) \in A\} = \bigcup_{k=0}^{\infty} [k/n, (k+1)/n) \times X^{-1}_{k/n}(A)$, so $X^{(n)}$ is measurable for each $n \geq 1$. By the right-continuity of the sample paths, $X^{(n)} \underset{n\to\infty}{\longrightarrow} X$ pointwise, so X is measurable. ∎

By definition, the laws $\mathcal{L}(X_{t_1},\dots,X_{t_n})$ with $0 \leq t_1 < \cdots < t_n$ are called the *finite-dimensional distributions* of X. If X and Y are stochastic processes with the same finite dimensional distributions then we say that Y is a *version* of X (and vice versa). Here X and Y need not be defined on the same probability space. If X and Y are stochastic processes defined on the same probability space then we say that Y is a *modification* of X if $X_t = Y_t$ a.s.

$\forall t \geq 0$. Note that if Y is a modification of X, then X and Y have the same finite dimensional distributions. We say that X and Y are *indistinguishable* if $X_t = Y_t$ $\forall t \geq 0$ a.s.[5]

Example (Modification). Let $(\Omega, \mathcal{F}, \mathbf{P}) = ([0,1], \mathcal{B}[0,1], \ell)$ where ℓ is the Lebesgue measure. For a given $x \in (0, \infty)$, define $[0,1]$-valued stochastic processes X^x and Y by

$$X^x(t) := \begin{cases} t, & \text{if } t \neq x, \\ 0, & \text{if } t = x, \end{cases} \tag{1.6}$$

and

$$Y(t) := t, \quad t \in [0,1]. \tag{1.7}$$

Then Y is a modification of X^x but X and Y are not indistinguishable. ∎

Lemma 1.5 (Right continuous modifications)**.** *If Y is a modification of X and X and Y have right-continuous sample paths, then X and Y are indistinguishable.*

Proof. If Y is a modification of X then $X_t = Y_t$ $\forall t \in \mathbb{Q}$ a.s. By right-continuity of sample paths, this implies that $X_t = Y_t$ $\forall t \geq 0$ a.s. ∎

We will usually be interested in stochastic processes with sample paths that have right limits $X_{t+} := \lim_{s \downarrow t} X_s$ for each $t \geq 0$ and left limits $X_{t-} := \lim_{s \uparrow t} X_s$ for each $t > 0$. In practise nobody can measure time with infinite precision, so when we model a real process it is a matter of taste whether we assume that the sample paths are right or left continuous; it is tradition to assume that they are right continuous. (Lemmas 1.4 and 1.5 hold equally well for processes with left continuous sample paths.) Note that a consequence of this assumption is that the sample paths cannot have a jump at time $t = 0$; this will actually be convenient later on. In the next section we study the space of all paths that are right-continuous with left limits in more detail.

1.3. **Cadlag sample paths.** Let $(E, \mathcal{O})$ be a metrizable space. A function $w : [0, \infty) \to E$ such that w is right continuous and $w(t-)$ exists for each $t > 0$ is called a *cadlag* function (from the French "<u>c</u>ontinu <u>à</u> <u>d</u>roit <u>l</u>imite <u>à</u> <u>g</u>auche"). The space of all such functions is denoted by

$$\begin{aligned} &\mathcal{D}_E[0,\infty) := \\ &\quad \big\{ w : [0,\infty) \to E : w(t) = w(t+) \ \forall t \geq 0, \ w(t-) \text{ exists } \forall t > 0 \big\}. \end{aligned} \tag{1.8}$$

[5] Note the order of the statements: If Y is a modification of X, then there is for each $t \geq 0$ a measurable set $\Omega^*_t \subset \Omega$ with $\mathbf{P}(\Omega^*) = 1$ such that $X_t(\omega) = Y_t(\omega)$ for all $\omega \in \Omega^*_t$. If X and Y are indistiguishable, then there exists a measurable set Ω^* (independent of t) such that $X_t(\omega) = Y_t(\omega)$ for all $\omega \in \Omega^*$.

We begin by observing that functions in $\mathcal{D}_E[0,\infty)$ are better behaved than one might suspect.

Lemma 1.6 (Only countably many jumps)**.** *If $w \in \mathcal{D}_E[0,\infty)$, then w has at most countably many points of discontinuity.*

Proof. For $n = 1, 2, \ldots$, and d a metric on $(E, \mathcal{O})$, let

$$A_n := \big\{t > 0 : d\big(w(t), w(t-)\big) > \frac{1}{n}\big\}. \tag{1.9}$$

Since w has limits from the right and the left, A_n can not possess cluster points. Hence A_n is countable for all $n = 1, 2, \ldots$, and the set of all discontinuities of $\cup_{n\geq 1} A_n$ is countable too. ∎

In order to be in a position to do probability on spaces of random variables with values in $\mathcal{D}_E[0,\infty)$ we want to equip $\mathcal{D}_E[0,\infty)$ with a topology so that in this topology $\mathcal{D}_E[0,\infty)$ is Polish. We wil see that this is possible provided that $(E, \mathcal{O})$ is Polish.

To motivate the topology that we will choose, we first take a look at the space

$$\mathcal{C}_E[0,\infty) := \big\{\text{continuous functions } w : [0,\infty) \to E\big\}. \tag{1.10}$$

Lemma 1.7 (Uniform convergence on compacta)**.** *Let (E, d) be metric space. Then the following conditions on functions $w_n, w \in \mathcal{C}_E[0,\infty)$ are equivalent.*

(a) *For all $T > 0$,*

$$\lim_{n\to\infty} \sup_{t\in[0,T]} d(w_n(t), w(t)) = 0. \tag{1.11}$$

(b) *For all $(t_n)_{n\in\mathbb{N}}, t \in [0,\infty)$ such that $t_n \underset{n\to\infty}{\longrightarrow} t$,*

$$\lim_{n\to\infty} w_n(t_n) = w(t). \tag{1.12}$$

Proof. (a)⇒(b). If $t_n \to t$ then there is a $T > 0$ such that $t_n, t \leq T$ for all n. Now

$$\begin{aligned} d\big(w_n(t_n), w(t)\big) &\leq d\big(w(t_n), w(t)\big) + d\big(w_n(t_n), w(t_n)\big) \\ &\leq d\big(w(t_n), w(t)\big) + \sup_{s\in[0,T]} d\big(w_n(s), w(s)\big) \underset{n\to\infty}{\longrightarrow} 0. \end{aligned} \tag{1.13}$$

by (a) and the continuity of w.

(b)⇒(a). Imagine that there exists a $T > 0$ such that

$$\limsup_{n\to\infty} \sup_{t\in[0,T]} d(w_n(t), w(t)) = \varepsilon > 0. \tag{1.14}$$

Then we can choose $s_n \in [0,T]$ such that $\limsup_{n\to\infty} d(w_n(s_n), w(s_n)) = \varepsilon$. By the compactness of $[0,T]$ we can choose $n_1 < n_2 < \cdots$ such that $\lim_{m\to\infty} s_{n_m} = t$ for some $t \in [0,T]$ and $d(w_{n_m}(s_{n_m}), w(s_{n_m})) \geq \frac{\varepsilon}{2}$ for each

m. Hence, $d\big(w_{n_m}(s_{n_m}), w(t)\big) + d\big(w(s_{n_m}), w(t)\big) \geq d(w_{n_m}(s_{n_m}), w(s_{n_m})) \geq \frac{\varepsilon}{2}$. By continuity, $d(w(s_{n_m}), w(t)) \underset{m\to\infty}{\longrightarrow} 0$. We therefore find that

$$\limsup_{n\to\infty} d(w_n(s_n), w(t)) \geq \frac{\varepsilon}{2} \tag{1.15}$$

which contradicts (1.12). ∎

If w_n, w are as in Lemma 1.7 then because of Property (a) we say that w_n converges to w *uniformly on compacta.* Property (b) shows that this definition does not depend on the choice of the metric on E, i.e., if d and $\tilde{d}$ are equivalent metrics on E then $w_n \to w$ uniformly on compacta w.r.t. d if and only if $w_n \to w$ uniformly on compacta w.r.t. $\tilde{d}$. The topology on $\mathcal{C}_E[0,\infty)$ of uniform convergence on compacta is metrizable. A possible choice of a metric on $\mathcal{C}_E[0,\infty)$ generating the topology of uniform convergence on compacta is for example:

$$d_{\text{u.c.}}(w_1, w_2) := \int_0^\infty \mathrm{d}s\, e^{-s} \sup_{t\in[0,\infty)} 1 \wedge d\big(w_1(t\wedge s), w_2(t\wedge s)\big) \tag{1.16}$$

Remark. If d is a metric on $(E, \mathcal{O})$, then $1 \wedge d$ is also a metric, and both metrics are equivalent. ∎

On $\mathcal{D}_E[0,\infty)$, we could also define uniform convergence on compacta as in Lemma 1.7, Property (a), but this topology would be too strong for our purposes. For example, if $E = \mathbb{R}$, we would like the functions $w_n := 1_{[1+\frac{1}{n},\infty)}$ to approximate the function $w := 1_{[1,\infty)}$ as $n \to \infty$, but $\sup_{t\in[0,2]} |w_n(t) - w(t)| = 1$ for each n. We wish to find a topology on $\mathcal{D}_E[0,\infty)$ such that $w_n \to w$ whenever the jump times of the functions w_n converge to the jump times of w while the "rest" of the paths converge uniformly in compacta. The main result of this section is that such a topology exists and has nice properties.

Theorem 1.8 (Skorohod topology)**.** *Let (E, d) be a metric space. Then there exists a metric $d^{\mathrm{d}}_{\mathrm{Sk}}$ on $\mathcal{D}_E[0,\infty)$ such that in this metric, $\mathcal{D}_E[0,\infty)$ is separable if E is separable, $\mathcal{D}_E[0,\infty)$ is complete if E is complete, and $w_n \to w$ if and only if for all $T \in [0,\infty)$ there exists a sequence λ_n of strictly increasing, continuous functions $\lambda_n : [0,T] \to [0,\infty)$ with $\lambda_n(0) = 0$, such that*

$$\lim_{n\to\infty} \sup_{t\in[0,T]} |\lambda_n(t) - t| = 0 \tag{1.17}$$

and for $(t_n)_{n\in\mathbb{N}}, t \in [0,T]$,

$$\lim_{n\to\infty} w_n(\lambda_n(t_n)) = \begin{cases} w(t), & \textit{whenever } (t_n) \downarrow t, \\ w(t-), & \textit{whenever } (t_n) \uparrow t. \end{cases} \tag{1.18}$$

Remark. The idea of the functions λ_n in (1.17) and (1.18) is to make two functions $w, \tilde{w}$ close in the topology on $\mathcal{D}_E[0,\infty)$ if a small deformation of the time scale makes them near in the uniform topology. The topology in Theorem 1.8 is called the Skorohod topology, after its inventor. ∎

Our proof of Theorem 1.8 will follow Section 3.5 in [EK86]. Let Λ' be the collection of strictly increasing functions λ mapping $[0,\infty)$ onto $[0,\infty)$. In particular, for all $\lambda \in \Lambda'$ we have $\lambda(0) = 0$, $\lim_{t\to\infty} \lambda(t) = \infty$, and λ is continuous. Furthermore, let Λ be the subclass of Lipschitz continuous functions $\lambda \in \Lambda'$ such that

$$\| \lambda \| := \sup_{0\le s<t} \left| \log \frac{\lambda(t) - \lambda(s)}{t - s} \right| < \infty. \tag{1.19}$$

In the literature $\| \lambda \|$ is refered to as the *dilatation* of $\lambda \in \Lambda$.

Lemma 1.9 (Properties of a dilatation)**.** *The dilatation $\| \cdot \| \colon \Lambda \to \mathbb{R}^+$ has the following properties:*

(i) *For all $\lambda \in \Lambda$,*

$$\| \lambda \| = \| \lambda^{-1} \|, \tag{1.20}$$

where λ^{-1} denotes the inverse function of λ, i.e., $\lambda^{-1}(\lambda(t)) = t$ for all $t \ge 0$.

(ii) *If $\lambda_1, \lambda_2 \in \Lambda$, then $\lambda_1 \circ \lambda_2 \in \Lambda$, and we have*

$$\| \lambda_1 \circ \lambda_2 \| \le \| \lambda_1 \| + \| \lambda_2 \| . \tag{1.21}$$

(iii) *If $(\lambda_n)_{n\in\mathbb{N}}$ is a sequence in Λ with $\| \lambda_n \| \underset{n\to\infty}{\longrightarrow} 0$, then for all $T \in [0,\infty)$,*

$$\lim_{n\to\infty} \sup_{t\in[0,T]} |\lambda_n(t) - t| = 0. \tag{1.22}$$

Proof of Lemma 1.9. (i) For all $\lambda \in \Lambda$,

$$\begin{aligned} \| \lambda \| &= \sup_{0\le s<t} \Big| \log \frac{\lambda(t) - \lambda(s)}{t - s} \Big| \\ &= \sup_{0\le s':=\lambda(s)<t':=\lambda(t)} \Big| -\log \frac{\lambda^{-1}(t') - \lambda^{-1}(s')}{t' - s'} \Big| = \| \lambda^{-1} \| . \end{aligned} \tag{1.23}$$

(ii) For $\lambda_1, \lambda_2 \in \Lambda$, $\lambda_1 \circ \lambda_2$ is also Lipschitz continuous, and

$$\begin{aligned} \| \lambda_1 \circ \lambda_2 \| &= \sup_{0\le s<t} \left| \log \frac{\lambda_1 \circ \lambda_2(t) - \lambda_1 \circ \lambda_2(s)}{t - s} \right| \\ &= \sup_{0\le s<t} \left| \log \frac{\lambda_1 \circ \lambda_2(t) - \lambda_1 \circ \lambda_2(s)}{\lambda_2(t) - \lambda_2(s)} + \log \frac{\lambda_2(t) - \lambda_2(s)}{t - s} \right| \\ &\le \| \lambda_1 \| + \| \lambda_2 \| . \end{aligned} \tag{1.24}$$

In particular, $\lambda_1 \circ \lambda_2 \in \Lambda$.

(iii) Since for all $\lambda \in \Lambda$,

$$\begin{aligned} \| \lambda \| &\geq 1 - e^{-\|\lambda\|} \\ &= \sup_{0 \leq s < t} \Big(1 - e^{-|\log \frac{\lambda(t)-\lambda(s)}{t-s}|} \Big) \\ &= \sup_{0 \leq s < t} \Big(1 - \frac{\lambda(t)-\lambda(s)}{t-s} \wedge \frac{t-s}{\lambda(t)-\lambda(s)} \Big), \end{aligned} \tag{1.25}$$

$\| \lambda_n \| \underset{n\to\infty}{\longrightarrow} 0$ implies that $\sup_{0 \leq s < t} |\frac{\lambda_n(t)-\lambda_n(s)}{t-s}| \underset{n\to\infty}{\longrightarrow} 1$. In particular, for all $T \geq 0$,

$$\lim_{n\to\infty} \sup_{t \in [0,T]} |\lambda_n(t) - t| = 0. \tag{1.26}$$

∎

(Counter-)Example. For $n \in \mathbb{N}$, let

$$\lambda_n(t) := \begin{cases} \frac{n(n-2)}{n^2-2} t, & \text{if } t \in [0, \frac{1}{2} - \frac{1}{n^2}], \\ nt + \frac{1}{2}(1-n), & \text{if } t \in [\frac{1}{2} - \frac{1}{n^2}, \frac{1}{2} + \frac{1}{n^2}], \\ \frac{n(n-2)}{n^2-2} t - \frac{2(n-1)}{n^2-2}, & t \in [\frac{1}{2} + \frac{1}{n^2}, 1]. \end{cases} \tag{1.27}$$

Then $(\lambda_n)_{n \in \mathbb{N}}$ in Λ, satisfies (1.22) but $\| \lambda_n \| = n \underset{n\to\infty}{\longrightarrow} \infty$. ∎

In analogy with (1.16), for $v, w \in \mathcal{D}_E[0,\infty)$, we define the *Skorohod metric* by

$$\begin{aligned} &d^d_{\mathrm{Sk}}(v,w) \\ &:= \inf_{\lambda \in \Lambda} \Big\{ \| \lambda \| \vee \int_{[0,\infty)} \mathrm{d}s\, e^{-s} \sup_{t \in [0,\infty)} 1 \wedge d\Big(v\big(t \wedge s\big), w\big(\lambda(t) \wedge s\big)\Big) \Big\}. \end{aligned} \tag{1.28}$$

The next lemma states that d^d_{Sk} is indeed a metric on $\mathcal{D}_E[0,\infty)$.

Lemma 1.10. $\big(\mathcal{D}_E[0,\infty), d^d_{\mathrm{Sk}}\big)$ *is a metric space.*

Proof. For *symmetry* recall Part (i) of Lemma 1.9, and notice that

$$\begin{aligned} &\sup_{t \in [0,\infty)} \Big\{ 1 \wedge d\Big(v\big(t \wedge s\big), w\big(\lambda(t) \wedge s\big)\Big) \Big\} \\ &\qquad = \sup_{t \in [0,\infty)} 1 \wedge d\Big(v\big(\lambda^{-1}(t) \wedge s\big), w\big(t \wedge s\big)\Big), \end{aligned} \tag{1.29}$$

for all $\lambda \in \Lambda$. This implies that $d_{\mathrm{Sk}}(v,w) = d_{\mathrm{Sk}}(w,v)$ for all $v, w \in \mathcal{D}_E[0,\infty)$.

If $d_{\mathrm{Sk}}(v,w) = 0$, then there exists a sequence $(\lambda_n)_{n \in \mathbb{N}}$ in Λ such that $\| \lambda_n \| \underset{n\to\infty}{\longrightarrow} 0$ and

$$\ell\{s \in [0, s_0] : \sup_{t \in [0,\infty)} 1 \wedge d\big(v(t \wedge s), w(\lambda_n(t) \wedge s)\big) \geq \varepsilon\} \underset{n\to\infty}{\longrightarrow} 0 \tag{1.30}$$

for all $\varepsilon > 0$ and $s_0 \in [0,\infty)$. Hence by Part (iii) of Lemma 1.9 and (1.30), $v(t) = w(t)$ for all continuity points t of w, and therefore by Lemma 1.6 and right continuity of v and w, $v = w$.

It remains to show the *triangle inequality.* Recall Part (ii) of Lemma 1.9, and notice that

for all $t \in [0,\infty)$,

$$\begin{aligned}
&\sup_{s\in[0,\infty)} \Big\{1 \wedge d\Big(w\big(t\wedge s\big), u\big(t\wedge\ \lambda_1\circ\lambda_2(s)\big)\Big)\Big\} \\
&\quad\le \sup_{s\in[0,\infty)} \Big\{1 \wedge d\Big(w\big(t\wedge s\big), v\big(t\wedge\ \lambda_2(s)\big)\Big)\Big\} \\
&\qquad + \sup_{s\in[0,\infty)} \Big\{1 \wedge d\Big(v\big(t\wedge \lambda_2(s)\big), u\big(t\wedge\ \lambda_1\circ\lambda_2(s)\big)\Big)\Big\} \\
&\quad= \sup_{s\in[0,\infty)} \Big\{1 \wedge d\Big(w\big(t\wedge s\big), v\big(t\wedge\ \lambda_2(s)\big)\Big)\Big\} \\
&\qquad + \sup_{s\in[0,\infty)} \Big\{1 \wedge d\Big(v\big(t\wedge s\big), u\big(t\wedge\ \lambda_1(s)\big)\Big)\Big\}.
\end{aligned} \tag{1.31}$$

Combining (1.24) and (1.31) implies that $d_{\mathrm{Sk}}(w,u) \le d_{\mathrm{Sk}}(w,v) + d_{\mathrm{Sk}}(v,u)$. ∎

Exercise 1.11. *For $n \in \mathbb{N}$, let $v_n := 1_{[0,1-2^{-n})}$ and $w_n := 1_{[0,2^{-n})}$. Decide whether the sequences $(v_n)_{n\in\mathbb{N}}$ and $(w_n)_{n\in\mathbb{N}}$ converge in $\big(\mathcal{D}_E[0,\infty), d^d_{\mathrm{Sk}}\big)$ and, if so, determine the limit function.*

Proposition 1.12 (A convergence criterion)**.** *Let $(w_n) \in \mathcal{D}_E[0,\infty)$ and $w \in \mathcal{D}_E[0,\infty)$. Then the following are equivalent:*

(a) $d^d_{\mathrm{Sk}}(w_n, w) \underset{n\to\infty}{\longrightarrow} 0$.

(b) *There exists a sequence $(\lambda_n)_{n\in\mathbb{N}} \in \Lambda$ such that $\|\ \lambda_n\ \| \underset{n\to\infty}{\longrightarrow} 0$ and*

$$\lim_{n\to\infty} \sup_{t\in[0,T]} d\big(w_n(\lambda_n(t)), w(t)\big) = 0, \tag{1.32}$$

for all $T \in [0,\infty)$.

(c) *For each $T > 0$, there exists a sequence $(\lambda_n)_{n\in\mathbb{N}}$ in Λ' (possibly depending on T) satisfying (1.26) and (1.32).*

(d) *For each $T > 0$, there exists a sequence $(\lambda_n)_{n\in\mathbb{N}}$ in Λ' (possibly depending on T) satisfying (1.17) and (1.18).*

Corollary 1.13. *The Skorohod topology does not depend on the choice of the metric on $(E,\mathcal{O})$.*

Proof of Corollary 1.13. If $d, \tilde{d}$ are two equivalent metrics on $(E, \mathcal{O})$ and d^d_{Sk} and $d^{\tilde{d}}_{\mathrm{Sk}}$ are the associated Skorohod metrics, then formula (1.18) shows that $w_n \underset{n\to\infty}{\longrightarrow} w$ in d_{Sk} if and only if $w_n \underset{n\to\infty}{\longrightarrow} w$ in $d^{\tilde{d}}_{\mathrm{Sk}}$. It is easy to see that two metrics are equivalent if every sequence that converges in one metric also converges in the other metric, and vice versa.[6] ∎

Proof of Proposition 1.12. (a)$\Longleftrightarrow$(b). We start showing that (a) is equivalent to (b). Assume first that $d^d_{\mathrm{Sk}}(w_n, w) \underset{n\to\infty}{\longrightarrow} 0$ for a metric d on $(E, \mathcal{O})$. By definition, then there exist sequences $(\lambda_n)_{n\in\mathbb{N}}$ in Λ such that $\| \lambda_n \| \underset{n\to\infty}{\longrightarrow} 0$ and

$$\ell\{s \in [0, s_0] : \sup_{t\in[0,\infty)} 1 \wedge d\big(w_n(\lambda_n(t) \wedge s), w(t \wedge s)\big) \geq \varepsilon\} \underset{n\to\infty}{\longrightarrow} 0 \tag{1.33}$$

for all $\varepsilon > 0$ and $s_0 \in [0, \infty)$.

Hence, there is a subsequence $(n_k)_{k\in\mathbb{N}}$ such that $d\big(w_{n_k}(\lambda_{n_k}(t) \wedge s), w(t \wedge s)\big) \underset{k\to\infty}{\longrightarrow} 0$ for almost every $s \in [0, \infty)$, and thus for all continuity points s of w. That is, there exist sequences $(\lambda_n)_{n\in\mathbb{N}}$ in Λ and $(s_n)_{n\in\mathbb{N}} \uparrow \infty$ in $[0, \infty)$ such that $\| \lambda_n \| \underset{n\to\infty}{\longrightarrow} 0$ and

$$\lim_{n\to\infty} \sup_{t\geq 0} d\big(w_n(\lambda_n(t) \wedge s_n), w(t \wedge s_n)\big) = 0. \tag{1.34}$$

Now for given $T \in [0, \infty)$, $s_n \geq T \vee \lambda_n(T)$ for all n sufficiently large. Therefore (1.34) implies (1.32).

On the other hand, let a sequence $(\lambda_n)_{n\in\mathbb{N}}$ in Λ satisfy the condition of (b). Let $s \in [0, \infty)$. Then for each $n \in \mathbb{N}$,

$$\begin{aligned}
&\sup_{t\geq 0} d\big(w_n(\lambda_n(t) \wedge s), w(t \wedge s)\big) \\
&\quad= \sup_{t'_n := \lambda_n(t) \geq 0} d\big(w_n(t'_n \wedge s), w(\lambda_n^{-1}(t'_n) \wedge s)\big) \\
&\quad\leq \sup_{t'_n \geq 0} d\big(w_n(t'_n \wedge s), w(\lambda_n^{-1}(t'_n \wedge s))\big) \\
&\qquad+ \sup_{t'_n \geq 0} d\big(w(\lambda_n^{-1}(t'_n \wedge s)), w(\lambda_n^{-1}(t'_n) \wedge s)\big).
\end{aligned} \tag{1.35}$$

[6] To see this, note that a set A is closed in the topology generated by a metric d if and only if $x \in A$ for all $x_n \in A$ with $x_n \to x$ in d. This shows that two metrics which define the same form of convergence have the same closed sets. Since open sets are the complements of closed sets, they also have the same open sets, i.e., they generate the same topology.

We can estimate this further by

$$\leq \sup_{r:=\lambda_n^{-1}(t'_n\wedge s)\in[0,\lambda_n^{-1}(s)]} d\big(w_n(\lambda_n(r)), w(r)\big)$$
$$+ \sup_{r:=\lambda_n^{-1}(t'_n)\in[s,\lambda_n^{-1}(s)\vee s]} d\big(w(r), w(s)\big)\vee$$
$$\sup_{r:=\lambda_n^{-1}(t'_n)\wedge s\in[\lambda_n^{-1}(s)\wedge s,s]} d\big(w(\lambda_n^{-1}(s)), w(r)\big),$$

where the second half of the last inequality follows by considering the cases $t'_n \leq s$ and $t'_n > s$ separately. Thus by (1.32),

$$\lim_{n\to\infty}\sup_{t\in[0,\infty)}\big\{1\wedge d\big(w_n(\lambda_n(t)\wedge s), w(t\wedge s)\big)\big\} = 0 \tag{1.36}$$

for every continuity point s of w. Hence, applying the dominated convergence theorem in (1.28) yields that $d^d_{\mathrm{Sk}}\big(w_n, w\big) \underset{n\to\infty}{\longrightarrow} 0$.

(b)$\Longleftrightarrow$(c). Obviously, assumption (c) is weaker than (b) (recall also (1.26)). To see the other direction, let N be a positive integer, and $(\lambda_n^N)_{n\in\mathbb{N}}$ in Λ' satisfying (1.26) with $T = N$ and such that

$$\lambda_n^N(t) := \lambda_n^N(N) + t - N, \quad t \geq N. \tag{1.37}$$

We want to construct a sequence $(\widehat{\lambda}_n)_{n\in\mathbb{N}}$ in Λ such that

- $\parallel \widehat{\lambda}_n \parallel \underset{n\to\infty}{\longrightarrow} 0$, and
- $\sup_{t\in[0,T]} d\big(w_n(\widehat{\lambda}_n(t)), w(t)\big) \underset{n\to\infty}{\longrightarrow} 0$, for all $T \in [0,\infty)$.

Notice that by (1.32) we can find a subsequence $(n_k)_{k\in\mathbb{N}}$ such that

$$\sup_{t\in[0,N]} d\big(w_n(\lambda_n^N(t)), w(t)\big) \leq \frac{1}{N} \tag{1.38}$$

for all $n \geq n_N$, while in general, we can not conclude from (1.26) that $\liminf_{n\to\infty} \parallel \lambda_n^N \parallel = 0$ (recall the counterexample given behind the proof of Lemma 1.9).

We proceed as follows. First we are therefore going to construct a sequence $(\widehat{\lambda_n^N})_{n\in\mathbb{N}}$ in Λ whose Lipschitz constant goes to 1 as $N \to \infty$ and which is obtained by disturbing $(\lambda_n^N)_{n\in\mathbb{N}}$ such that the dilatation of λ_n^N converges to zero as $N \to \infty$ but mildly enough that we can ensure that $\sup_{t\in[0,N]} d\big(w_n(\widehat{\lambda_n^N}(t)), w_n(\lambda_n(t))\big) \underset{N\to\infty}{\longrightarrow} 0$.

For that, define $\tau_0^N := 0$, and for all $k \geq 1$,

$$\tau_k^N := \begin{cases} \inf\{t > \tau_{k-1}^N : d\big(w(t), w(\tau_{k-1}^N)\big) > \frac{1}{N}\}, & \text{if } \tau_{k-1}^N < \infty, \\ \infty, & \text{if} \tau_{k-1}^N = \infty. \end{cases} \tag{1.39}$$

Since w is right continuous, the sequence $(\tau_k^N)_{k\in\mathbb{N}}$ is strictly increasing as long as its terms remain finite. Since w has limits from the left, the sequence has no cluster point. Now let for each $n \in \mathbb{N}$,

$$s_{k,n}^N := (\lambda_n^N)^{-1}(\tau_k^N), \tag{1.40}$$

where by convention $(\lambda_n^N)^{-1}(\infty)=\infty$.

Define a sequence $(\widehat{\lambda_n^N})_{n\in\mathbb{N}}$ in Λ by

$$(1.41)\quad \widehat{\lambda_n^N}(t):=\begin{cases}\tau_k^N+\frac{\tau_{k+1}^N-\tau_k^N}{s_{k+1,n}^N-s_{k,n}^N}(t-s_{k,n}^N), & \text{if} \quad t\in[s_{k,n}^N,s_{k+1,n}^N\wedge N),\\ \widehat{\lambda_n^N}(N)+t-N, & \text{if} \quad t\in(N,\infty),\\ \text{arbitrary}, & \text{otherwise},\end{cases}$$

where, by convention $\infty^{-1}\infty=1$. With this convention and by (1.26),

$$(1.42)\qquad \|\,\widehat{\lambda_n^N}\,\|=\max_{s_{k,n}^N\le N}\big|\log\big(\tau_{k+1}^N-\tau_k^N\big)-\log\big(s_{k+1,n}^N-s_{k,n}^N\big)\big|\underset{n\to\infty}{\longrightarrow}0,$$

and

$$(1.43)\qquad \sup_{t\in[0,N]}d\big(w_n(\widehat{\lambda_n^N}(t)),w_n(\lambda_n(t))\big)\le\frac{2}{N}.$$

Since

(1.44)

$$\begin{aligned}&\sup_{t\in[0,N]}d\big(w_n(\widehat{\lambda_n^N}(t)),w(t)\big)\\ &\qquad\le\sup_{t\in[0,N]}d\big(w_n(\lambda_n^N(t)),w(t)\big)+\sup_{t\in[0,N]}d\big(w_n(\widehat{\lambda_n^N}(t)),w_n(\lambda_n(t))\big)\\ &\qquad\le\sup_{t\in[0,N]}d\big(w_n(\lambda_n^N(t)),w(t)\big)+\frac{2}{N},\end{aligned}$$

for all $n\in\mathbb{N}$, (1.18) implies that we can choose a subsequence $(n_k)_{k\in\mathbb{N}}$ such that $\|\,\widehat{\lambda_n^N}\,\|\le\frac{1}{N}$ and $\sup_{t\in[0,N]}d\big(w_n(\widehat{\lambda_n^N}(t)),w(t)\big)\le\frac{3}{N}$ for all $n\ge n_N$. For $1\le n<n_1$, let $\widetilde{\lambda_n}$ be arbitrary. For $n_N\le n<n_{N+1}$, $N\ge1$, let $\widetilde{\lambda_n}:=\widehat{\lambda_n^N}$. Then the sequence $(\widetilde{\lambda_n})_{n\in\mathbb{N}}$ satisfies the conditions of (b).

(c)$\Longleftrightarrow$(d). To finish the proof we must show that (c) is equivalent to (d). Fix $T>0$ and $\lambda_n\in\Lambda'$ satisfying (1.26). Define $\tilde{w}_n(t):=w_n(\lambda_n(t))$ $(t\in[0,T])$. Then we must show that the following conditions are equivalent

(i) $\lim_{n\to\infty}\sup_{t\in[0,T]}d(\tilde{w}_n(t),w(t))=0$.

(ii) $\lim_{n\to\infty}\tilde{w}_n(t_n)=\begin{cases}w(t) & \text{whenever} \quad t_n\downarrow t\\ w(t-) & \text{whenever} \quad t_n\uparrow t\end{cases},\quad t_n,t\in[0,T].$

This is very similar to the proof of Lemma 1.7, with w_n replaced by $\tilde{w}_n$. The implication (i)$\Rightarrow$(ii) can be proved as in (1.13) using the facts that $w(t_n)\to w(t)$ if $t_n\downarrow t$ and $w(t_n)\to w(t-)$ if $t_n\uparrow t$. To prove the implication (ii)$\Rightarrow$(i) we assume that (i) does not hold and show that there exist $n_1<n_2<\cdots$ such that $\lim_{m\to\infty}s_{n_m}=t$ for some $t\in[0,T]$ and $d(\tilde{w}_{n_m}(s_{n_m}),w(s_{n_m}))\ge\frac{\varepsilon}{2}$ for each m. Since either $s_{n_m}>t$ infinitely often, or $s_{n_m}<t$ infinitely often, or $s_{n_m}=t$ infinitely often, by going to a further subsequence we can assume that either $s_{n_m}\downarrow t$ or $s_{n_m}\uparrow t$. Now the proof proceeds as before. ∎

We next state that if the underlying space $(E, \mathcal{O})$ is Polish then $\mathcal{D}_E[0,\infty)$ is Polish.

Proposition 1.14 (Andrei N. Kolmogorov [Kol56])**.** *If $(E, \mathcal{O})$ is separable, then $\big(\mathcal{D}_E[0,\infty), d^d_{\mathrm{Sk}}\big)$ is separable. If (E,d) is complete, then $\big(\mathcal{D}_E[0,\infty), d^d_{\mathrm{Sk}}\big)$ is complete.*

Remark. If E is Polish then $\mathcal{D}_E[0,\infty)$ is separable and we can choose d such that E is complete in d, hence $\mathcal{D}_E[0,\infty)$ is complete in d_{Sk}, hence $\mathcal{D}_E[0,\infty)$ is Polish. ∎

We prepare the proof with the following problem:

Exercise 1.15. *Let $(E, \mathcal{O})$ be a separable topological space, and $(\alpha_n)_{n\in\mathbb{N}}$ a countable dense subset of E. Show that the collection Γ of all functions of the form*

$$w(t) := \begin{cases} \alpha_{n_k}, & t \in [t_{k-1}, t_k), \\ \alpha_{n_K}, & t \in [t_K, \infty), \end{cases} \tag{1.45}$$

where $0 = t_0 < t_1 < \cdots < t_K$ are rational numbers, $K \geq 1$, and $n_1, \ldots, n_K \in \mathbb{N}$, is dense in $\big(\mathcal{D}_E[0,\infty), d^d_{\mathrm{Sk}}\big)$.

Proof of Proposition 1.14. Separability is covered by Exercise 1.15.

To prove *completeness*, it is enough to show that every Cauchy sequence has a subsequential limit. If $(w_n)_{n\in\mathbb{N}}$ is Cauchy, then for all $k \in \mathbb{N}$ there exists a N_k such that for all $m, n \geq N_k$,

$$d^d_{\mathrm{Sk}}\big(w_m, w_n\big) \leq 2^{-(k+1)} e^{-k}. \tag{1.46}$$

That is, for $k \geq 1$, we can choose $\lambda_k \in \Lambda$ and $s_k > k$ such that

$$\parallel \lambda_k \parallel \vee \sup_{t\in[0,\infty)} 1 \wedge d\big(w_{N_k}(\lambda_k(t) \wedge s_k), w_{N_{k+1}}(t \wedge s_k)\big) \leq 2^{-k}. \tag{1.47}$$

Let then

$$\mu_k := \lim_{n\to\infty} \lambda_{k+n} \circ \cdots \circ \lambda_{k+1} \circ \lambda_k, \tag{1.48}$$

and notice that μ_k exists uniformly on bounded intervals, is Lipschitz continuous and satisfies

$$\parallel \mu_k \parallel \leq \sum_{l=k}^{\infty} \parallel \lambda_l \parallel \leq 2^{-k+1}, \tag{1.49}$$

and hence, in particular, belongs to Λ. Since by (1.47), for all $k \geq 1$,

$$\begin{aligned}
&\sup_{t\in[0,\infty)} 1 \wedge d\big(w_{N_k}(\mu_k^{-1}(t) \wedge s_k), w_{N_{k+1}}(\mu_{k+1}^{-1}(t) \wedge s_k)\big) \\
&\quad= \sup_{t\in[0,\infty)} 1 \wedge d\big(w_{N_k}(\mu_k^{-1}(t) \wedge s_k), w_{N_{k+1}}(\lambda_k(\mu_k^{-1})(t) \wedge s_k)\big) \\
&\quad= \sup_{t\in[0,\infty)} 1 \wedge d\big(w_{N_k}(t \wedge s_k), w_{N_{k+1}}(\lambda_k(t) \wedge s_k)\big) \\
&\quad\leq 2^{-k},
\end{aligned} \tag{1.50}$$

completeness of E implies that $u_k := w_{N_k} \circ \mu_k^{-1}$ converges uniformly on bounded intervals to a function $w : [0,\infty) \to E$. Moreover, since $u_k \in \mathcal{D}_E[0,\infty)$, for all $k \geq 1$, also $w \in \mathcal{D}_E[0,\infty)$. Therefore $(w_{N_k})_{k\in\mathbb{N}}$ and w satisfy the conditions of part (b) of Proposition 1.12, and hence we conclude that $d^d_{\mathrm{Sk}}(w_{N_k}, w) \underset{k\to\infty}{\longrightarrow} 0$. ∎

Let $\mathcal{S}_E$ denote the Borel σ-algebra on $\big(\mathcal{D}_E[0,\infty), d^d_{\mathrm{Sk}}\big)$. Since we are going to talk about probability measures on $\big(\mathcal{D}_E[0,\infty), \mathcal{S}_E\big)$ it is important to know more about $\mathcal{S}_E$.

The following result states that $\mathcal{S}_E$ is just the σ-algebra generated by the coordinate variables.

Proposition 1.16 (Borel σ-field)**.** *If $(E, \mathcal{O})$ is Polish, then the Borel-σ-field on $\mathcal{D}_E[0,\infty)$ coincides with the σ-field generated by the coordinate projections $(\xi_t)_{t\geq 0}$, defined as*

$$\xi_t : \quad \mathcal{D}_E[0,\infty) \ni w \mapsto w(t), \qquad t \geq 0. \tag{1.51}$$

Proof. Let $\mathcal{S}_E^{\mathrm{coor}}$ denote the σ-algebra generated by the coordinate maps, i.e.,

$$\mathcal{S}_E^{\mathrm{coor}} := \sigma(\xi_t : t \in [0,\infty)). \tag{1.52}$$

We start showing that $\mathcal{S}_E^{\mathrm{coor}} \subseteq \mathcal{S}_E$. For a given $\varepsilon > 0$, $t \in [0,\infty)$ and f a bounded continuous function on E, consider the following map:

$$f_t^\varepsilon : \quad \mathcal{D}_E[0,\infty) \ni w \mapsto \frac{1}{\varepsilon}\int_t^{t+\varepsilon} \mathrm{d}s\, f(\xi_s(w)) \in \mathbb{R}. \tag{1.53}$$

It is easy to check that f_t^ε is continuous on $\mathcal{D}_E[0,\infty)$, and hence Borel measurable. Moreover, since $\lim_{\varepsilon\downarrow 0} f_t^\varepsilon = f \circ \xi_t$, we find that $f \circ \xi_t$ is Borel measurable for every bounded and continuous function f, and hence also for all bounded functions f. Consequently,

$$\xi_t^{-1}(\Gamma) := \{w \in \mathcal{D}_E[0,\infty) : \xi_t(w) \in \Gamma\} \in \mathcal{S}_E, \quad \Gamma \in \mathcal{B}(E). \tag{1.54}$$

That is, $\mathcal{S}_E^{\mathrm{coor}} \subseteq \mathcal{S}_E$.

To prepare the other direction, notice first that if $D \subseteq [0,\infty)$ is dense then

$$\mathcal{S}_E^{\mathrm{coor}} = \sigma(\xi_t : t \in D). \tag{1.55}$$

Indeed, for each $t \in [0, \infty)$, there exists a sequence $(t_n)_{n\in\mathbb{N}}$ in $D \cap [t, \infty)$ with $(t_n) \downarrow t$, as $n \to \infty$. Therefore, $\xi_t = \lim_{n\to\infty} \xi_{t_n}$ is $\sigma(\xi_t : t \in D)$-measurable.

Assume now that $(E, \mathcal{O})$ is separable. Fix $n \in \mathbb{N}$ and $0 =: t_0 < t_1 < t_2 < \cdots < t_n < t_{n+1} < \infty$. Consider the function

$$\text{(1.56)} \quad \eta : \quad E^{n+1} \ni (\alpha_0, \dots, \alpha_n) \mapsto \sum_{k=0}^{n-1} \alpha_k 1_{[t_k, t_{k+1})} + \alpha_n 1_{[t_n, \infty)} \in \mathcal{D}_E[0, \infty).$$

Since for a metric d on E,

$$\text{(1.57)} \qquad d^d_{\mathrm{Sk}}\big(\eta(\alpha_0, \dots, \alpha_n), \eta(\beta_0, \dots, \beta_n)\big) \le \max_{0\le k\le n} d(\alpha_k, \beta_k),$$

η is continuous. Moreover, since ξ_t is by definition $\mathcal{S}^{\mathrm{coor}}_E$-measurable and $(E, \mathcal{O})$ is separable, for a given $u \in \mathcal{D}_E[0, \infty)$, the following map

$$\text{(1.58)} \qquad \kappa^{u,(t_0,\dots,t_n)} : \quad \mathcal{D}_E[0, \infty) \ni w \mapsto d_{\mathrm{Sk}}\big(u, \eta \circ (\xi_{t_0}, \dots, \xi_{t_n})(w)\big) \in \mathbb{R}$$

is $\mathcal{S}^{\mathrm{coor}}_E$-measurable.

Finally, for each $m \in \mathbb{N}$, let η_m be defined as η was with the special choice $n = m^2$ and $t_i := \frac{i}{m}$, $i = 0, \dots, m^2$. Then for all $w \in \mathcal{D}_E[0, \infty)$,

$$\text{(1.59)} \qquad \lim_{m\to\infty} d_{\mathrm{Sk}}\big(u, \eta_m \circ (\xi_{t_0}, \dots, \xi_{t_{m^2}})(w)\big) = d_{\mathrm{Sk}}\big(u, w\big).$$

Thus, also the map $d_{\mathrm{Sk}}(u, \cdot) : w \mapsto d_{\mathrm{Sk}}\big(u, w\big)$ is $\mathcal{S}^{\mathrm{coor}}_E$-measurable for all fixed $u \in \mathcal{D}_E[0, \infty)$. In particular, every open ball

$$\text{(1.60)} \qquad B(u, \varepsilon) := \big\{w \in \mathcal{D}_E[0, \infty) : d_{\mathrm{Sk}}\big(u, w\big) < \varepsilon\big\}$$

belongs to $\mathcal{S}^{\mathrm{coor}}_E$, and since $(E, \mathcal{O})$ (and by Proposition 1.14 also $\mathcal{D}_E[0, \infty)$) is separable, $\mathcal{S}^{\mathrm{coor}}_E$ contains all open sets in $\mathcal{D}_E[0, \infty)$, and hence contains $\mathcal{S}_E$. ∎

If E is a metrizable space, then we denote the space of continuous functions $w : [0, \infty) \to E$ by $\mathcal{C}_E[0, \infty)$.

Lemma 1.17 (Continuous functions)**.** *The space $\mathcal{C}_E[0, \infty)$ is a closed subset of $\mathcal{D}_E[0, \infty)$. The induced topology on $\mathcal{C}_E[0, \infty)$ is the topology of uniform convergence on compact sets.*

Proof. For *closedness*, let $(w_n)_{n\in\mathbb{N}}$ be a sequence of functions in $\mathcal{C}_E[0, \infty)$, and $w \in \mathcal{D}_E[0, \infty)$ such that $d_{\mathrm{Sk}}(w_n, w) \underset{n\to\infty}{\longrightarrow} 0$. We have to show that $w \in \mathcal{C}_E[0, \infty)$. By condition (c) of Proposition 1.12, for all $T \in [0, \infty)$, there exists a sequence $(\lambda^T_n)_{n\in\mathbb{N}}$ in Λ' satisfying (1.26) and (1.32). Hence, for all $T \in [0, \infty)$ and $\varepsilon > 0$,

- by (1.32), there exists $N = N(T, \varepsilon)$ such that for all $n \ge N$ and $t \in [0, T]$, $|\lambda_n(t) - t| < \varepsilon$, and
- by continuity of w_n, there exists $\delta = \delta(\varepsilon) > 0$ such that for all $s, t \in [0, T]$ with $|s - t| < \delta$, $d(w_n(t), w_n(s)) < \varepsilon$.

Combining both yields that for all $n \geq N(T, \delta(\varepsilon))$ and $t \in [0, T]$,

$$d(w_n(t), w_n(\lambda_n(t))) < \varepsilon. \tag{1.61}$$

Thus, (1.61) together with (1.26) implies

$$\begin{aligned}
&\sup_{t\in[0,T]} d(w_n(t), w(t)) \\
&\quad \leq \sup_{t\in[0,T]} d(w_n(\lambda_n(t)), w_n(t)) + \sup_{t\in[0,T]} d(w_n(\lambda_n(t)), w(t)) \\
&\quad \underset{n\to\infty}{\longrightarrow} 0.
\end{aligned} \tag{1.62}$$

This is equivalent to uniform convergence of $(w_n)_{n\in\mathbb{N}}$ against w on compacta. In particular, the limit function w is continuous. ∎

The next lemma shows that stochastic processes with cadlag sample paths are just random variables with values in a rather large and complicated space.

Lemma 1.18 (Processes with cadlag sample paths). *A function $(t, \omega) \mapsto X_t(\omega)$ is a stochastic process with Polish state space E and cadlag sample paths if and only if $\omega \mapsto (X_t(\omega))_{t\geq 0}$ is a $\mathcal{D}_E[0, \infty)$-valued random variable. Two E-valued stochastic processes X and $\tilde{X}$ with cadlag sample paths have the same finite dimensional distributions if and only if, considered as $\mathcal{D}_E[0, \infty)$-valued random variables, they have the same laws $\mathcal{L}(X)$ and $\mathcal{L}(\tilde{X})$.*

Proof. Let $X : \Omega \to \mathcal{D}_E[0, \infty)$ denote the function $\omega \mapsto X(\omega) := (X_t(\omega))_{t\geq 0}$. By Proposition 1.16, the Borel-σ-field on $\mathcal{D}_E[0, \infty)$ is generated by the coordinate projections $(\xi_t)_{t\geq 0}$. Therefore, the function X is measurable if and only if $X^{-t}(\xi_t^{-1}(A)) \in \mathcal{F}$ for all $A \in \mathcal{B}(E)$. Since $X^{-t}(\xi_t^{-1}(A)) = (\xi_t \circ X)^{-1}(A) = X_t^{-1}(A)$ this is equivalent to the statement that the $(X_t)_{t\geq 0}$ are random variables.

The finite dimensional distributions of an E-valued stochastic process X are uniquely determined by all probabilities of the form

$$\mathbf{P}\{X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n\} \tag{1.63}$$

with $0 \leq t_1 \leq \cdots \leq t_n$ and $A_1, \ldots, A_n \in \mathcal{B}(E)$. The class of all subsets of $\mathcal{D}_E[0, \infty)$ of the form $\{w \in \mathcal{D}_E[0, \infty) : w_{t_1} \in A_1, \ldots, w_{t_n} \in A_n\}$ is closed under finite intersections and generates the Borel-σ-field on $\mathcal{D}_E[0, \infty)$, so the probabilities of the form (1.63) uniquely determine the law $\mathcal{L}(X)$ of X, considered as $\mathcal{D}_E[0, \infty)$-valued random variable. ∎

1.4. **Compactification of Polish spaces.** In this section we collect some important facts about Polish spaces that will be useful later on. In particular, we will see that every Polish space can be embedded in a compact space.

Compact metrizable spaces are, in a sense, the "nicest" topological spaces. A countable product $\prod_{i\in\mathbb{N}} E_i$ of compact metrizable spaces, equipped with the product topology, is compact and metrizable [Kel55, Theorem 4.14].[7] Every compact metrizable space is separable.[8] Conversely, every separable metrizable space can be embedded in a compact metrizable space.

Definition 1.19. *By definition, a* compactification *of a topological space E is a compact topological space $\overline{E}$ such that $E \subseteq \overline{E}$, the topology on E is the induced topology from $\overline{E}$, and $\overline{E}$ is the closure of E.*

Proof. Notice that if $\overline{E}$ is a compactification of E then E is compact and only if $E = \overline{E}$ ∎

The next proposition can be found in [Kel55, Theorem 4.17] or [Cho69, Theorem 6.3].

Proposition 1.20 (Metrizable compactifications)**.** *Every separable metrizable space E has a metrizable compactification $\overline{E}$.*

Definition 1.21 (Product topology)**.** *Let $((E_k, \mathcal{O}_k))_{k\in\mathbb{N}}$ be metrizable topological spaces. The* product topology *$\mathcal{O}$ on $\prod_{k=1}^{\infty} E_k$ is the roughest topology on $\prod_{k=1}^{\infty} E_k$ such that all projections π from $E \to E_i$ are continuous.*

Remark. Let $((E_k, d_k))_{k\in\mathbb{N}}$ be metric spaces. Then the product topology $\mathcal{O}$ on $\prod_{k=1}^{\infty} E_k$ can be metrized by

$$d(x, y) := \prod_{k=1}^{\infty} 2^{-k}\, 1 \wedge d_k(x_k, y_k) \tag{1.64}$$

for all $x := (x_1, x_2, \dots)$ and $y := (y_1, y_2, \dots)$ in E. ∎

Proof. (Sketch) Equip $[0, 1]^{\mathbb{N}}$ with the product topology. Then $[0, 1]^{\mathbb{N}}$ is compact and metrizable. Using Urysohn's lemma, it can be shown that there exist a countable family $(f_i)_{i\in\mathbb{N}}$ of continuous functions $f_i : E \to [0, 1]$ such that the map $f : E \to [0, 1]^{\mathbb{N}}$ defined by $f(x) := (f_i(x))_{i\in\mathbb{N}}$ is open and one-to-one. Since f is obviously continuous, it follows that f is a homeomorphism between E and $f(E)$. Identifying E with its image $f(E)$ and taking for $\overline{E}$ the closure of $f(E)$ in $[0, 1]^{\mathbb{N}}$ we obtain the required compactification. ∎

Unfortunately, for general separable metrizable spaces, E may be a very 'bad' (even non-measurable) subset of its compactification $\overline{E}$. For Polish spaces, and in particular for locally compact spaces, the situation is better. **In what follows, all spaces are separable and metrizable.**

[7] Uncountable products of compact metrizable spaces are still compact but no longer metrizable.

[8] This follows from the fact that for a metric space compact ⇒ totally bounded ⇒ countable basis for the topology ⇒ separable.

Definition 1.22 (Locally compact)**.** *We say that E is* locally compact *if for each $x \in E$ there exists an open set O and a compact set C such that $x \in O \subset C$.*

Exercise 1.23. *Let E be a locally compact space. Show that E is separable, and that there exists compact sets $(C_i)_{i\in\mathbb{N}}$ such that $E = \bigcup_{i\in\mathbb{N}} C_i$.*

We need the following facts.

Proposition 1.24 (Subsets of locally compact and Polish spaces)**.**

(i) *A subset F of a locally compact space E is itself locally compact in the induced topology if and only if $F \subset E$ is the intersection of an open set with a closed set.*

(ii) *A subset F of a Polish space E is itself Polish in the induced topology if and only if $F \subset E$ is a countable intersection of open sets.*

Proof. For Part (i), see [Bou64, §8.16]. For Part (ii), see [Bou58, Section 6, Theorem 1]. ∎

Remark. Sets that are the intersection of an open set with a closed set are called locally closed. Sets that are a countable intersection of open sets are called G_δ-sets. Every closed set is a G_δ-set.[9] ∎

The following is a immediate consequence of Proposition 1.24.

Corollary 1.25. *Let E be a separable metrizable space and $\overline{E}$ is a metrizable compactification of E. Then*

- *E is locally compact if and only if E is an open subset of $\overline{E}$.*
- *E is Polish if and only if E is a countable intersection of open sets in $\overline{E}$.*

Exercise 1.26. *Prove Corollary 1.25.*

In particular,

$$E \text{ compact } \Rightarrow \ E \text{ locally compact } \Rightarrow \ E \text{ Polish.}^{10} \tag{1.65}$$

If E is locally compact but not compact, then there exists a metrizable compactification $\overline{E}$ of E such that $\overline{E}\backslash E$ consists of one point (usually denoted by ∞). In this case, $\overline{E}$ is the set

$$E^\infty := E \cup \{\infty\} \tag{1.66}$$

and by definition a subset $U \subseteq \bar{E}$ is open if either

[9] If (E, d) is a metric space and $A \subseteq E$ is closed, then $O_n := \{x \in E : d(x, A) < \frac{1}{n}\}$ are open sets with $A = \bigcap_n O_n$.

[10] If E is a separable metrizable space and there exists a metrizable compactification $\overline{E}$ of E such that $E \subset \overline{E}$ is a Borel measurable set, an analytic set, or a universally measurable set, then E is called a Lusin space, a Souslin space, or a Radon space, respectively.

- $\infty \notin U$ and U is open in the original topology.
- $\infty \in U$ and $\bar{E} \setminus U$ is compact in the original topology of E.

We call E^∞ the *one-point compactification* of E.

As an application of Proposition 1.24, we prove:

Proposition 1.27 (Product spaces)**.**
(i) *A finite product $E_1 \times \cdots \times E_n$ of locally compact spaces is locally compact but a countably infinite product $\prod_{i\in\mathbb{N}} E_i$ is not, unless all but finitely many E_i are compact.*
(ii) *A countable product $\prod_{i\in\mathbb{N}} E_i$ of Polish spaces is Polish.*

Proof. Let E_i $(i \in \mathbb{N})$ be locally compact spaces and let $\overline{E}_i$ be metrizable compactifications of the E_i. Then $\prod_{i\in\mathbb{N}} \overline{E}_i$ is a metrizable compactification of $\prod_{i\in\mathbb{N}} E_i$. Let π_i denote the projection on $\overline{E}_i$. If all but finitely many E_i are compact, then there is an n such that $E_i = \overline{E}_i$ for all $i > n$. Therefore $\prod_{i\in\mathbb{N}} E_i = \bigcap_{i=1}^n \pi_i^{-1}(E_i)$ is an open subset of $\prod_{i\in\mathbb{N}} \overline{E}_i$, hence $\prod_{i\in\mathbb{N}} E_i$ is locally compact. If there are infinitely many E_{i_k} that are not compact, then choose $x = (x_i)_{i\in\mathbb{N}} \in \prod_{i\in\mathbb{N}} E_i$ and $x^{(k)} \in \prod_{i\in\mathbb{N}} \overline{E}_i$ such that $x_i^{(k)} = x_i$ for all $i \neq i_k$ and $x_{i_k}^{(k)} \in \overline{E}_{i_k} \backslash E_{i_k}$. Then $x^{(k)} \notin \prod_{i\in\mathbb{N}} E_i$ and $x^{(k)} \to x$ in the product topology, which proves that $\prod_{i\in\mathbb{N}} \overline{E}_i \backslash \prod_{i\in\mathbb{N}} E_i$ is not closed, hence $\prod_{i\in\mathbb{N}} E_i$ is not open, hence $\prod_{i\in\mathbb{N}} E_i$ is not locally compact.

If the E_i are Polish, then each E_i is a countable intersection of open subsets of $\overline{E}_i$, say $E_i = \bigcap_j O_{ij}$. Then $\prod_{i\in\mathbb{N}} E_i = \bigcap_{i,j} \pi_i^{-i}(O_{ij})$ is a countable intersection of open subsets of $\prod_{i\in\mathbb{N}} \overline{E}_i$, hence $\prod_{i\in\mathbb{N}} E_i$ is Polish. ∎

Definition 1.28 (Separating points)**.** *We say that a family $(f_i)_{i\in I}$ of functions on a space E* separates points *if for each $x \neq y$ there exists an $i \in I$ such that $f_i(x) \neq f_i(y)$.*

The next (deep) result is often very useful.

Proposition 1.29 (Borel σ-field)**.** *Let E, $(E_i)_{i\in\mathbb{N}}$ be Polish spaces and let $(f_i)_{i\in\mathbb{N}}$ be a countable family of measurable functions $f_i : E \to E_i$ that separates points. Then $\sigma(f_i : i \in \mathbb{N}) = \mathcal{B}(E)$.*

Proof. See [Sch73, Lemma II.18]. Warning: the statement is false for uncountable families $(f_i)_{i\in I}$. For example, if $E = [0,1]$, then the functions $(1_{\{x\}})_{x\in[0,1]}$ separate points, but they generate the σ-field $\mathcal{S} := \{A \subset [0,1] :$ A countable or $[0,1]\backslash A$ countable$\}$. ∎

A simple application is:

Corollary 1.30 (Product σ-field)**.** *If $(E_i)_{i\in\mathbb{N}}$ are Polish spaces, then the Borel-σ-field $\mathcal{B}(\prod_{i\in\mathbb{N}} E_i)$ coincides with the product-σ-field $\prod_{i\in\mathbb{N}} \mathcal{B}(E_i)$.*

Proof. Let π_i denote the projection on E_i. Then the functions $(\pi_i)_{i\in\mathbb{N}}$ are continuous (hence certainly measurable) and separate points. ∎

Note that Proposition 1.29 also implies that if E is Polish, then the Borel-σ-field on $\mathcal{D}_E[0,\infty)$ coincides with the σ-field generated by the coordinate projections $\{\xi_t : t \in \mathbb{Q} \cap [0,\infty)\}$. This strengthens Proposition 1.16!

2. Markov processes

In the previous section, we have studied stochastic processes in general, and stochastic processes with cadlag sample paths in particular. In the present section we take a look at a special class of stochastic processes, namely those which have the Markov property, and in particular at those whose transition probabilities are time-homogeneous. We will see how such time-homogeneous transition probabilities can be interpreted as semigroups. In the next sections we will then see how a certain type of these semigroups, namely those which have the Feller property, may be constructed from their generators, and how such semigroups give rise to Markov processes with cadlag sample paths.

2.1. **The Markov property.** We start by recalling the notion of conditional expectation.

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be our underlying probability space. For any σ-field $\mathcal{H}$, let

$$B(\mathcal{H}) := \{f : \Omega \to \mathbb{R} \, : \, f \ \mathcal{H}\text{-measurable and bounded}\}. \tag{2.1}$$

Definition 2.1. *The* conditional expectation *of a random variable $F \in B(\mathcal{F})$ given $\mathcal{H}$, denoted by $\mathbf{E}_{\mathcal{H}}[F]$ or $\mathbf{E}[F|\mathcal{H}]$, is a random variable such that*

$$\begin{aligned} &(1) \quad \mathbf{E}_{\mathcal{H}}(F) \in B(\mathcal{H}), \\ &(2) \quad \mathbf{E}[\mathbf{E}_{\mathcal{H}}(F)H] = \mathbf{E}[FH] \quad \forall H \in B(\mathcal{H}). \end{aligned} \tag{2.2}$$

The random variable $\mathbf{E}_{\mathcal{H}}[F]$ is almost surely defined through these two conditions (with respect to the restriction of $\mathbf{P}$ to $\mathcal{H}$). Some elementary properties of the conditional expectation are:

$$\begin{aligned} &(\text{“continuity”}) \quad \mathbf{E}_{\mathcal{H}}(F_i) \uparrow \mathbf{E}_{\mathcal{H}}(F) \quad \text{a.s.} \quad \forall F_i \uparrow F, \\ &(\text{“projection”}) \quad \mathbf{E}_{\mathcal{G}}[\mathbf{E}_{\mathcal{H}}[F]] = \mathbf{E}_{\mathcal{G}}[F] \quad \text{a.s.} \quad \forall \mathcal{G} \subset \mathcal{H}, \\ &(\text{“pull out”}) \quad \mathbf{E}_{\mathcal{H}}(F)H = \mathbf{E}_{\mathcal{H}}(FH) \quad \text{a.s.} \quad \forall H \in B(\mathcal{H}). \end{aligned} \tag{2.3}$$

We write $\mathbf{P}(A|\mathcal{H}) := \mathbf{E}_{\mathcal{H}}[1_A]$ $(A \in \mathcal{F})$ and for any random variable G we abbreviate $\mathbf{E}_G[F] = \mathbf{E}[F|G] := \mathbf{E}[F|\sigma(G)]$ and $\mathbf{P}(A|G) := \mathbf{P}(A|\sigma(G))$.

Proof of the pull out property. We need to check that $\mathbf{E}_{\mathcal{H}}[F]H$ satisfies (2.2). Indeed, $\mathbf{E}_{\mathcal{H}}[F]H \in B(\mathcal{H})$ since $\mathbf{E}_{\mathcal{H}}[F] \in B(\mathcal{H})$ and $H \in B(\mathcal{H})$, and applying (2.2) (2) twice we see that $\mathbf{E}[\mathbf{E}_{\mathcal{H}}(F)HH'] = \mathbf{E}[FHH'] = \mathbf{E}[\mathbf{E}_{\mathcal{H}}[FH]H']$ for all $H \in B(\mathcal{H})$, which shows that $\mathbf{E}_{\mathcal{H}}[F]H$ satisfies (2.2) (2). ∎

Lemma 2.2 (Conditional expectation)**.** *It suffices to check (2.2) (2) for H of the form $H = 1_A$ with $A \in \mathcal{G}$, where $\mathcal{G}$ is closed under finite intersections, there exists $A_i \in \mathcal{G}$ such that $A_i \uparrow \Omega$, and $\sigma(\mathcal{G}) = \mathcal{H}$.*

Before we prove Lemma 2.2, we recall a basic fact from measure theory. A subset $\mathcal{D}$ of the set of all subsets of Ω is called a *Dynkin system* if

(1) $\Omega \in \mathcal{D}$,
(2) $A, B \in \mathcal{D},\ A \supseteq B \ \Rightarrow\ A \backslash B \in \mathcal{D}$, and

(3) $A_n \in \mathcal{D},\ A_n \uparrow A\ \Rightarrow\ A \in \mathcal{D}$.

Lemma 2.3. *Let $\mathcal{C}$ be a collection of subsets of Ω which is closed under finite intersections. Then the smallest Dynkin system which contains $\mathcal{C}$ is equal to $\sigma(\mathcal{C})$.*

Proof. See any book on measure theory. ∎

Proof of Lemma 2.2. Set $\mathcal{D} := \{A \in \mathcal{H} : \mathbf{E}[\mathbf{E}_{\mathcal{H}}(F)1_A] = \mathbf{E}[F1_A]\}$. By the linearity and continuity of the conditional expectation, $A, B \in \mathcal{D},\ A \supseteq B\ \Rightarrow A\backslash B \in \mathcal{D}$ and $A_n \in \mathcal{D},\ A_n \uparrow A\ \Rightarrow\ A \in \mathcal{D}$. Since we are assuming that there exists $A_i \in \mathcal{G}$ such that $A_i \uparrow \Omega$, we also have $\Omega \in \mathcal{D}$, so $\mathcal{D}$ is a Dynkin system. Therefore, by Lemma 2.3, $\mathbf{E}[\mathbf{E}_{\mathcal{H}}(F)1_A] = \mathbf{E}[F1_A]$ for all $A \in \mathcal{G}$ implies $\mathbf{E}[\mathbf{E}_{\mathcal{H}}(F)1_A] = \mathbf{E}[F1_A]$ for all $A \in \mathcal{H}$. The general statement follows by approximation with simple functions, using the linearity and continuity of $E_{\mathcal{H}}$. ∎

Example. Let Z be uniformly distributed on $[0,1]$, $X := \cos(2\pi Z)$ and $Y := \sin(2\pi Z)$. Then (X,Y) is uniform distributed on $\{(x,y) : x^2+y^2 = 1\}$, and hence a version of the conditioned distribution of X given Y is

$$P(\{X \in C\}|Y) = \frac{1}{2}\delta_{\sqrt{1-Y^2}}(C) + \frac{1}{2}\delta_{-\sqrt{1-Y^2}}(C). \tag{2.4}$$

Moreover, we find $\mathbf{E}_Y[X] = 0$, and $\mathbf{E}_Y[X^2] = 1 - Y^2$. ∎

Let X be a stochastic process with values in a Polish space $(E, \mathcal{O})$. For each $t \geq 0$, we introduce the σ-fields

$$\mathcal{F}_t^X := \sigma\big(X_s;\ 0 \leq s \leq t\big), \tag{2.5}$$

and

$$\mathcal{G}_t^X := \sigma\big(X_u;\ t \leq u\big). \tag{2.6}$$

Note that $\mathcal{F}_t^X$ is the collection of events that refer to the behavior of the process X up to time t. That is, $\mathcal{F}_t^X$ contains all "information" that can be obtained by observing the process X up to time t. Likewise, $\mathcal{G}_t^X$ contains all information that can be obtained by observing the process X after time t.

Proposition 2.4 (Markov property)**.** *The following four conditions on X are equivalent.*

(a) *For all $A \in \mathcal{F}_t^X$, $B \in \mathcal{G}_t^X$, and $t \geq 0$,*

$$\mathbf{P}(A \cap B|X_t) = \mathbf{P}(A|X_t)\mathbf{P}(B|X_t) \quad a.s., \tag{2.7}$$

(b) *For all $B \in \mathcal{G}_t^X$, $t \geq 0$,*

$$\mathbf{P}(B|X_t) = \mathbf{P}(B|\mathcal{F}_t^X) \quad a.s., \tag{2.8}$$

(c) *For all* $C \in \mathcal{B}(E)$, *and* $0 \le s \le t$,

$$\mathbf{P}(\{X_t \in C\}|\mathcal{F}_s^X) = \mathbf{P}(\{X_t \in C\}|X_s) \quad a.s., \tag{2.9}$$

(d) *For all* $C_1, C_2, \ldots \in \mathcal{B}(E)$, *and* $0 \le t_1 \le \cdots \le t_n$,

$$\begin{aligned} &\mathbf{P}\{X_{t_1} \in C_1, \ldots, X_{t_n} \in C_n\} \\ &= \mathbf{E}\Big[1_{\{X_{t_1} \in C_1\}} \mathbf{E}_{X_{t_1}}\big[1_{\{X_{t_2} \in C_2\}} \mathbf{E}_{X_{t_2}}[\cdots \mathbf{E}_{X_{t_{n-1}}}[1_{\{X_{t_n} \in C_n\}}]]\big]\Big]. \end{aligned} \tag{2.10}$$

Remark. If X satisfies the equivalent conditions from Proposition 2.4 then we say that X has the *Markov property*. Note that condition (a) says that the future and the past are conditionally independent given the present. Condition (b) says that the behavior of X after time t depends only on the behavior of X before time t through the state of X at time t. ∎

Exercise 2.5 (Gaussian processes with Markov property). *A stochastic process* $X := (X_t)_{t \in [0,\infty)}$ *is called* Gaussian process *if for all* $n \in \mathbb{N}$, *and* $(t_1, \ldots, t_n) \in [0,\infty)^n$, *the random vector* $(X_{t_1}, \ldots, X_{t_n})^t$ *is normal distributed with mean* $(\mu_{t_1}, \ldots, \mu_{t_n}) \in \mathbb{R}^n$ *and covariance function* $\Gamma(s,t) := \mathbf{E}[(X_s - \mu_s)(X_t - \mu_t)]$.

Show that a centered (i.e., $\mu \equiv 0$*) Gaussian process has the Markov property if and only if for all* $s, t, u \in [0,\infty)$ *with* $s < t < u$,

$$\Gamma(s,u)\Gamma(t,t) = \Gamma(s,t)\Gamma(t,u). \tag{2.11}$$

Proof of Proposition 2.4. (a)⇒(b): We have to show that for all $A \in \mathcal{F}_t^X$ and $B \in \mathcal{G}_t^X$

$$\mathbf{E}\big[1_A \mathbf{P}(B|X_s)\big] = \mathbf{P}\big(A \cap B\big). \tag{2.12}$$

By the projection property and by the pull out property applied to $H := \mathbf{P}(B|X_s)$,

$$\begin{aligned} \mathbf{E}\big[1_A \mathbf{P}(B|X_s)\big] &= \mathbf{E}\big[\mathbf{E}_{X_s}[1_A \mathbf{P}(B|X_s)]\big] \\ &= \mathbf{E}\big[\mathbf{P}(A|X_s)\mathbf{P}(B|X_s)\big] \end{aligned} \tag{2.13}$$

By (a),

$$\begin{aligned} \mathbf{E}\big[\mathbf{P}(A|X_s)\mathbf{P}(B|X_s)\big] &= \mathbf{E}\big[\mathbf{P}(A \cap B|X_s)\big] \\ &= \mathbf{P}\big(A \cap B\big). \end{aligned} \tag{2.14}$$

(b)⇒(c): This follows by applying (2.8) to $t := s$ and $B := \{X_t \in C\} \in \mathcal{G}_s^X$.
(c)⇒(d): Since $\mathcal{F}_{t_1} \subseteq \mathcal{F}_{t_2} \subseteq \ldots$, repeated use of the projection property and the pull out property give

$$\begin{aligned} &\mathbf{P}\{X_{t_1} \in C_1, \ldots, X_{t_n} \in C_n\} = \mathbf{E}[1_{\{X_{t_1} \in C_1\}} \cdots 1_{\{X_{t_n} \in C_n\}}] \\ &= \mathbf{E}\Big[1_{\{X_{t_1} \in C_1\}} \mathbf{E}_{\mathcal{F}_{t_1}^X}\big[1_{\{X_{t_2} \in C_2\}} \mathbf{E}_{\mathcal{F}_{t_2}^X}[\cdots \mathbf{E}_{\mathcal{F}_{t_{n-1}}^X}[1_{\{X_{t_n} \in C_n\}}]]\big]\Big], \end{aligned} \tag{2.15}$$

and the right hand side of (2.15) equals the right hand side of (2.10) by (b).

(d)⇒(c): By approximation with simple functions it follows from (d) that for any $0 \leq t_1 \leq \cdots \leq t_n$ and $F_1 \in B(\sigma(X_{t_1}), \ldots, F_n \in B(\sigma(X_{t_n}),$

$$\begin{aligned} &\mathbf{E}[F_1 \cdots F_n] \\ &\quad = \mathbf{E}\Big[F_1 \mathbf{E}_{X_{t_1}}\big[F_2 \mathbf{E}_{X_{t_2}}\big[\cdots \mathbf{E}_{X_{t_{n-1}}}[F_n]\big]\big]\Big]. \end{aligned} \tag{2.16}$$

Let $0 \leq s_1 \leq \cdots \leq s_m = s \leq t$, $C_1, \ldots, C_m, C \in \mathcal{B}(E)$. Using (c) and applying (2.16) to $n = m-1$ and $F_n = 1_{\{X_{s_m} \in C_m\}} \mathbf{E}_{X_s}[1_{\{X_t \in C\}}]$, we find that

$$\begin{aligned} &\mathbf{E}\big[1_{\{X_{s_1} \in C_1, \ldots, X_{s_m} \in C_m\}} 1_{\{X_t \in C\}}\big] \\ &\quad = \mathbf{E}\Big[1_{\{X_{s_1} \in C_1\}} \mathbf{E}_{X_{s_1}}\big[\cdots \mathbf{E}_{X_{s_{m-1}}}\big[1_{\{X_{s_m} \in C_m\}} \mathbf{E}_{X_s}[1_{\{X_t \in C\}}]\big]\big]\big]\Big] \\ &\quad = \mathbf{E}\big[1_{\{X_{s_1} \in C_1, \ldots, X_{s_m} \in C_m\}} \mathbf{E}_{X_s}[1_{\{X_t \in C\}}]\big]. \end{aligned} \tag{2.17}$$

It follows from Lemma 2.2 that

$$\mathbf{E}_{X_s}[1_{\{X_t \in C\}}] = \mathbf{E}_{\mathcal{F}_s^X}[1_{\{X_t \in C\}}]. \tag{2.18}$$

(c)⇒(b): By approximation with simple functions it follows from (c) that for all $F \in B(\sigma(X_t))$ and $0 \leq s \leq t$,

$$\mathbf{E}[F|\mathcal{F}_s^X] = \mathbf{E}[F|X_s] \quad \text{a.s.} \tag{2.19}$$

Let $0 \leq t \leq u_1 \leq \cdots \leq u_m$ and $C_1, \ldots, C_m \in \mathcal{B}(E)$. Then repeated use of the projection property, the pull out property, and (2.19) give

$$\begin{aligned} &\mathbf{E}_{\mathcal{F}_t^X}\big[1_{\{X_{u_1} \in C_1, \ldots, X_{u_m} \in C_m\}}\big] \\ &\quad = \mathbf{E}_{\mathcal{F}_t^X}\big[1_{\{X_{u_1} \in C_1\}} \mathbf{E}_{\mathcal{F}_{u_1}^X}\big[\cdots \mathbf{E}_{\mathcal{F}_{u_{m-1}}^X}[1_{\{X_{u_m} \in C_m\}}]\big]\big] \\ &\quad = \mathbf{E}_{X_t}\big[1_{\{X_{u_1} \in C_1\}} \mathbf{E}_{X_{u_1}}\big[\cdots \mathbf{E}_{X_{u_{m-1}}}[1_{\{X_{u_m} \in C_m\}}]\big]\big]. \end{aligned} \tag{2.20}$$

In the last step we have applied (2.19) first to $1_{\{X_{u_m} \in C_m\}} \in B(\sigma(X_{u_n}))$, then to $1_{\{X_{u_{m-1}} \in C_m\}} \mathbf{E}_{X_{u_{m-1}}}[1_{\{X_{u_m} \in C_m\}}] \in B(\sigma(X_{u_{n-1}}))$, and so on. It follows that $\mathbf{E}_{\mathcal{F}_t^X}\big[1_{\{X_{u_1} \in C_1, \ldots, X_{u_m} \in C_m\}}\big]$ is $\sigma(X_t)$-measurable. Therefore, by the projection property,

$$\begin{aligned} &\mathbf{E}_{X_t}\big[1_{\{X_{u_1} \in C_1, \ldots, X_{u_m} \in C_m\}}\big] \\ &\quad = \mathbf{E}_{X_t}\Big[\mathbf{E}_{\mathcal{F}_t^X}\big[1_{\{X_{u_1} \in C_1, \ldots, X_{u_m} \in C_m\}}\big]\Big] \\ &\quad = \mathbf{E}_{\mathcal{F}_t^X}\big[1_{\{X_{u_1} \in C_1, \ldots, X_{u_m} \in C_m\}}\big]. \end{aligned} \tag{2.21}$$

The class of all sets A such that $\mathbf{E}_{X_t}[1_A] = \mathbf{E}_{\mathcal{F}_t^X}[1_A]$ forms a Dynkin system, so by Lemma 2.3 we arrive at (b).

(b)⇒(a): Indeed, by (2.8) and the pull out property, for all $D \in \sigma(X_t)$,

$$\begin{aligned} \mathbf{P}\big(A \cap B \cap D\big) &= \mathbf{E}\big[\mathbf{P}(B|\mathcal{F}_t^X)1_{A\cap D}\big] \\ &= \mathbf{E}\big[\mathbf{P}(B|X_t)1_{A\cap D}\big] \\ &= \mathbf{E}\Big[\mathbf{E}_{X_t}\big[\mathbf{P}(B|X_t)1_{A\cap D}\big]\Big] \\ &= \mathbf{E}\big[\mathbf{P}(A|X_t)\mathbf{P}(B|X_t)1_D\big], \end{aligned} \tag{2.22}$$

which proves (2.7) because $\mathbf{P}(A|X_t)\mathbf{P}(B|X_t) \in B(\sigma(X_t))$. ∎

2.2. **Transition probabilities.** Let E, F be Polish spaces. By definition, a *probability kernel* from E to F is a function $K : E \times \mathcal{B}(F) \to [0,1]$ such that

1. For fixed $x \in E$, $K(x,\cdot)$ is a probability measure on F.
2. For fixed $A \in \mathcal{B}(F)$, $K(\cdot, A)$ is a measurable function on E.

If $E = F$ then we say that K is a probability kernel on E.

Example. For all $x \in \mathbb{R}$ and $A \in \mathcal{B}(\mathbb{R})$, set

$$K\big(x, A\big) := \frac{1}{\sqrt{2\pi}} \int_A \mathrm{d}y \, \exp\big[-\frac{(x-y)^2}{2}\big]. \tag{2.23}$$

Then K is a probability kernel on $\mathbb{R}$. ∎

There is another way of looking at probability kernels that is often very useful. For any Polish space E we define

$$B(E) := \{f : E \to \mathbb{R} \, : \, f \text{ Borel measurable and bounded}\}. \tag{2.24}$$

Lemma 2.6 (Probability kernels)**.** *If K is a probability kernel from E to F then the operator $K : B(F) \to B(E)$ defined by*

$$Kf(x) := \int_F K(x, \mathrm{d}y)\, f(y), \qquad x \in E, \; f \in \mathcal{B}(F), \tag{2.25}$$

satisfies

1. *K is conservative, i.e., $K1 = 1$.*
2. *K is positive, i.e., $Kf \geq 0$ for all $f \geq 0$.*
3. *K is linear, i.e., $K(\lambda_1 f_1 + \lambda_2 f_2) = \lambda_1 K(f_1) + \lambda_2 K(f_2)$ for all $f_1, f_2 \in B(F)$ and $\lambda_1, \lambda_2 \in \mathbb{R}$.*
4. *K is continuous with respect to monotone sequences, i.e., $K(f_i) \uparrow K(f)$ for all $f_i \uparrow f$, $f_i, f \in B(F)$.*

Conversely, every operator $K : B(F) \to B(E)$ with these properties corresponds to a probability kernel from E to F as in (2.25).

Proof. If K is a probability kernel from E to F then the operator $K : B(F) \to B(E)$ defined in (2.25) maps $B(F)$ into $B(E)$ since $K(\cdot, A)$ is measurable for each $A \in \mathcal{B}(F)$, and the operator K has the properties (1)–(4)) since $K(x, \cdot)$ is a probability measure for each $x \in E$. Conversely, if $K : B(F) \to B(E)$ satisfies (1)–(4) then $K(x, A) := K1_A(x)$ is measurable as a function of x for each $A \in \mathcal{B}(F)$ since the operator K maps $B(F)$ into $B(E)$ and $K(x, \cdot)$ is a probability measure by (1)–(4). ∎

Remark. If E is a set consisting of one point, say $E = \{0\}$, then a probability kernel from E to F is just a probability measure $K(0, \cdot) = \mu$, say. In this case $B(E)$ is isomorphic to $\mathbb{R}$ and the operator in (2.25), considered as an operator from $B(F)$ to $\mathbb{R}$, is given by

$$\mu f := \int_F \mu(\mathrm{d}y) f(y), \qquad f \in \mathcal{B}(F). \tag{2.26}$$

∎

If E, F, and G are Polish spaces, K is a probability kernel from E to F, and L is a probability kernel from F to G, then the composition of the operators $L : B(G) \to B(F)$ and $K : B(F) \to B(E)$ yields an operator $KL : B(G) \to B(E)$ that corresponds to the *composite kernel* KL from E to G given by

$$(KL)(x, A) := \int_F K(x, \mathrm{d}y) L(y, A) \qquad (x \in E,\ A \in \mathcal{B}(G)). \tag{2.27}$$

The following result states that conditional probabilities of random variables with values in Polish spaces are associated with probability kernels.

Proposition 2.7 (Conditional probability kernel)**.** *Let X, Y be random variables with values in Polish spaces E and F, respectively. Then there exists a probability kernel P from E to F such that for all $A \in \mathcal{B}(F)$,*

$$\mathbf{P}\{Y \in A | X\} = P(X, A), \qquad \text{a.s.}$$

The kernel P is unique up to a.s. equality with respect to $\mathcal{L}(X)$.

Proof (sketch). Let $\mathcal{M}(F)$ be the space of finite measures on F, equipped with σ-field generated by the mappings $\mu \mapsto \mu(A)$ with $A \in \mathcal{B}(F)$. Define a function $M : \mathcal{B}(E) \to \mathcal{M}(F)$ by

$$M(B)(A) := \mathbf{P}\{X \in B,\ Y \in A\}. \tag{2.28}$$

Then $M(\emptyset) = 0$, the zero measure, and M is σ-additive, so we may interpret M as a measure on $(E, \mathcal{B}(E))$ with values in $\mathcal{M}(F)$. Moreover, $\mathbf{P}\{X \in B\} = 0$ implies $M(B) = 0$, so M is absolutely continuous with respect to P_X, the law of X. It follows from the fact that F is a Polish space that the space $\mathcal{M}(F)$ has the Radon-Nikodym property, i.e., the Radon-Nikodym theorem also holds for $\mathcal{M}(F)$-valued measures and functions. As a result, there exists a $\mathcal{M}(F)$-valued measurable function $x \mapsto P(x,\cdot)$ from E to $\mathcal{M}(F)$, unique up to a.s. equality with respect to P_X, such that

$$M(B) = \int_B P(x,\cdot)P_X(\mathrm{d}x). \tag{2.29}$$

It is not hard to check that a function $x \mapsto P(x,\cdot)$ from E to $\mathcal{M}(F)$ is measurable if and only if P is a probability kernel from E to F. Now (2.29) says that

$$\mathbf{E}[P(X,A)1_B] = \int_B P(x,A)P_X(\mathrm{d}x) = M(B)(A) = \mathbf{P}\{X \in B,\ Y \in A\}, \tag{2.30}$$

which is equivalent to (2.7). ∎

Remark. Proposition 2.7 remains true if only F is Polish and E is any measurable space. ∎

It follows from Proposition 2.7 that for any stochastic process X in E there exist probability kernels $(P_{s,t})_{0\leq s\leq t}$ on E such that for all $A \in \mathcal{B}(E)$, and $0 \leq s \leq t$,

$$\mathbf{P}\{X_t \in A | X_s\} = P_{s,t}(X_s, A), \qquad \text{a.s.} \tag{2.31}$$

We call $(P_{s,t})_{0\leq s\leq t}$ the *transition probabilities* of X.

Proposition 2.8 (Markov transition probabilities)**.** *Let X be a stochastic process with values in E and let $(P_{s,t})_{0\leq s\leq t}$ be probability kernels on E. Then the following conditions are equivalent:*

(a) *For all $C \in \mathcal{B}(E)$ and $0 \leq s \leq t$,*

$$\mathbf{P}(\{X_t \in C\}|\mathcal{F}^X_s) = P_{s,t}(X_s, C), \qquad a.s. \tag{2.32}$$

(b) *X has the Markov property, and for all $C \in \mathcal{B}(E)$ and $0 \leq s \leq t$,*

$$\mathbf{P}(\{X_t \in C\}|X_s) = P_{s,t}(X_s, C), \qquad a.s. \tag{2.33}$$

(c) *For all $C_1, \ldots, C_n \in \mathcal{B}(E)$ and $0 = t_0 \leq t_1 \leq \cdots \leq t_n$,*

$$\begin{aligned} &\mathbf{P}\{X_{t_1} \in C_1, \ldots, X_{t_n} \in C_n\} \\ &\quad = \int \mathbf{P}\{X_0 \in \mathrm{d}x_0\} \int_{C_1} P_{t_0,t_1}(x_0, \mathrm{d}x_1) \cdots \int_{C_n} P_{t_{n-1},t_n}(x_{n-1}, \mathrm{d}x_n). \end{aligned} \tag{2.34}$$

Proof. (a)⇒(b): It follows from (a) that $\mathbf{P}(\{X_t \in C\}|\mathcal{F}^X_s)$ is measurable with respect to $\sigma(X_t)$, and therefore $\mathbf{P}(\{X_t \in C\}|X_s) = \mathbf{E}[\mathbf{P}(\{X_t \in C\}|\mathcal{F}^X_s)|X_s] = \mathbf{P}(\{X_t \in C\}|\mathcal{F}^X_s)$, a.s. By condition (c) of Proposition 2.4, X has the Markov property.

(b)⇒(a): Since X has the Markov property, by condition (c) of Proposition 2.4, $\mathbf{E}[\mathbf{P}(\{X_t \in C\}|\mathcal{F}^X_s)|X_t] = \mathbf{P}(\{X_t \in C\}|X_s) = \mathbf{P}(\{X_t \in C\}|\mathcal{F}^X_s)$, a.s.

(b)⇒(c): Since X has the Markov property, X satisfies condition (d) of Proposition 2.4. Using the fact that $\mathbf{P}(\{X_t \in C\}|X_s) = P_{s,t}(X_s, C)$, a.s., we arrive at (c).

(c)⇒(b): We start by showing that for all $C \in \mathcal{B}(E)$ and $0 \le s \le t$),

$$\mathbf{P}(\{X_t \in C\}|X_s) = P_{s,t}(X_s, C), \tag{2.35}$$

where $(P_{s,t})_{0\le s\le t}$ are the probability kernels in (c). Since $P_{s,t}(X_s, C)$ is measurable with respect to $\sigma(X_s)$, by the definition of the conditional probability, it suffices to show that $\mathbf{E}[P_{s,t}(X_s, C)1_{\{X_s\in B\}}] = \mathbf{P}\{X_s \in B, X_t \in C\}$ for all $B, C \in \mathcal{B}(E)$. Indeed, by (c),

(2.36)
$$\begin{aligned}\mathbf{E}[P_{s,t}(X_s, C)\}1_{\{X_s\in B\}}] &= \int \mathbf{P}\{X_0 \in \mathrm{d}x_0\} \int_B P_{0,s}(x_0, \mathrm{d}x_1)P_{s,t}(x_1, C)\\ &= \mathbf{P}\{X_s \in B, X_t \in C\}.\end{aligned}$$

This proves (2.35), i.e., the $(P_{s,t})_{0\le s\le t}$ are the transition probabilities of X. It follows that X satisfies condition (d) from Proposition 2.4, so X has the Markov property. ∎

2.3. **Transition functions and Markov semigroups.** Condition (c) of Proposition 2.8 shows that the finite dimensional distributions of a process X with the Markov probability are uniquely determined by its transition probabilities $(P_{s,t})_{0\le s\le t}$ and its *initial law* $\mathcal{L}(X_0)$. We will mainly be interested in the case that the transition probabilities can be chosen in such a way that $P_{s,t}$ is a function of $t-s$ only. This leads to the following definition. Recall that the *delta-measure* δ_x in a point x is defined as

$$\delta_x(A) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases} \tag{2.37}$$

Definition 2.9 (Transition function)**.** *By definition, a* transition function *on E is a collection $(P_t)_{t\ge0}$ of probability kernels on E such that*

(1) *(Initial law) For all $x \in E$,*

$$P_0(x, \cdot) := \delta_x, \tag{2.38}$$

(2) *(Chapman-Kolmogorov equation) For all $x \in E$, $C \in \mathcal{B}(E)$, and $0 \le s \le t$,*

$$\int P_s(x, \mathrm{d}y)P_t(y, A) = P_{s+t}(x, A). \tag{2.39}$$

We make the following observation.

Lemma 2.10 (Markov semigroups)**.** *A collection $(P_t)_{t\ge0}$ of probability kernels on E is a transition function if and only if the associated operators $P_t : B(E) \to B(E)$ defined by*

$$P_tf(x) := \int_E P_t(x, \mathrm{d}y)f(y) \qquad x \in E,\ f \in B(E) \tag{2.40}$$

satisfy

(1) $P_0 f = f \quad (f \in B(E))$,
(2) $P_s P_t = P_{s+t} \quad (s, t \geq 0)$.

Properties (1) and (2) from Lemma 2.10 say that the operators $(P_t)_{t\geq 0}$ form a semigroup. If $(P_t)_{t\geq 0}$ is a transition function then we call the associated semigroup of operators on $B(E)$ a *Markov semigroup.*

Proposition 2.11 (Markov processes)**.** *Let X be a stochastic process with values in E and let $(P_t)_{t\geq 0}$ be a transition function on E. Then the following conditions are equivalent:*

(a) *For all $f \in B(E)$ and $0 \leq s \leq t$,*

$$\mathbf{E}[f(X_t)|\mathcal{F}^X_s] = P_{t-s} f(X_s), \qquad a.s. \tag{2.41}$$

(b) *X has the Markov property, and for all $A \in \mathcal{B}(E)$ and $0 \leq s \leq t$,*

$$\mathbf{P}(\{X_t \in A\}|X_s) = P_{t-s}(X_s, A), \qquad a.s. \tag{2.42}$$

(c) *For all $A_1, \ldots, A_n \in \mathcal{B}(E)$ and $0 = t_0 \leq t_1 \leq \cdots \leq t_n$,*

$$\begin{aligned}&\mathbf{P}\{X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n\}\\ &\quad = \int \mathbf{P}\{X_0 \in \mathrm{d}x_0\} \int_{A_1} P_{t_1-t_0}(x_0, \mathrm{d}x_1) \cdots \int_{A_n} P_{t_n - t_{n-1},}(x_{n-1}, \mathrm{d}x_n).\end{aligned} \tag{2.43}$$

Proof. We claim that condition (a) is equivalent to

(a)' For all $A \in \mathcal{B}(E)$ and $0 \leq s \leq t$,

$$\mathbf{P}(\{X_t \in A\}|\mathcal{F}^X_s) = P_{t-s}(X_s, A), \qquad \text{a.s.} \tag{2.44}$$

Indeed, the implication (a)⇒(a)' is obvious, while the converse follows by approximation with simple functions. Therefore the statement follows directly from Proposition 2.8. ∎

Proposition 2.12 (Construction from semigroup)**.** *Let E be a Polish space, $(P_t)_{t\geq 0}$ a transition function on E, and μ a probability measure on E. Then there exists a stochastic process X, unique in finite dimensional distributions, such that X satisfies the equivalent conditions (a)–(c) from Proposition 2.11.*

Proof. By condition (c) from Proposition 2.11, it suffices to show that there exists a stochastic process X with finite dimensional distributions given by

$$\begin{aligned}&\mathbf{P}\{X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n\}\\ &\quad = \int \mu(\mathrm{d}x_0) \int_{A_1} P_{t_1-t_0}(x_0, \mathrm{d}x_1) \cdots \int_{A_n} P_{t_n - t_{n-1}}(x_{n-1}, \mathrm{d}x_n)\end{aligned} \tag{2.45}$$

for all $n \in \mathbb{N}$, $0 = t_0 \leq t_1 \leq \cdots \leq t_n$ and $A_1, \ldots, A_n \in \mathcal{B}(E)$. By the Chapman Kolmogorov equation for transition functions, these finite dimensional distributions are consistent in the sense of Theorem 1.2, so there exists a stochastic process $(X_t)_{t\geq 0}$ with the finite dimensional distributions in (2.45). ∎

If X is a stochastic process with the Markov property and there exists a transition function $(P_t)_{t\geq 0}$ such that X satisfies the equivalent conditions (a)–(c) from Proposition 2.11, then we say that X is *time-homogeneous*. Note that by (c), the finite dimensional distributions of X are uniquely determined by $\mathcal{L}(X_0)$ and $(P_t)_{t\geq 0}$. We call X the *Markov process* with semigroup $(P_t)_{t\geq 0}$, started in the initial law $\mathcal{L}(X_0)$.

Example. (Transition function of Brownian motion). Very often it is not possible to give the transition function of a process explicitly. An exception is Brownian motion. Here, for all $0 \leq s \leq t$, $x \in \mathbb{R}$, and $A \in \mathcal{B}(\mathbb{R})$,

$$P_{s,t}(x, A) = \frac{1}{\sqrt{2\pi t}} \int_A \mathrm{d}y \, \exp\big[- \frac{(x-y)^2}{2(t-s)} \big]. \tag{2.46}$$

∎

Exercise 2.13 (Time reversal and time-homogeneity)**.** *Let $T > 0$ and let $(X_t)_{t\in[0,T]}$ be a stochastic process with index set $[0, T]$. How would you define the Markov property for such a process? Show that if $(X_t)_{t\in[0,T]}$ has the Markov property then the* time-reversed process $(X_{T-t})_{t\in[0,T]}$ *also has the Markov property. If $(X_t)_{t\in[0,T]}$ is time-homogeneous, then is $(X_{T-t})_{t\in[0,T]}$ in general also time-homogeneous? (Hint: it may be easier to investigate the latter question for Markov chains $(X_i)_{i\in\{0,\dots,n\}}$.)*

2.4. **Forward and backward equations.** In Proposition 2.12 we have seen that for a given initial law $\mathcal{L}(X_0)$ and transition function (Markov semigroup) $(P_t)_{t\geq 0}$, there exists a Markov process X, which is unique in finite dimensional distributions. There are two reasons why we are not satisfied with this result. The first reason is that Proposition 2.12 says nothing about the sample paths of X, which we would like to be cadlag. The second reason is that Proposition 2.12 says nothing about how to construct transition functions $(P_t)_{t\geq 0}$ in the first place. Examples such as (2.46) where we can explicitly write down a transition function are rare. There are basically two more general approaches towards obtaining transition functions.

Identify probability kernels K on E with operators $K : B(E) \to B(E)$ as in Lemma 2.6 and probability measures μ on E with functions $\mu : B(E) \to \mathbb{R}$. In a first attempt to obtain a transition function $(P_t)_{t\geq 0}$, we fix $x \in E$, and we consider the probability measures

$$\mu_t := P_t(x, \cdot) \qquad (t \geq 0).$$

Then

$$\mu_t f = \int P_t(x, \mathrm{d}y) f(y) = P_t f(x) \qquad (t \geq 0)$$

and therefore

$$\mu_{t+\varepsilon}f = P_{t+\varepsilon}f(x) = P_tP_\varepsilon f(x) = \mu_tP_\varepsilon f \qquad (t,\varepsilon \geq 0).$$

Therefore, we can try to define an operator H, acting on probability measures, by

$$H\mu := \lim_{\varepsilon\to 0}\varepsilon^{-1}\big(\mu P_\varepsilon - \mu\big),$$

and then try to solve

$$\begin{cases} \frac{\partial}{\partial t}\mu_t = H\mu_t, \\ \mu_0 = \delta_x \end{cases} \tag{2.47}$$

for fixed $x \in E$. Equation (2.47) is called the *forward equation*.

In the second approach, we fix $f \in B(E)$, and consider the functions

$$u_t := P_tf \qquad (t \geq 0).$$

Then

$$u_{t+\varepsilon} = P_{t+\varepsilon}f = P_\varepsilon P_tf = P_\varepsilon u_t \qquad (t,\varepsilon \geq 0).$$

Therefore, we can try to define an operator G, acting on functions f, by

$$Gf := \lim_{\varepsilon\to 0}\varepsilon^{-1}\big(P_\varepsilon f - f\big),$$

and then try to solve

$$\begin{cases} \frac{\partial}{\partial t}u_t = Gu_t, \\ u_0 = f \end{cases} \tag{2.48}$$

for fixed $f \in B(E)$. Equation (2.48) is called the *backward equation*.

3. Feller semigroups

3.1. **Weak convergence.** Let E be a Polish space. By definition,

$$\mathcal{C}_{\mathrm{b}}(E) := \{f : f : E \to \mathbb{R} \text{ bounded continuous}\} \tag{3.1}$$

is the space of all bounded continuous real-valued functions on E. We equip $\mathcal{C}_{\mathrm{b}}(E)$ with the *supremumnorm*

$$\|f\| := \sup_{x \in E} |f(x)|. \tag{3.2}$$

With this norm, $\mathcal{C}_{\mathrm{b}}(E)$ is a *Banach space.* If E is compact then every continuous function is bounded so we simply write $\mathcal{C}(E) = \mathcal{C}_{\mathrm{b}}(E)$. In this case $\mathcal{C}(E)$ is moreover separable. By definition

$$\mathcal{M}_1(E) := \{\mu : \mu \text{ probability measure on } (E, \mathcal{B}(E))\}. \tag{3.3}$$

is the space of all probability measures on E. We equip $\mathcal{M}_1(E)$ with the topology of weak convergence. We say that a sequence of measures $\mu_n \in \mathcal{M}_1(E)$ converges weakly to a limit $\mu \in \mathcal{M}_1(E)$, denoted as $\mu_n \Rightarrow \mu$, if

$$\mu_n f \underset{n\to\infty}{\longrightarrow} \mu f \qquad \forall f \in \mathcal{C}_{\mathrm{b}}(E). \tag{3.4}$$

(Recall the notation $\mu f := \int f \mathrm{d}\mu$ from (2.26).) This notion of convergence indeed comes from a topology.

Proposition 3.1 (Prohorov metric)**.** *Let (E, d) be a separable metric space. For any $A \subseteq E$ and $r > 0$, put $A^r := \{x \in E : \inf_{y\in A} d(x, y) < r\}$. Then*

$$\begin{aligned} d_{\mathrm{Pr}}(\mu_1, \mu_2) &:= \inf\{r > 0 : \mu_1(A) \le \mu_2(A^r) + r \ \forall A \subseteq E \ \textit{closed}\} \\ &= \inf\{r > 0 : \exists \mu \in \mathcal{M}_1(E \times E) \ \textit{s.t.} \\ &\qquad \mu(A \times E) = \mu_1(A),\ \mu(E \times A) = \mu_2(A) \ \forall A \in \mathcal{B}(E), \\ &\qquad \mu(\{(x_1, x_2) \in E \times E : d(x_1, x_2) \ge r\}) \le r\} \end{aligned} \tag{3.5}$$

defines a metric on $\mathcal{M}_1(E)$ generating the topology of weak convergence. The space $(\mathcal{M}_1(E), d_{\mathrm{Pr}})$ is separable. If (E, d) is complete, then $(\mathcal{M}_1(E), d_{\mathrm{Pr}})$ is complete.

Proof. See [EK86, Theorems 3.1.2, 3.1.7, and 3.3.1]. ∎

The second formula for d_{Pr} in (3.5) says that

$$\begin{aligned} d_{\mathrm{Pr}}(\mu_1, \mu_2) = \inf\{r > 0 : &\mathbf{P}\{d(X_1, X_2) \ge r\} \le r, \\ &\mathcal{L}(X_1) = \mu_1,\ \mathcal{L}(X_1) = \mu_1\}, \end{aligned} \tag{3.6}$$

where the infimum is over all pairs of random variables (X_1, X_2) with laws μ_1 and μ_2, respectively.

Formula (3.4) shows that the topology of weak convergence on $\mathcal{M}_1(E)$ does not depend on the choice of the metric on E. In other words, if $d, \tilde{d}$ are equivalent metrics on E and d_{Pr} and $\tilde{d}_{\mathrm{Pr}}$ are the associated Prohorov metrics on $\mathcal{M}_1(E)$, then d_{Pr} and $\tilde{d}_{\mathrm{Pr}}$ are equivalent. Proposition 3.1 moreover shows that $\mathcal{M}_1(E)$ is Polish if E is Polish.

The next proposition can be found in [EK86, Theorem 3.2.2].

Proposition 3.2 (Prohorov)**.** *Let E be Polish. Then a set $\mathcal{K} \subseteq \mathcal{M}_1(E)$ is compact if and only if $\mathcal{K}$ is closed and*

$$\forall \varepsilon > 0\ \exists K \subset E \text{ compact s.t. } \mu(E \backslash K) \leq \varepsilon\ \forall \mu \in \mathcal{K}. \tag{3.7}$$

Property (3.7) is called the *tightness* if the set $\mathcal{K}$. Note that Proposition 3.2 implies in particular that $\mathcal{M}_1(E)$ is compact if E is compact.

Exercise 3.3. *Let E be Polish. Show $\mathcal{K} \subset \mathcal{M}_1(E)$ tight $\Rightarrow \overline{\mathcal{K}}$ compact.*

3.2. **Continuous kernels and Feller semigroups.** For a proof of the following proposition, see for example [RS80, Theorem IV.14].

Proposition 3.4 (Probability measures as positive linear forms)**.** *Let E be compact and metrizable. A probability measure $\mu \in \mathcal{M}_1(E)$ defines through (2.26) a function $\mu : \mathcal{C}(E) \to \mathbb{R}$ with the following properties*

1. *(normalization) $\mu 1 = 1$.*
2. *(positivity) $\mu f \geq 0$ for all $f \geq 0$.*
3. *(linearity) $\mu(\lambda_1 f_1 + \lambda_2 f_2) = \lambda_1 \mu(f_1) + \lambda_2 \mu(f_2)$ for all $\lambda_1, \lambda_2 \in \mathbb{R},\ f_1, f_2 \in \mathcal{C}_{\rm b}(E)$.*

Conversely, each function $\mu : \mathcal{C}(E) \to \mathbb{R}$ with these properties corresponds through (2.26) to a probability measure $\mu \in \mathcal{M}_1(E)$.

Let E, F be *compact* metrizable spaces and let $\mathcal{M}_1(E), \mathcal{M}_1(F)$ be the spaces of probability measures on E and F, respectively, equipped with the topology of weak convergence. By definition, a probability kernel K from E to F is *continuous* if the map $x \mapsto K(x, \cdot)$ from E to $\mathcal{M}_1(F)$ is continuous.

Proposition 3.5 (Continuous probability kernels)**.** *A continuous probability kernel K from E to F defines through (2.25) an operator $K : \mathcal{C}(F) \to \mathcal{C}(E)$ with the following properties*

1. *(conservativeness) $K1 = 1$.*
2. *(positivity) $Kf \geq 0$ for all $f \geq 0$.*
3. *(linearity) $K(\lambda_1 f_1 + \lambda_2 f_2) = \lambda_1 K(f_1) + \lambda_2 K(f_2)$ for all $\lambda_1, \lambda_2 \in \mathbb{R},\ f_1, f_2 \in \mathcal{C}(E)$.*

Conversely, each operator $K : \mathcal{C}(F) \to \mathcal{C}(E)$ with these properties corresponds through (2.25) to a continuous probability kernel K from E to F.

Proof. By Proposition 3.4, the properties (1)–(4) from Proposition 3.5 are equivalent to the statement that for fixed $x \in E$, $K(x, \cdot)$ is a probability measure on F. (Note that tightness is automatic since F is compact.) The statement that K maps $\mathcal{C}(F)$ into $\mathcal{C}(E)$ means that $\int K(x_n, \mathrm{d}y) f(y) \to \int K(x, \mathrm{d}y) f(y)$ whenever $x_n \to x$ and $f \in \mathcal{C}(F)$. This is equivalent to the statement that $K(x_n, \cdot) \Rightarrow K(x, \cdot)$ whenever $x_n \to x$, i.e., $x \mapsto K(x, \cdot)$ is continuous. ∎

Exercise 3.6. *Show that properties (1)–(3) from Proposition 3.5 imply that $K : \mathcal{C}(F) \to \mathcal{C}(E)$ is continuous, i.e., $Kf_n \to Kf$ whenever $\|f_n - f\| \to 0$.*

It is easy to see (for example from Proposition 3.5) that the composition (in the sense of (2.27)) of two continuous probability kernels is again continuous.

Let E be a *compact* metrizable space. By definition, we say that a transition probability $(P_t)_{t\geq 0}$ on E is *continuous* if the map $(t,x) \mapsto P_t(x,\cdot)$ from $[0,\infty) \times E$ into $\mathcal{M}_1(E)$ is continuous. Here we equip $[0,\infty)\times E$ with the product topology and $\mathcal{M}_1(E)$ with the topology of weak convergence.

Proposition 3.7 (Feller semigroups)**.** *Let $(P_t)_{t\geq 0}$ be a continuous transition probability on E. Then the operators $(P_t)_{t\geq 0}$ defined in (2.40) map $\mathcal{C}(E)$ into $\mathcal{C}(E)$ and, considered as operators from $\mathcal{C}(E)$ into $\mathcal{C}(E)$, they satisfy*

1. *P_t is conservative for each $t \geq 0$, i.e., $P_t 1 = 1$.*
2. *P_t is positive for each $t \geq 0$, i.e., $P_t f \geq 0$ for all $f \geq 0$.*
3. *P_t is linear for each $t \geq 0$.*
4. *The $(P_t)_{t\geq 0}$ form a semigroup, i.e., $P_0 f = f$ for all $f \in \mathcal{C}(E)$ and $P_s P_t = P_{s+t}$ for all $s,t \geq 0$.*
5. *$(P_t)_{t\geq 0}$ is strongly continuous, i.e., $\lim_{t\to 0} \|P_t f - f\| = 0$ for all $f \in \mathcal{C}(E)$.*

Conversely, each collection of operators $(P_t)_{t\geq 0}$ from $\mathcal{C}(E)$ into $\mathcal{C}(E)$ with these properties corresponds through (2.40) to a continuous transition probability on E.

A collection of operators $(P_t)_{t\geq 0}$ from $\mathcal{C}(E)$ into $\mathcal{C}(E)$ with the properties (1)–(5) from Proposition 3.7 is called a *Feller semigroup*.

Proof of Proposition 3.7. By the definition of weak convergence of probability measures, a transition probability $(P_t)_{t\geq 0}$ on E is continuous if and only if the function $(t,x) \mapsto P_t f(x)$ from $[0,\infty)\times E$ into $\mathbb{R}$ is continuous for each $f \in \mathcal{C}(E)$. We claim that this is equivalent to the statement that $P_t f \in \mathcal{C}(E)$ for all $t \geq 0$ and

(5)' $\lim_{s\to t} \|P_s f - P_t f\| = 0$ for all $f \in \mathcal{C}(E)$, $t \geq 0$.

Assume that $P_t f \in \mathcal{C}(E)$ for all $t \geq 0$ and (5)' holds. Choose $(t_n, x_n) \to (t,x)$. Then

$$\begin{aligned} |P_{t_n} f(x_n) - P_t f(x)| &\leq |P_{t_n} f(x_n) - P_t f(x_n)| + |P_t f(x_n) - P_t(x)| \\ &\leq \|P_{t_n} f - P_t f\| + |P_t f(x_n) - P_t(x)| \underset{n\to\infty}{\longrightarrow} 0, \end{aligned} \tag{3.8}$$

which shows that $(t,x) \mapsto P_t f(x)$ is continuous. Conversely, if $(t,x) \mapsto P_t f(x)$ is continuous then obviously we must have $P_t f \in \mathcal{C}(E)$ for all $t \geq 0$. Now assume that (5)' does not hold. Then we can find $\varepsilon > 0$, $t_n \to t$, and $x_n \in E$ such that

$$|P_{t_n} f(x_n) - P_t f(x_n)| \geq \varepsilon. \tag{3.9}$$

Since E is compact, we can choose a convergent subsequence $x_{n_m} \to x$. Then $(t_{m_n}, x_{m_n}) \to (t,x)$ but since

(3.10)
$$|P_{t_{m_n}} f(x_{m_n}) - P_t f(x)| \geq |P_{t_{m_n}} f(x_{m_n}) - P_t f(x_{m_n})| - |P_t f(x_{m_n}) - P_t f(x)|$$

we have $\liminf_{n\to\infty} |P_{t_{m_n}} f(x_{m_n}) - P_t f(x)| \geq \varepsilon$ by te continuity of $P_t f$, which shows that $P_{t_{m_n}} f(x_{m_n}) \not\to P_t f(x)$, i.e., $(t,x) \mapsto P_t f(x)$ is not continuous. (Note that this is very similar to the proof of Lemma 1.7.)

It follows from Proposition 3.5 that a collection $(P_t)_{t\geq 0}$ of operators on $\mathcal{C}(E)$ satisfying (1)–(4) corresponds to a transition probability on E with the property that P_t is a continuous probability kernel for each fixed $t \geq 0$.

It therefore suffices to show that (5) is equivalent to (5)'. The implication (5)'⇒(5) is trivial. Conversely, if (5) holds then

$$\lim_{t_n \downarrow t} \|P_{t_n} f - P_t f\| = \lim_{t_n \downarrow t} \|P_{t_n - t}(P_t f) - (P_t f)\| = 0 \tag{3.11}$$

by the semigroup property and (5) applied to $P_t f$. This shows that $t \mapsto P_t f$, considered as a function from $[0,\infty)$ into $\mathcal{C}(E)$, is continuous from the right. To prove also continuity from the left, we note that

$$\lim_{t_n \uparrow t} \|P_{t_n} f - P_t f\| = \lim_{t_n \uparrow t} \|P_{t_n}(f - P_{t-t_n} f)\| \leq \lim_{t_n \uparrow t} \|f - P_{t-t_n} f\| = 0, \tag{3.12}$$

where we have the semigroup property, (5), and the fact that

$$\begin{aligned} \|P_t f\| &= \sup_{x\in E} \Big| \int_E P_t(x, \mathrm{d}y) f(y) \Big| \\ &\leq \sup_{x\in E} \int_E P_t(x, \mathrm{d}y) \big| f(y) \big| \leq \sup_{y\in E} |f(y)| = \|f\|. \end{aligned} \tag{3.13}$$

∎

3.3. **Banach space calculus.** Let $(V, \|\cdot\|)$ be a Banach space, equipped with the topology generated by the norm. We need to develop calculus for V-valued functions. The next proposition defines the Riemann integral for continuous V-valued functions. Since this is very similar to the usual Riemann integral, we skip the proof.

Proposition 3.8 (Riemann integral)**.** *Let $u : [a,b] \to V$ be continuous and let*

$$a = t_0^{(n)} \leq s_1^{(n)} \leq t_t^{(n)} \leq \cdots \leq t_{m_n-1}^{(n)} \leq s_{m_n}^{(n)} \leq t_{m_n}^{(n)} = b \tag{3.14}$$

satisfy

$$\lim_{n\to\infty} \sup\{t_k^{(n)} - t_{k-1}^{(n)} : k = 1, \ldots, m_n\} = 0. \tag{3.15}$$

Then the limit

$$\int_a^b u(t)\,\mathrm{d}t := \lim_{n\to\infty} \sum_{k=1}^{m_n} u(s_k^{(n)})(t_k^{(n)} - t_{k-1}^{(n)}) \tag{3.16}$$

exists and does not depend on the choice of the $t_k^{(n)}$ and $s_k^{(n)}$.

If $a < b \leq \infty$ and $u : [a, b) \to V$ is continuous then we define

$$\int_a^b u(t)\,\mathrm{d}t := \lim_{c \uparrow b} \int_a^c u(t)\,\mathrm{d}t, \tag{3.17}$$

whenever the limit exists. In this case we say that u is integrable over $[a, b)$. In case $b < \infty$ and $u : [a, b] \to V$ is continuous this coincides with our earlier definition of $\int_a^b u(t)\,\mathrm{d}t$.

Lemma 3.9 (Infinite integrals)**.** *Let $a < b \leq \infty$, let $u : [a, b) \to V$ be continuous and $\int_a^b \|u(t)\|\,\mathrm{d}t < \infty$. Then u is integrable over $[a, b)$ and*

$$\Big\| \int_a^b u(t)\,\mathrm{d}t \Big\| \leq \int_a^b \|u(t)\|\,\mathrm{d}t. \tag{3.18}$$

Proof. Since u is continuous and $f \mapsto \|f\|$ is continuous, the function $t \mapsto \|u(t)\|$ is continuous. First consider the case that $b < \infty$ and that $u : [a, b] \to V$ is continuous. Choose $t^{(n)}_k$ and $s^{(n)}_k$ as in (3.14) and (3.15). Then

$$\Big\| \sum_{k=1}^{m_n} u(s^{(n)}_k)(t^{(n)}_k - t^{(n)}_{k-1}) \Big\| \leq \sum_{k=1}^{m_n} \|u(s^{(n)}_k)\|(t^{(n)}_k - t^{(n)}_{k-1}). \tag{3.19}$$

Taking the limit $n \to \infty$ we arrive at (3.18). If $a < b \leq \infty$ and $u : [a, b) \to V$ is continuous then it follows that for each $a \leq c \leq c' < b$

$$\Big\| \int_a^{c'} u(t)\,\mathrm{d}t - \int_a^c u(t)\,\mathrm{d}t \Big\| \leq \int_c^{c'} \|u(t)\|\,\mathrm{d}t. \tag{3.20}$$

If $\int_a^\infty \|u(t)\|\,\mathrm{d}t < \infty$ and $c_i \uparrow b$ then (3.20) implies that $\big(\int_a^{c_i} u(t)\,\mathrm{d}t\big)_{i \geq 1}$ is a Cauchy sequence, and hence, by the completeness of V, u is integrable over $[a, \infty)$. Taking the limit in (3.18) we see that this estimate holds in the more general case as well. ∎

Let I be an interval. We say that a function $u : I \to V$ is continuously differentiable if for each $t \in I$ the limit

$$\tfrac{\partial}{\partial t} u(t) := \lim_{h \to 0} h^{-1}(u(t + h) - u(t)) \tag{3.21}$$

exists, and $t \mapsto \frac{\partial}{\partial t} u(t)$ is continuous on I. We skip the proof of the next result.

Proposition 3.10 (Fundamental theorem of calculus)**.** *Assume that $u : [a, b] \to V$ is continuously differentiable. Then*

$$\int_a^b \tfrac{\partial}{\partial t} u(t)\,\mathrm{d}t = u(b) - u(a). \tag{3.22}$$

So far, when we talked about a linear operator A on a normed linear space N, we always meant a linear map $A : N \to N$ that is defined on all of N. It will be convenient to generalize this definition such that operators need no longer be defined on the whole space.

Definition 3.11 (Linear operators)**.** *A* linear operator *on a normed space* $(N, \|\cdot\|)$ *is a pair* $(\mathcal{D}(A), A)$ *where* $\mathcal{D}(A) \subseteq N$ *is a linear subspace of* N *and* $A : \mathcal{D}(A) \to N$ *is a linear map. The* graph *of such a linear operator is the linear space*

$$\mathcal{G}(A) := \{(f, Af) : f \in \mathcal{D}(A)\} \subseteq N \times N. \tag{3.23}$$

We say that a linear operator is closed *if its graph* $\mathcal{G}(A)$ *is a closed subspace of* $N \times N$*, equipped with the product topology.*

Note that a linear operator (including its domain!) is uniquely characterized by its graph. In fact, every linear subspace $\mathcal{G} \subset N \times N$ with the property that

$$(f, g) \in \mathcal{G},\ (f, \tilde{g}) \in \mathcal{G} \quad \Rightarrow \quad g = \tilde{g} \tag{3.24}$$

is the graph of a linear operator $(\mathcal{D}(A), A)$.[11] Note that the fact that A is closed means that if $f_i \in \mathcal{D}(A)$ are such that $\lim_{i\to\infty} f_i =: f$ and $\lim_{i\to\infty} Af_i =: g$ exist, then $f \in \mathcal{D}(A)$ and $Af = g$.

We recall a few facts from functional analysis.

Theorem 3.12 (Closed graph theorem)**.** *Let* $(N, \|\cdot\|)$ *be a normed linear space and let* $(\mathcal{D}(A), A)$ *be a linear operator on* N *with* $\mathcal{D}(A) = N$*. Then one has the relations (a)⇔(b)⇒(c) between the statements:*

(a) *A is continuous, i.e.,* $\|Af_n - Af\| \to 0$ *whenever* $\|f_n - f\| \to 0$.
(b) *A is bounded, i.e., there exists a constant K such that* $\|Af\| \leq K\|f\|$ *for all* $f \in L$.
(c) *A is closed.*

If N is complete then all statements are equivalent.

To see that unbounded closed operators have nice properties, we prove the following fact, that will be useful later.

Lemma 3.13 (Closed operators and integrals)**.** *Let V be a Banach space and let* $(\mathcal{D}(A), A)$ *be a closed linear operator on V. Let* $a < b \leq \infty$*, let* $u : [a, b) \to V$ *be continuous,* $u(t) \in \mathcal{D}(A)$ *for all* $t \in [a, b)$*,* $t \mapsto Au(t)$ *continuous,* $\int_a^b \|u(t)\|\,\mathrm{d}t < \infty$*, and* $\int_a^b \|Au(t)\|\,\mathrm{d}t < \infty$*. Then*

$$\int_a^b u(t)\,\mathrm{d}t \in \mathcal{D}(A) \quad \textit{and} \quad A\int_a^b u(t)\,\mathrm{d}t = \int_a^b A\,u(t)\,\mathrm{d}t. \tag{3.25}$$

Proof. We first prove the statement for the case that u and Au are continuous functions on a bounded time interval $[a, b]$. Choose $t^{(n)}_k$ and $s^{(n)}_k$ as in (3.14) and (3.15). Define

$$f_n := \sum_{k=1}^{m_n} u(s^{(n)}_k)(t^{(n)}_k - t^{(n)}_{k-1}). \tag{3.26}$$

[11] Sometimes the concept of a linear operator is generalized even further in the sense that condition (3.24) is dropped. In this case, one talks about multi-valued operators.

Then $f_n \in \mathcal{D}(A)$ and

$$Af_n = \sum_{k=1}^{m_n} Au(s_k^{(n)})(t_k^{(n)} - t_{k-1}^{(n)}). \tag{3.27}$$

By our assumptions,

$$f_n \underset{n\to\infty}{\longrightarrow} \int_a^b u(t)\,\mathrm{d}t \quad \text{and} \quad Af_n \underset{n\to\infty}{\longrightarrow} \int_a^b Au(t)\,\mathrm{d}t. \tag{3.28}$$

Since A is closed, it follows that (3.25) holds. The statement for intervals of the form $[a, b)$ follows by approximation with compact intervals, again using the fact that A is closed. ∎

3.4. **Semigroups and generators.** Let $(V, \|\cdot\|)$ be a Banach space. By definition, a (linear) *semigroup* on V is a collection of everywhere defined linear operators $(S_t)_{t\geq 0}$ on V such that $S_0 f = f$ for all $f \in V$ and $S_s S_t = S_{s+t}$ for all $s, t \geq 0$. We say that $(S_t)_{t\geq 0}$ is a *contraction semigroup* if S_t is a contraction for each $t \geq 0$, i.e.,

$$\|S_t f\| \leq \|f\| \qquad \forall f \in V. \tag{3.29}$$

We say that a semigroup $(S_t)_{t\geq 0}$ is *strongly continuous* if

$$\lim_{t\to 0} S_t f = f \qquad \forall f \in V. \tag{3.30}$$

Example. A Feller semigroup on $\mathcal{C}(E)$, where E is compact and metrizable and $\mathcal{C}(E)$ is equipped with the supremumnorm, is a strongly continuous contraction semigroup. (See Proposition 3.7 and (3.13).) ∎

Remark. If $(S_t)_{t\geq 0}$ is a strongly continuous contraction semigroup on a Banach space V, then exactly the same proof as in (3.11)–(3.12) shows that $t \mapsto S_t f$ is a continuous map from $[0, \infty)$ into V for each $f \in V$. ∎

Let $(S_t)_{t\geq 0}$ be a strongly continuous contraction semigroup on a Banach space V. By definition, the *generator* of $(S_t)_{t\geq 0}$ is the linear operator $(\mathcal{D}(G), G)$, where

$$\mathcal{D}(G) := \big\{ f \in L : \lim_{t\to 0} t^{-1}(S_t f - f) \text{ exists} \big\}, \tag{3.31}$$

and

$$Gf := \lim_{t\to 0} t^{-1}(S_t f - f). \tag{3.32}$$

Exercise 3.14 (Generator of a deterministic process)**.** *Define a continuous transition probability $(P_t)_{t\geq 0}$ on $[-1, 1]$ by*

$$P_t(x, \cdot\,) := \delta_{xe^{-t}}(\cdot) \qquad (x \in [0, 1]). \tag{3.33}$$

Determine the generator $(\mathcal{D}(G), G)$ of the corresponding Feller semigroup $(P_t)_{t\geq 0}$ on $\mathcal{C}[-1, 1]$.

Proposition 3.15 (Generators). *Let V be a Banach space, let $(S_t)_{t\geq 0}$ be a strongly continuous contraction semigroup on V, and let $(\mathcal{D}(G), G)$ be its generator. Then $\mathcal{D}(G)$ is dense in V and $(\mathcal{D}(G), G)$ is closed. For each $f \in \mathcal{D}(G)$, the function $t \mapsto S_t f$ from $[0,\infty)$ to V is continuously differentiable, $S_t f \in \mathcal{D}(G)$ for all $t \geq 0$, and*

$$\tag{3.34} \tfrac{\partial}{\partial t} S_t f = G S_t f = S_t G f \qquad (t \geq 0).$$

Proof. For each $h > 0$ and $f \in V$ we have

$$\tag{3.35} h^{-1}(S_{t+h} - S_t) f = \big\{h^{-1}(S_h - S_0)\big\} S_t f = S_t \big\{h^{-1}(S_h - S_0)\big\} f.$$

If $f \in \mathcal{D}(G)$ then $\lim_{h\downarrow 0} h^{-1}(S_h - S_0) f = Gf$. Since S_t is a contraction it is continuous, so the right-hand side of (3.35) converges to $S_t G f$. It follows that the other expressions converge as well, so $S_t f \in \mathcal{D}(G)$ for all $t \geq 0$ and

$$\tag{3.36} \lim_{h\downarrow 0} h^{-1}(S_{t+h} - S_t) f = G S_t f = S_t G f \qquad (t \geq 0).$$

If $t > 0$ and $0 < h \leq t$ then

$$\tag{3.37} h^{-1}(S_t f - S_{t-h}) f = S_{t-h}\big\{h^{-1}(S_h - S_0)\big\} f \to S_t G f \quad \text{as } h \downarrow 0.$$

Here the convergence follows from the estimates

$$\tag{3.38} \begin{aligned} &\|S_{t-h}\{h^{-1}(S_h - S_0)\} f - S_t G f\| \\ &\quad \leq \|S_{t-h}\{h^{-1}(S_h - S_0)\} f - S_{t-h} G f\| + \|S_{t-h} G f - S_t G f\| \\ &\quad \leq \|\{h^{-1}(S_h - S_0)\} f - G f\| + \|S_{t-h} G f - S_t G f\|. \end{aligned}$$

Formula (3.37) shows that the time derivatives of $S_t f$ from the left exist and are equal to the derivatives from the right. It follows that $t \mapsto S_t f$ is continuously differentiable and (3.34) holds.

To prove the other statements, we start by showing that for any $f \in V$,

$$\tag{3.39} \int_0^t S_s f \,\mathrm{d}s \in \mathcal{D}(G) \quad \text{and} \quad G \int_0^t S_s f \,\mathrm{d}s = S_t f - f.$$

We have already seen that for any $f \in V$ the function $t \mapsto S_t f$ is continuous, so $\int_0^t S_s f \,\mathrm{d}s$ is well-defined. For each $t \geq 0$, S_t is a contraction, hence continuous, hence closed, so by Lemma 3.13

$$\tag{3.40} \begin{aligned} &h^{-1}(S_h - S_0) \int_0^t S_s f \,\mathrm{d}s = h^{-1} \int_0^t \big(S_{s+h} - S_s\big) f \,\mathrm{d}s \\ &\quad = h^{-1}\Big\{ \int_h^{t+h} S_s f \,\mathrm{d}s - \int_0^t S_s f \,\mathrm{d}s \Big\} = h^{-1} \int_t^{t+h} S_s f \,\mathrm{d}s - h^{-1} \int_0^h S_s f \,\mathrm{d}s. \end{aligned}$$

Letting $h \to 0$ we arrive at (3.39).

Since for each $f \in V$

$$\tag{3.41} \lim_{t\downarrow 0} t^{-1} \int_0^t S_s f \,\mathrm{d}s = f,$$

formula (3.39) shows that $\mathcal{D}(G)$ is dense in V. To show that $(\mathcal{D}(G), G)$ is closed, choose $f_n \in \mathcal{D}(G)$ such that $\lim_{n\to\infty} f_n =: f$ and $\lim_{n\to\infty} Gf_n =: g$ exist. By (3.34) and the fundamental theorem of calculus,

$$S_t f_n - f_n = \int_0^t S_s G f_n \,\mathrm{d}s \qquad (t > 0). \tag{3.42}$$

Letting $n \to \infty$, using the fact that $\|S_s(Gf_n - f_n)\| \leq \|Gf_n - f_n\|$ for each $s \geq 0$, we find that

$$S_t f - f = \int_0^t S_s g \,\mathrm{d}s \qquad (t > 0). \tag{3.43}$$

Dividing by t and letting $t \to 0$ we conclude that $f \in \mathcal{D}(G)$ and $Gf = g$. ∎

3.5. **Dissipativity and the maximum principle.** Let E be a compact metrizable space. We say that a linear operator $(\mathcal{D}(A), A)$ on $\mathcal{C}(E)$ satisfies the *positive maximum principle* if

$$Af(x) \leq 0 \text{ whenever } f(x) \geq 0 \text{ and } f(y) \leq f(x) \ \forall y \in E. \tag{3.44}$$

This says that $Af(x) \leq 0$ whenever f assumes a positive maximum over E in x.

Proposition 3.16 (Generators of Feller semigroups)**.** *Let E be compact and metrizable, let $(P_t)_{t\geq 0}$ be a Feller semigroup on $\mathcal{C}(E)$, and let $(\mathcal{D}(G), G)$ be its generator. Then*

1. *$1 \in \mathcal{D}(G)$ and $G1 = 0$.*
2. *$\mathcal{D}(G)$ is dense in $\mathcal{C}(E)$.*
3. *$(\mathcal{D}(G), G)$ is closed.*
4. *$(\mathcal{D}(G), G)$ satisfies the positive maximum principle.*

Proof. Property (1) follows from the fact that $P_t 1 = 1$ for all $t \geq 0$. Properties (2)–(3) follow from Proposition 3.15. To prove (4), assume that $f \in \mathcal{D}(G)$, $x \in E$, $f(x) \geq 0$, and $f(y) \leq f(x)$ $\forall y \in E$. Then

$$P_t f(x) = \int P_t(x, \mathrm{d}y) f(y) \leq f(x) \qquad (t \geq 0), \tag{3.45}$$

and therefore

$$Gf(x) := \lim_{t\to 0} t^{-1}(P_t f(x) - f(x)) \leq 0. \tag{3.46}$$

(Note that the limit exists by our assumption that $f \in \mathcal{D}(G)$.) ∎

Generators of strongly continuous contraction semigroups have an important property that we have not mentioned so far.

Definition 3.17. *A linear operator $(\mathcal{D}(A), A)$ on a Banach space V is called* dissipative *if $\|(\lambda - A)f\| \geq \lambda\|f\|$ for every $f \in \mathcal{D}(A)$ and $\lambda > 0$.*

Note that an equivalent formulation of dissipativity is that

$$\|f\| \le \|(1-\varepsilon A)f\| \qquad (f \in \mathcal{D}(A),\ \varepsilon > 0). \tag{3.47}$$

This follows by setting $\varepsilon = \lambda^{-1}$ in $\|(\lambda - A)f\| \ge \lambda\|f\|$ and multiplying both sides of the inequality by ε.

Lemma 3.18 (Contractions and dissipativity)**.** *If C is an (everywhere defined) contraction and $r > 0$ then $r(C-1)$ is dissipative.*

Proof. If C is a contraction then for each $\varepsilon > 0$ and $f \in V$, one has $\|(1 - \varepsilon r(C-1))f\| = \|f - \varepsilon rCf + \varepsilon rf\| \ge (1 + r\varepsilon)\|f\| - r\varepsilon\|Cf\| \ge \|f\|$. ∎

Lemma 3.19 (Maximum principle and dissipativity)**.** *Let E be compact and metrizable and let $(\mathcal{D}(A), A)$ be a linear operator on $\mathcal{C}(E)$. If A satisfies the positive maximum principle, then A is dissipative.*

Proof. Assume that $f \in \mathcal{D}(A)$. Since E is compact there exists an $x \in E$ with $|f(x)| \ge |f(y)|$ for all $y \in E$. If $f(x) \ge 0$, then by the positive maximum principle $Af(x) \le 0$ and therefore $\|(1-\varepsilon A)f\| \ge |f(x) - \varepsilon Af(x)| \ge f(x) = \|f\|$. If $f(x) \le 0$, then by the fact that A is linear also $-f \in \mathcal{D}(A)$ and $\|(1-\varepsilon A)f\| = \|(1-\varepsilon A)(-f)\| \ge \|-f\| = \|f\|$. ∎

Exercise 3.20. *Let $(\mathcal{D}(A_{\rm WF}), A_{\rm WF})$ be the operator on $\mathcal{C}[0,1]$ given by*

$$\begin{aligned} \mathcal{D}(A_{\rm WF}) &:= \mathcal{C}^2[0,1], \\ A_{\rm WF} f(x) &:= \tfrac{1}{2}x(1-x)\tfrac{\partial^2}{\partial x^2} f(x), \qquad x \in [0,1]. \end{aligned} \tag{3.48}$$

Show that $A_{\rm WF}$ satisfies the positive maximum principle. For which values of c does the operator

$$\tfrac{1}{2}x(1-x)\tfrac{\partial^2}{\partial x^2} + c(\tfrac{1}{2} - x)\tfrac{\partial}{\partial x} \tag{3.49}$$

satisfy the positive maximum principle?

Lemma 3.21 (Laplace equation and dissipativity)**.** *Let V be a Banach space, let $(S_t)_{t\ge 0}$ be a strongly continuous contraction semigroup on V, and let $(\mathcal{D}(G), G)$ be its generator. Then G is dissipative. Moreover, for each $\lambda > 0$ and $f \in V$, the* Laplace equation

$$p \in \mathcal{D}(G) \quad \textit{and} \quad (\lambda - G)p = f \tag{3.50}$$

has a unique solution, which is given by

$$p = \int_0^\infty S_t f e^{-\lambda t}\,\mathrm{d}t. \tag{3.51}$$

Proof. For each $\lambda > 0$, define $U_\lambda : V \to V$ by

$$U_\lambda f := \int_0^\infty S_t f e^{-\lambda t}\,\mathrm{d}t. \tag{3.52}$$

Since $\int_0^\infty e^{-\lambda t}\,\mathrm{d}t = \lambda^{-1}$ we have

$$\|U_\lambda f\| \le \lambda^{-1}\|f\| \qquad (\lambda > 0,\ f \in V). \tag{3.53}$$

Since (compare (3.40))

$$h^{-1}(S_h - S_0)U_\lambda f = h^{-1}\int_0^\infty (S_{t+h} - S_t)fe^{-\lambda t}\,\mathrm{d}t$$
$$= h^{-1}(e^{\lambda h} - 1)\int_0^\infty S_t f e^{-\lambda t}\mathrm{d}t - h^{-1}e^{\lambda h}\int_0^h S_t f e^{-\lambda t}\mathrm{d}t, \tag{3.54}$$

letting $h \to 0$ we find that $U_\lambda f \in \mathcal{D}(G)$ and $GU_\lambda f = \lambda U_\lambda f - f$, i.e.,

$$(\lambda - G)U_\lambda f = f \qquad (\lambda > 0,\ f \in V). \tag{3.55}$$

If $f \in \mathcal{D}(G)$, then, using Lemma 3.13 and Proposition 3.15,

$$U_\lambda G f = \int_0^\infty S_t G f e^{-\lambda t}\,\mathrm{d}t = \int_0^\infty G S_t f e^{-\lambda t}\,\mathrm{d}t = G\int_0^\infty S_t f e^{-\lambda t}\,\mathrm{d}t = GU_\lambda f, \tag{3.56}$$

so that by (3.55) we also have

$$U_\lambda(\lambda - G)f = f \qquad (\lambda > 0,\ f \in \mathcal{D}(G)). \tag{3.57}$$

It follows that $(\lambda - G)$ is a bijection from $\mathcal{D}(G)$ to V and that U_λ is its inverse. Now (3.53) implies that

$$\|f\| = \|U_\lambda(\lambda - G)f\| \leq \lambda^{-1}\|(\lambda - G)f\| \qquad (\lambda > 0,\ f \in \mathcal{D}(G)), \tag{3.58}$$

which shows that G is dissipative. ∎

Lemma 3.22 (Cauchy equation)**.** *Let $(\mathcal{D}(A), A)$ be a dissipative linear operator on a Banach space V. Assume that $f \in \mathcal{D}(A)$, $u : [0,\infty) \to V$ is continuously differentiable, $u(t) \in \mathcal{D}(A)$ for all $t \geq 0$, and that u solves the* Cauchy equation

$$\begin{cases} \frac{\partial}{\partial t}u(t) = Au(t) & (t \geq 0), \\ u(0) = f. \end{cases} \tag{3.59}$$

Then $\|u(t)\| \leq \|f\|$ for all $t \geq 0$. In particular, by linearity, solutions to the Cauchy equation (3.59) are unique.

Proof. Since A is dissipative, by (3.47), $\|f\| \leq \|(1 - tA)f\|$ for all $f \in \mathcal{D}(A)$ and $t > 0$, so

$$\begin{aligned} &\|u(t)\| - \|u(0)\| \\ &\quad \leq \|(1 - tA)u(t)\| - \|u(t) - (u(t) - u(0))\| \\ &\quad = \|u(t) - tAu(t)\| - \Big\|u(t) - \int_0^t Au(s)\,\mathrm{d}s\Big\| \\ &\quad \leq \Big\|tAu(t) - \int_0^t Au(s)\,\mathrm{d}s\Big\| \leq \int_0^t \|Au(t) - Au(s)\|\mathrm{d}s. \end{aligned} \tag{3.60}$$

Choose $0 = t_0^{(n)} \leq t_1^{(n)} \leq \cdots \leq t_{m_n}^{(n)} = t$ such that $\lim_{n\to\infty} \sup\{t_k^{(n)} - t_{k-1}^{(n)} : k = 1, \ldots, m_n\} = 0$. Applying (3.60) to $\|u(t_k^{(n)})\| - \|u(t_{k-1}^{(n)})\|$ we see that

$$\begin{aligned} \|u(t)\| - \|u(0)\| &= \sum_{k=1}^{m_n} \|u(t_k^{(n)})\| - \|u(t_{k-1}^{(n)})\| \\ &\leq \sum_{k=1}^{m_n} \int_{t_{k-1}^{(n)}}^{t_k^{(n)}} \Big\| Au(t_k^{(n)}) - Au(s) \Big\| \, \mathrm{d}s. \end{aligned} \tag{3.61}$$

It is not hard to check that

$$\lim_{n\to\infty} \sup_{k=1,\ldots,m_n} \sup_{t_{k-1}^{(n)} \leq s < t_k^{(n)}} \|Au(t_k^{(n)}) - Au(s)\| = 0, \tag{3.62}$$

so the right-hand side of (3.61) tends to zero as $n \to \infty$. ∎

Lemma 3.22 has two useful corollaries.

Corollary 3.23 (Generator characterizes semigroup)**.** *Let V be a Banach space. If two strongly continuous contraction semigroups on V have the same generator, then they are equal.*

Proof. Let $(S_t)_{t\geq 0}$ and $(\tilde{S}_t)_{t\geq 0}$ be strongly continuous contraction semigroups on V with the same generator $(\mathcal{D}(G), G)$. By Proposition 3.15, for each $f \in \mathcal{D}(G)$, the functions $u(t) := S_t f$ and $\tilde{u}(t) := \tilde{S}_t f$ solve the Cauchy equation

$$\begin{cases} \frac{\partial}{\partial t} u(t) = Gu(t) & (t \geq 0), \\ u(0) = f. \end{cases} \tag{3.63}$$

By Lemmas 3.21 and 3.22, $S_t f = \tilde{S}_t f$ for all $t \geq 0$, $f \in \mathcal{D}(G)$. By Proposition 3.15, $\mathcal{D}(G)$ is dense in V so using the continuity of S_t and $\tilde{S}_t$ we find that $S_t f = \tilde{S}_t f$ for all $t \geq 0$, $f \in V$. ∎

Remark. An alternative proof of Corollary 3.23 uses the fact that the Laplace equation (3.50) has a unique solution. ∎

Corollary 3.24 (Bounded generators)**.** *Let V be a Banach space and let $A : V \to V$ be a bounded dissipative linear operator. Then A generates a strongly continuous contraction semigroup $(S_t)_{t\geq 0}$ on V, which is given by*

$$S_t f = e^{At} f := \sum_{n=0}^{\infty} \frac{1}{n!} (At)^n f \qquad (t \geq 0). \tag{3.64}$$

Proof. Using the fact that A is bounded it is not hard to prove that the infinite sequence in (3.64) converges, defines a strongly continuous semigroup $(S_t)_{t\geq 0}$ on V, and that $S_t f$ solves the Cauchy equation

$$\begin{cases} \frac{\partial}{\partial t} u(t) = Au(t) & (t \geq 0), \\ u(0) = f. \end{cases} \tag{3.65}$$

It follows from Lemma 3.22 that $(S_t)_{t\geq 0}$ is a contraction semigroup. ∎

Exercise 3.25. *Let E be compact and metrizable, let K be a continuous probability kernel on E, and $r \geq 0$ a constant. Define an (everywhere defined) linear operator on $\mathcal{C}(E)$ by*

$$Gf := r(Kf - f) \qquad (f \in \mathcal{C}(E)). \tag{3.66}$$

Show that G generates a Feller semigroup. How would you describe the corresponding Markov process on E?

3.6. **Hille-Yosida: different formulations.** By definition, the *range* of a linear operator $(A, \mathcal{D}(A))$ on a Banach space V is the space

$$\mathcal{R}(A) := \{Af : f \in \mathcal{D}(A)\}. \tag{3.67}$$

Here is a version of the celebrated Hille-Yosida theorem:

Theorem 3.26 (Hille-Yosida)**.** *A linear operator $(\mathcal{D}(G), G)$ on a Banach space V is the generator of a strongly continuous contraction semigroup if and only if*

1. *$\mathcal{D}(G)$ is dense.*
2. *G is dissipative.*
3. *There exists a $\lambda > 0$ such that $\mathcal{R}(\lambda - G) = V$.*

Note that condition (3) says that there exists a $\lambda > 0$ such that for each $f \in V$, there exists a solution $p \in \mathcal{D}(G)$ to the Laplace equation $(\lambda - G)p = f$. Thus, the necessity of the conditions (1)–(3) follows from Proposition 3.15 and Lemma 3.21.

Before we turn to the proof of Theorem 3.26, we first discuss some of its merits, drawbacks, and consequences. The Hille-Yosida theorem is actually seldomly applied in the form in which we have stated it above. The reason is that in most cases of interest, the domain $\mathcal{D}(G)$ of the generator of the semigroup that one is interested in is not known explicitly. Rather, one knows the action of G on certain well-behaved elements of the Banach space (for example sufficiently differentiable functions) and wishes to extend this action to a generator of a strongly continuous semigroup. Since generators are always closed (recall Proposition 3.15) one is naturally led to the following definition.

Definition 3.27. *Let $(\mathcal{D}(A), A)$ be a linear operator on a Banach space V and let $\mathcal{G} = \mathcal{G}(A) := \{(f, Af) : f \in \mathcal{D}(A)\}$ be its graph. Let $\overline{\mathcal{G}}$ denote the closure of $\mathcal{G}$ in $V \times V$, equipped with the product topology. If $\overline{\mathcal{G}}$ is itself the graph of a linear operator $(\mathcal{D}(\overline{A}), \overline{A})$ then we say that $(\mathcal{D}(A), A)$ is* closable *and we call $(\mathcal{D}(\overline{A}), \overline{A})$ the* closure *of $(\mathcal{D}(A), A)$.*

Here is the form of the Hille-Yosida theorem that it is usually applied in:

Theorem 3.28 (Hille-Yosida, second version). *A linear operator* $(\mathcal{D}(A), A)$ *on a Banach space* V *is closable and its closure generates a strongly continuous contraction semigroup if and only if*

(1) $\mathcal{D}(A)$ *is dense.*
(2) A *is dissipative.*
(3) *There exists a* $\lambda > 0$ *such that* $\mathcal{R}(\lambda - A)$ *is dense in* V*.*

Since we are mainly interested in Feller semigroups, we will usually need the following version of Theorem 3.28:

Theorem 3.29 (Hille-Yosida for Feller semigroups). *Let* E *be compact and metrizable. A linear operator* $(\mathcal{D}(A), A)$ *on* $\mathcal{C}(E)$ *is closable and its closure* $\overline{A}$ *generates a Feller semigroup if and only if*

(1) *There exist* $f_n \in \mathcal{D}(A)$ *such that* $f_n \to 1$ *and* $Af_n \to 0$*.*
(2) $\mathcal{D}(A)$ *is dense.*
(3) A *satisfies the positive maximum principle.*
(4) *There exists a* $\lambda > 0$ *such that* $\mathcal{R}(\lambda - A)$ *is dense in* $\mathcal{C}(E)$*.*

In (1), the convergence is in $\mathcal{C}(E)$, i.e., $\|f_n - 1\| \to 0$ and $\|Af_n\| \to 0$. It suffices, of course, if $1 \in \mathcal{D}(A)$ and $A1 = 0$.

We have already seen how to check the positive maximum principle in an explicit set-up. To check that a subset of $\mathcal{C}(E)$ is dense, the next theorem is often useful. For a proof, see [RS80, Theorem IV.9].

Theorem 3.30 (Stone-Weierstrass). *Let* E *be compact and metrizable. Assume that* $\mathcal{D} \subset \mathcal{C}(E)$ *separates points and*

(1) $1 \in \mathcal{D}$.
(2) $f_1 f_2 \in \mathcal{D}$ *for all* $f_1, f_2 \in \mathcal{D}$.
(3) $\lambda_1 f_1 + \lambda_2 f_2 \in \mathcal{D}$ *for all* $f_1, f_2 \in \mathcal{D}$ *and* $\lambda_1, \lambda_2 \in \mathbb{R}$.

Then $\mathcal{D}$ *is dense in* $\mathcal{C}(E)$*.*

In view of Theorem 3.30, the Conditions (1)–(3) from Theorem 3.29 are usually easy to check. The hard condition is usually condition (4), which says that there exists a dense set $\mathcal{D} \subset \mathcal{C}(E)$ and a $\lambda > 0$ such that for each $f \in \mathcal{D}$, there exists a solution $p \in \mathcal{D}(A)$ to the Laplace equation $(\lambda - A)p = f$. It actually suffices to find solutions to a Cauchy equation. This is not easier but perhaps a bit more intuitive:

Lemma 3.31 (Cauchy and Laplace equations). *Let* $(A, \mathcal{D}(A))$ *be a densely defined dissipative linear operator on a Banach space* V*,* $f \in V$*, and assume that* $u : [0, \infty) \to V$ *is continuously differentiable,* $u(t) \in \mathcal{D}(A)$ *for all* $t \geq 0$*, and* u *solves the Cauchy equation*

$$\begin{cases} \frac{\partial}{\partial t} u(t) = Au(t) & (t \geq 0), \\ u(0) = f. \end{cases} \tag{3.68}$$

Then A *is closable and* $p := \int_0^\infty u(t) e^{-\lambda t} \mathrm{d}t$ *satisfies the Laplace equation*

$$p \in \mathcal{D}(\overline{A}) \quad \text{and} \quad (\lambda - \overline{A})p = f. \tag{3.69}$$

Proof. By Lemma 3.22, $\|u(t)\| \leq \|f\|$ for all $t \geq 0$ so $\int_0^\infty \|u(t)e^{-\lambda t}\|\mathrm{d}t < \infty$. By Lemma 3.36 below, $(A, \mathcal{D}(A))$ is closable. By Proposition 3.13, $p \in \mathcal{D}(\overline{A})$ and

$$\begin{aligned} \overline{A}p &= \int_0^\infty \overline{A}\,u(t)e^{-\lambda t}\mathrm{d}t = \int_0^\infty (\tfrac{\partial}{\partial t}u(t))e^{-\lambda t}\mathrm{d}t \\ &= \Big|_{t=0}^\infty u(t)e^{-\lambda t} - \int_0^\infty u(t)(\tfrac{\partial}{\partial t}e^{-\lambda t})\mathrm{d}t = -f + \lambda p, \end{aligned} \tag{3.70}$$

which shows that $(\lambda - \overline{A})p = f$. ∎

Exercise 3.32. *Show that the closure of the operator A_{WF} from Exercise 3.20 generates a Feller semigroup on $\mathcal{C}[0,1]$. Hint: use the space of all polynomials on $[0,1]$.*

3.7. **Dissipative operators.** Before we embark on the proofs of the various versions of the Hille-Yosida theorem we study dissipative operators in more detail. In doing so, it will be convenient to use the formalism of multi-valued operators. By definition, a multi-valued (linear) operator on a Banach space V is a linear subspace

$$\mathcal{G} \subseteq V \times V. \tag{3.71}$$

We say that $\mathcal{G}$ is *single-valued* if $\mathcal{G}$ satisfies (3.24). In this case, $\mathcal{G}$ is the graph of some linear operator $(\mathcal{D}(A), A)$ on V. We call

$$\begin{aligned} \mathcal{D}(\mathcal{G}) &:= \{f : \exists g \in V \text{ s.t. } (f,g) \in \mathcal{G}\}, \\ \mathcal{R}(\mathcal{G}) &:= \{g : \exists f \in V \text{ s.t. } (f,g) \in \mathcal{G}\} \end{aligned} \tag{3.72}$$

the *domain* and *range* of $\mathcal{G}$. We say that $\mathcal{G}$ is *bounded* if there exists a constant K such that

$$\|g\| \leq K\|f\| \qquad \forall (f,g) \in \mathcal{G}. \tag{3.73}$$

We say that $\mathcal{G}$ is a *contraction* if $\|g\| \leq \|f\|$ for all $(f,g) \in \mathcal{G}$. Note that if $\mathcal{G}$ is single-valued and $\mathcal{G}$ is the graph of $(\mathcal{D}(A), A)$, then these definitions coincide with the corresponding definitions for A.

Lemma 3.33 (Bounded operators)**.** *Let V be a Banach space and let $\mathcal{G}$ be a bounded (possibly multivalued) linear operator on $\mathcal{G}$. Then $\mathcal{G}$ is single-valued. Moreover, $\overline{\mathcal{D}(\mathcal{G})} = \mathcal{D}(\overline{\mathcal{G}})$ and $\mathcal{G}$ is closed if and only if $\mathcal{D}(\mathcal{G})$ is closed.*

Proof. Assume that $(f,g), (f,\tilde{g}) \in \mathcal{G}$. Then by linearity $(0, g - \tilde{g}) \in \mathcal{G}$, and by boundedness $\|g - \tilde{g}\| \leq K\|0\| = 0$, hence $g = \tilde{g}$. It follows that $\mathcal{G}$ is the graph of a bounded linear operator $(\mathcal{D}(A), A)$ on V.

One has

$$\begin{aligned} \overline{\mathcal{D}(A)} &= \big\{f \in V : \exists f_n \in \mathcal{D}(A) \text{ s.t. } f_n \to f\big\}, \\ \mathcal{D}(\overline{A}) &= \big\{f \in V : \exists f_n \in \mathcal{D}(A),\ g \in V \text{ s.t. } f_n \to f,\ Af_n \to g\big\}. \end{aligned} \tag{3.74}$$

Therefore, the inclusion $\overline{\mathcal{D}(A)} \supseteq \mathcal{D}(\overline{A})$ is obvious. Conversely, assume that $f_n \in \mathcal{D}(\mathcal{G})$, $f_n \to f$ for some $f \in V$. Then $\|Af_n - Af_m\| \leq K\|f_n - f_m\|$, which shows that the Af_n form a Cauchy sequence. Therefore $Af_n \to g$ for some $g \in V$, which shows that $\overline{\mathcal{D}(A)} \subseteq \mathcal{D}(\overline{A})$.

If $(\mathcal{D}(A), A)$ is closed then by what we have just proved $\overline{\mathcal{D}(A)} = \mathcal{D}(\overline{A}) = \mathcal{D}(A)$, so $\mathcal{D}(A)$ is closed. Conversely, if $\mathcal{D}(A)$ is closed and $f_n \in \mathcal{D}(\mathcal{G})$, $f_n \to f$, $Af_n \to g$, then $f \in \mathcal{D}(A)$ by the fact that $\mathcal{D}(A)$ is closed and $\|A(f_n - f)\| \leq K\|f_n - f\| \to 0$ by the boundedness of $(\mathcal{D}(A), A)$, which shows that $g = \lim_{n\to\infty} Af_n = Af$, and therefore $(f, g) \in \mathcal{G}(A)$. This shows that $(\mathcal{D}(A), A)$ is closed. ∎

Let $\mathcal{G} \subseteq V \times V$ again be a multivalued operator and let λ_1, λ_2 be constants. We define

$$\lambda_1 + \lambda_2\mathcal{G} := \{(f, \lambda_1 f + \lambda_2 g) : (f, g) \in \mathcal{G}\}. \tag{3.75}$$

Note that if $\mathcal{G}$ is the graph of a single-valued operator $(\mathcal{D}(A), A)$, then $\lambda_1 + \lambda_2\mathcal{G}$ is the graph of $(\mathcal{D}(A), \lambda_1 + \lambda_2 A)$. We define

$$\mathcal{G}^{-1} := \{(g, f) : (f, g) \in \mathcal{G}\}. \tag{3.76}$$

If $\mathcal{G}$ is the graph of a single-valued operator $(\mathcal{D}(A), A)$ and A is a bijection from $\mathcal{D}(A)$ to $\mathcal{R}(A)$, then $\mathcal{G}^{-1}$ is the graph of $(\mathcal{R}(A), A^{-1})$. Extending our earlier definition (see (3.47)), we say that $\mathcal{G}$ is *dissipative* if

$$\|f\| \leq \|f - \varepsilon g\| \qquad \forall (f, g) \in \mathcal{G},\ \varepsilon > 0. \tag{3.77}$$

Lemma 3.34 (Closures)**.** *Let V be a Banach space and let $\mathcal{G} \subseteq V \times V$ be a multivalued linear operator on V. Then*

(i) $\overline{\lambda_1 + \lambda_2\mathcal{G}} = \lambda_1 + \lambda_2\overline{\mathcal{G}}$ *for all* $\lambda_1, \lambda_2 \in \mathbb{R}$, $\lambda_2 \neq 0$.
(ii) $\overline{\mathcal{G}^{-1}} = \overline{\mathcal{G}}^{-1}$.
(iii) *If $\mathcal{G}$ is dissipative then $\overline{\mathcal{G}}$ is dissipative.*

Proof. Since $\lambda_2 \neq 0$,

$$\begin{aligned}\overline{\lambda_1 + \lambda_2\mathcal{G}} &= \big\{(f, h) : \exists (f_n, \lambda_1 f + \lambda_2 g_n) \in (\lambda_1 + \lambda_2\mathcal{G}),\\ &\qquad\qquad f_n \to f,\ \lambda_1 f + \lambda_2 g_n \to h\big\}\\ &= \big\{(f, \lambda_1 + \lambda_2 g) : \exists (f_n, g_n) \in \mathcal{G},\\ &\qquad\qquad f_n \to f,\ g_n \to g\big\} = \lambda_1 + \lambda_2\overline{\mathcal{G}}\big\}.\end{aligned} \tag{3.78}$$

The proof of (ii) is similar but easier. To prove (iii), note that if $(f, g) \in \overline{\mathcal{G}}$, then there exist $(f_n, g_n) \in \mathcal{G}$ such that $f_n \to f$ and $g_n \to g$, and therefore, by the dissipativity of $\mathcal{G}$, $\|f\| = \lim_{n\to\infty} \|f_n\| \leq \lim_{n\to\infty} \|f_n - \varepsilon g_n\| = \|f - \varepsilon g\|$. ∎

Lemma 3.35 (Dissipativity and range)**.** *Let $\mathcal{G}$ be dissipative and $\varepsilon > 0$. Then $(1 - \varepsilon\mathcal{G})^{-1}$ is a contraction. Moreover, $\overline{\mathcal{R}(1 - \varepsilon\mathcal{G})} = \mathcal{R}(1 - \varepsilon\overline{\mathcal{G}})$ and $\mathcal{G}$ is closed if and only if $\mathcal{R}(1 - \varepsilon\mathcal{G})$ is closed.*

Proof. If $\mathcal{G}$ is dissipative then $\|f\| \leq \|f - \varepsilon g\|$ for all $(f, g) \in \mathcal{G}$. This means that $\|h\| \leq \|f\|$ for all $(h, f) \in (1 - \varepsilon\mathcal{G})^{-1}$. This shows that $(1 - \varepsilon\mathcal{G})^{-1}$ is a contraction. Therefore, by Lemmas 3.33 and 3.34, $\overline{\mathcal{R}(1 - \varepsilon\mathcal{G})} =$

$\overline{\mathcal{D}((1-\varepsilon\mathcal{G})^{-1})} = \mathcal{D}(\overline{(1-\varepsilon\mathcal{G})^{-1}}) = \mathcal{D}((1-\varepsilon\overline{\mathcal{G}})^{-1}) = \mathcal{R}(1-\varepsilon\overline{\mathcal{G}})$, and $\mathcal{G}$ is closed $\Leftrightarrow$ $(1-\varepsilon\mathcal{G})^{-1}$ is closed $\Leftrightarrow$ $\mathcal{D}((1-\varepsilon\mathcal{G})^{-1})$ is closed $\Leftrightarrow$ $\mathcal{R}(1-\varepsilon\mathcal{G})$ is closed. ∎

Lemma 3.36 (Dissipativity and closability)**.** *Let $(\mathcal{D}(A), A)$ be dissipative and assume that $\mathcal{D}(A)$ is dense in V. Then $(\mathcal{D}(A), A)$ is closable.*

Proof. Let $\mathcal{G}$ be the graph of $(\mathcal{D}(A), A)$. By Lemma 3.34, $\overline{\mathcal{G}}$ is dissipative, while obviously $\mathcal{D}(\overline{\mathcal{G}})$ is dense in V. We need to show that $\overline{\mathcal{G}}$ is single-valued. By linearity, it suffices to show that $(0, g) \in \overline{\mathcal{G}}$ implies $g = 0$. So imagine that $(0, g) \in \overline{\mathcal{G}}$. Since $\mathcal{D}(\overline{\mathcal{G}})$ is dense in V there exist $(g_n, h_n) \in \overline{\mathcal{G}}$ such that $g_n \to g$. Since $\overline{\mathcal{G}}$ is dissipative, $\|0+\varepsilon g_n\| \leq \|(0+\varepsilon g_n) - \varepsilon(g+\varepsilon h_n)\|$ for each $\varepsilon > 0$. It follows that $\|g_n\| \leq \|g_n - g - \varepsilon h_n\|$ for each $\varepsilon > 0$. Letting $\varepsilon \to 0$ and then $n \to \infty$ we find that $\|g\| = \lim_{n\to\infty} \|g_n\| \leq \lim_{n\to\infty} \|g_n - g\| = 0$. ∎

3.8. **Resolvents.**

Definition 3.37 (Resolvents)**.** *By definition, the* resolvent set *of a closed linear operator $(\mathcal{D}(A), A)$ on a Banach space V is the set*

$$\rho(A) := \big\{\lambda \in \mathbb{R} : (\lambda - A) : \mathcal{D}(A) \to V \textit{ is a bijection,} \\ (\lambda - A)^{-1} \textit{ is a bounded operator}\big\}. \tag{3.79}$$

If $\lambda \in \rho(A)$ then the bounded operator $(\lambda - A)^{-1} : V \to \mathcal{D}(A)$ is called the resolvent *of A (at λ).*

Note that $\lambda \in \rho(A)$ implies that λ is not an eigenvalue of A. For imagine that $Ap = \lambda p$ for some $p \in \mathcal{D}(A)$. Then $p = (\lambda - A)^{-1}(\lambda - A)p = (\lambda - A)^{-1}0 = 0$. Note furthermore that the generator $(\mathcal{D}(G), G)$ of a strongly continuous contraction semigroup $(S_t)_{t\geq 0}$ never has eigenvalues $\lambda > 0$. For if $Gf = \lambda f$ with $f \geq 0$ then $u(t) := fe^{\lambda t}$ solves the Cauchy equation $\frac{\partial}{\partial t}u(t) = Gu(t)$ and therefore $S_t f = e^{\lambda t} f$, contradicting contractiveness.

Exercise 3.38. *Show that if $(\mathcal{D}(A), A)$ is not closed then the set $\rho(A)$ in (3.79) is always empty.*

Lemma 3.39 (Resolvent set is open)**.** *Let A be a closed linear operator on a Banach space V. Then the resolvent set $\rho(A)$ is an open subset of $\mathbb{R}$.*

Proof. Assume that $\lambda \in \rho(A)$. Then $(\lambda - A)^{-1}$ is a bounded operator, so there exists a K such that $\|(\lambda - A)^{-1} f\| \leq K\|f\|$ for all $f \in V$. Now let $|\lambda' - \lambda| < K^{-1}$. Then the infinite sum

$$Sf := \sum_{n=0}^{\infty} (\lambda - \lambda')^n (\lambda - A)^{-(n+1)} f \qquad (f \in V) \tag{3.80}$$

defines a bounded operator $S : V \to \mathcal{D}(A)$, and

(3.81)
$$\begin{aligned}(\lambda' - A)Sf &= \big((\lambda - A) - (\lambda - \lambda')\big)\sum_{n=0}^{\infty}(\lambda - \lambda')^n(\lambda - A)^{-n+1}f \\ &= \sum_{n=0}^{\infty}(\lambda - \lambda')^n(\lambda - A)^{-n}f - \sum_{n=0}^{\infty}(\lambda - \lambda')^{n+1}(\lambda - A)^{-n+1}f = f\end{aligned}$$

for each $f \in V$. In the same way we see that $S(\lambda' - A)f = f$ for each $f \in \mathcal{D}(A)$ so $(\lambda' - A) : \mathcal{D}(A) \to V$ is a bijection and its inverse $S = (\lambda' - A)^{-1}$ is a bounded operator. ∎

Exercise 3.40. *Let A be a closed linear operator on a Banach space V and $\lambda, \lambda' \in \rho(A)$, $\lambda \neq \lambda'$. Prove the* resolvent identity

(3.82)
$$(\lambda - A)^{-1}(\lambda' - A)^{-1} = \frac{(\lambda - A)^{-1} - (\lambda' - A)^{-1}}{\lambda' - \lambda} = (\lambda' - A)^{-1}(\lambda - A)^{-1}.$$

According to [EK86, page 11]*: 'Since $(\lambda - A)(\lambda' - A) = (\lambda' - A)(\lambda - A)$ for all $\lambda, \lambda' \in \rho(A)$, we have $(\lambda' - A)^{-1}(\lambda - A)^{-1} = (\lambda - A)^{-1}(\lambda' - A)^{-1}$'. Do you agree with this argument?*

Lemma 3.41 (Resolvent set of dissipative operator)**.** *Let $\rho(A)$ be the resolvent set of a closed dissipative operator $(\mathcal{D}(A), A)$ on a Banach space V and let $\rho^+(A) := \rho(A) \cap (0, \infty)$. Then*

$$\rho^+(A) = \{\lambda > 0 : \mathcal{R}(\lambda - A) = V\} \tag{3.83}$$

and either $\rho^+(A) = \emptyset$ or $\rho^+(A) = (0, \infty)$.

Proof. If A is a dissipative operator and $\lambda > 0$ then by Lemma 3.35 $(\lambda - A)^{-1} = \lambda(1 - \lambda^{-1}A)$ is a bounded operator (and therefore single-valued by Lemma 3.33). This proves (3.83). To see that $\rho^+(A)$ is either $\emptyset$ of $(0, \infty)$, by Lemma 3.39, it suffices to show that $\rho^+(A) \subset (0, \infty)$ is closed. Choose $\lambda_n \in \rho^+(A)$, $\lambda_n \to \lambda \in (0, \infty)$. We need to show that $\mathcal{R}(\lambda - A) = V$. Since A is closed, by Lemma 3.35, it suffices to show that $\mathcal{R}(\lambda - A)$ is dense in V. Choose $g \in V$ and define $g_n := (\lambda - A)(\lambda_n - A)^{-1}g$. Then $g_n \in \mathcal{R}(\lambda - A)$ and, since A is dissipative,

$$\begin{aligned}\|g - g_n\| &= \|\big((\lambda_n - A) - (\lambda - A)\big)(\lambda_n - A)^{-1}g\| \\ &= |\lambda_n - \lambda|\|(\lambda_n - A)^{-1}g\| \leq |\lambda_n - \lambda|\lambda_n^{-1}\|g\| \underset{n\to\infty}{\longrightarrow} 0.\end{aligned} \tag{3.84}$$

∎

3.9. **Hille-Yosida: proofs.** Let $(V, \|\cdot\|)$ be a Banach space. By definition, the *operator norm* of an everywhere defined bounded linear operator $A : V \to V$ is

$$\|A\| := \inf\{K > 0 : \|Af\| \leq K\|f\| \ \forall f \in V\}. \tag{3.85}$$

We say that a collection $\mathcal{A}$ of (everywhere defined) bounded linear operators on V is *uniformly bounded* if $\sup\{\|A\| : A \in \mathcal{A}\} < \infty$. The following fact is well-known, see for example [RS80, Theorem III.9]

Proposition 3.42 (Principle of uniform boundedness)**.** *Let $(V, \|\cdot\|)$ be a Banach space and let $\mathcal{A}$ be a collection of bounded linear operators $A : V \to V$. Assume that $\sup\{\|Af\| : f \in V\} < \infty$ for each $f \in V$. Then $\mathcal{A}$ is uniformly bounded.*

Lemma 3.43 (Order of limits)**.** *Let C, C_n be bounded linear operators on a Banach space V. Assume that $\lim_{n\to\infty} C_n f = Cf$ for all $f \in V$. Then the C_n are uniformly bounded and*

(3.86)
$$\lim_{n\to\infty}\lim_{m\to\infty} C_n f_m = \lim_{m\to\infty}\lim_{n\to\infty} C_n f_m = \lim_{n\to\infty} C_n f_n = Cf \qquad \forall f_n \to f.$$

Proof. By the continuity of the C_n,

$$\lim_{n\to\infty}\lim_{m\to\infty} C_n f_m = \lim_{n\to\infty} C_n f = Cf. \tag{3.87}$$

By the continuity of C,

$$\lim_{m\to\infty}\lim_{n\to\infty} C_n f_m = \lim_{m\to\infty} C f_m = Cf, \tag{3.88}$$

Since $\lim_{n\to\infty} \|C_n f\| = \|Cf\|$ one has $\sup_n \|C_n f\| < \infty$ for each $f \in V$, so by the principle of uniform boundedness the $\{C_n\}$ are uniformly bounded. Set $K := \sup_n \|C_n\|$. Then

$$\begin{aligned} \|C_n f_n - Cf\| &\leq \|C_n f_n - C_n f\| + \|C_n f - Cf\| \\ &\leq K\|f_n - f\| + \|C_n f - Cf\|, \end{aligned} \tag{3.89}$$

which shows that $\lim_{n\to\infty} C_n f_n = Cf$. ∎

Exercise 3.44. *Let V be a Banach space, let C_n be uniformly bounded linear operators on V. Let $f, f_m \in V$, $f_m \to f$. Assume that the limit $\lim_{n\to\infty} C_n f_m =: g_m$ exists for all m. Show that the limit $\lim_{m\to\infty} g_m$ exists and*

$$\lim_{n\to\infty} C_n f = \lim_{m\to\infty} g_m. \tag{3.90}$$

Corollary 3.45. *Let V be a Banach space, let $\mathcal{D} \subset V$ be dense and let C_n be (everywhere defined) uniformly bounded linear operators on V. Assume that $\lim_{n\to\infty} C_n f$ exists for all $f \in \mathcal{D}$. Then there exists a bounded linear operator C on V such that $C_n f \to Cf$ for all $f \in V$.*

Proof. By Exercise 3.44, the set $\{f \in V : \lim_{n\to\infty} C_n f \text{ exists}\}$ is closed, so by our assumptions the limit $\lim_{n\to\infty} C_n f$ exists for all $f \in V$. Define $Cf := \lim_{n\to\infty} C_n f$. It is easy to see that C is linear and bounded. ∎

Proof of Theorem 3.26. By Proposition 3.15 and Lemma 3.21, the conditions (1)–(3) are necessary. Conversely, if (1)–(3) hold, then by Lemma 3.41, $(1-\varepsilon G)^{-1}: V \to \mathcal{D}(G)$ is a bounded operator for each $\varepsilon > 0$. By definition, the *Yosida approximation* of G (at $\varepsilon > 0$) is the everywhere defined bounded (by Lemma 3.35) linear operator

$$G_\varepsilon f := \varepsilon^{-1}\big((1-\varepsilon G)^{-1} - 1\big) f \qquad (f \in V). \tag{3.91}$$

One has

$$\lim_{\varepsilon\to 0} G_\varepsilon f = Gf \qquad (f \in \mathcal{D}(G)). \tag{3.92}$$

To see this, recall that $(1-\varepsilon G)^{-1}(1-\varepsilon G)f = f$ for all $f \in \mathcal{D}(G)$ so that $(1-\varepsilon G)^{-1}f - f = \varepsilon(1-\varepsilon G)^{-1}Gf$ for all $f \in \mathcal{D}(G)$, and therefore by (3.91)

$$G_\varepsilon f = (1-\varepsilon G)^{-1}Gf \qquad (f \in \mathcal{D}(G)). \tag{3.93}$$

In order to prove (3.92), by (3.93) it suffices to show that

$$\lim_{\varepsilon\to 0}(1-\varepsilon G)^{-1}f = f \qquad (f \in V). \tag{3.94}$$

By (3.91), (3.93), and the fact that $(1-\varepsilon G)^{-1}$ is a contraction

$$\|(1-\varepsilon G)^{-1}f - f\| = \varepsilon\|G_\varepsilon f\| = \varepsilon\|(1-\varepsilon G)^{-1}Gf\| \le \varepsilon\|Gf\| \tag{3.95}$$

$(f \in \mathcal{D}(G))$. This proves (3.94) for $f \in \mathcal{D}(G)$. By Corollary 3.45 and the fact that $\mathcal{D}(G)$ is dense, we conclude that (3.94) holds for each $f \in V$.

By Lemma 3.35, $(1-\varepsilon G)^{-1}$ is a contraction, so by (3.91) and Lemma 3.18, G_ε is dissipative. Therefore, by Corollary 3.24, G_ε generates a strongly continuous contraction semigroup $(S^\varepsilon_t)_{t\ge 0} = (e^{G_\varepsilon t})_{t\ge 0}$ on V. We will show that the limit

$$S_t f := \lim_{\varepsilon\to 0} S^\varepsilon_t f \tag{3.96}$$

exists for all $t \ge 0$ and $f \in V$ and defines a strongly continuous contraction semigroup $(S_t)_{t\ge 0}$ with generator G.

It follows from Exercise 3.40 that $G_\varepsilon G_{\varepsilon'} = G_{\varepsilon'} G_\varepsilon$ for all $\varepsilon, \varepsilon' > 0$. Consequently also G_ε and $e^{G_{\varepsilon'}t}$ commute, so

$$\begin{aligned}
&\|e^{G_\varepsilon t}f - e^{G_{\varepsilon'}t}f\| \\
&= \Big\|\int_0^t \tfrac{\partial}{\partial s}\big(e^{G_\varepsilon s}e^{G_{\varepsilon'}(t-s)}\big) f\,\mathrm{d}s\Big\| \\
&\le \int_0^t \big\|(\tfrac{\partial}{\partial s}e^{G_\varepsilon s})e^{G_{\varepsilon'}(t-s)}f + e^{G_\varepsilon s}(\tfrac{\partial}{\partial s}e^{G_{\varepsilon'}(t-s)})f\big\|\mathrm{d}s \\
&= \int_0^t \big\|e^{G_\varepsilon s}e^{G_{\varepsilon'}(t-s)}(G_\varepsilon - G_{\varepsilon'})f\big\|\mathrm{d}s \\
&\le \int_0^t \|(G_\varepsilon - G_{\varepsilon'})f\|\mathrm{d}s = t\|(G_\varepsilon - G_{\varepsilon'})f\|.
\end{aligned} \tag{3.97}$$

Note that we have used commutativity in the last equality. It follows from (3.92) and (3.97) that for each $f \in \mathcal{D}(G)$, $t \ge 0$, and $\varepsilon_n \to 0$, $(e^{G_{\varepsilon_n}t}f)_{n\ge 0}$ is a Cauchy sequence, and therefore the limit in (3.96) exists for all $f \in \mathcal{D}(G)$

and $t \geq 0$. By Corollary 3.45 and the fact that the S^ε_t are contractions, the limit in (3.96) exists for all $f \in V$. With a bit more effort it is possible to see that the limit is locally uniform in t, i.e.,

$$\lim_{\varepsilon\to 0} \sup_{0\leq s\leq T} \|S^\varepsilon_t f - S_t f\| = 0 \qquad \forall T > 0,\ f \in V. \tag{3.98}$$

It remains to show that the operators $(S_t)_{t\geq 0}$ defined in (3.96) form a strongly continuous contraction semigroup with generator G. It is easy to see that they are contractions. For the semigroup property, we note that by Lemma 3.43

$$S_t S_s f = \lim_{\varepsilon\to 0} S^\varepsilon_t S^\varepsilon_s f = \lim_{\varepsilon\to 0} S^\varepsilon_{t+s} f = S_{t+s} f \qquad (f \in V). \tag{3.99}$$

To see that $(S_t)_{t\geq 0}$ is strongly continuous, we note that

$$\lim_{t\to 0} \|S_t f - f\| = \lim_{t\to 0} \lim_{\varepsilon\to 0} \|S^\varepsilon_t f - f\| = \lim_{\varepsilon\to 0} \lim_{t\to 0} \|S^\varepsilon_t f - f\| = 0, \tag{3.100}$$

where the interchanging of limits is allowed by (3.98). In order to prove that G is the generator of $(S_t)_{t\geq 0}$ it suffices to show that $\lim_{t\to 0} t^{-1}(S_t f - f) = Gf$ for all $f \in \mathcal{D}(G)$. For if this is the case, then the generator $(\mathcal{D}(\tilde G), \tilde G)$ of $(S_t)_{t\geq 0}$ is an extension of the operator $(\mathcal{D}(G), G)$. Since both $(\lambda - G) : \mathcal{D}(G) \to \mathcal{R}(\lambda - G) = V$ and $(\lambda - \tilde G) : \mathcal{D}(\tilde G) \to V$ are bijections, this is only possible if $(\mathcal{D}(\tilde G), \tilde G) = (\mathcal{D}(G), G)$.

In order to show that $\lim_{t\to 0} t^{-1}(S_t f - f) = Gf$ for all $f \in \mathcal{D}(G)$ it suffices to show that

$$S_t f - f = \int_0^t S_s G f \,\mathrm{d}s \qquad (f \in \mathcal{D}(G)). \tag{3.101}$$

By Proposition 3.15,

$$S^\varepsilon_t f - f = \int_0^t S^\varepsilon_s G_\varepsilon f \,\mathrm{d}s \qquad (f \in V). \tag{3.102}$$

Using (3.98) and (a simple extension of) Lemma 3.43,

$$\lim_{\varepsilon\to 0} \sup_{0\leq s\leq t} \|S^\varepsilon_s G_\varepsilon f - S_s G f\| = 0 \qquad (f \in \mathcal{D}(G)). \tag{3.103}$$

Inserting this into (3.102) we arrive at (3.101). ∎

Proof of Theorem 3.28. By Theorem 3.26, conditions (1) and (2) are obviously necessary. (Note that in general $\mathcal{D}(\overline{A}) \subseteq \overline{\mathcal{D}(A)}$ so if $\mathcal{D}(A)$ is not dense then $\mathcal{D}(\overline{A})$ is not dense.) By Lemma 3.35, condition (3) is also necessary.

Conversely, if $(\mathcal{D}(A), A)$ satisfies (1)–(3) then by Lemma 3.36, A is closable, while by Lemma 3.35, $\mathcal{R}(\lambda - \overline{A}) = V$, so that by Lemma 3.41, $\overline{A}$ satisfies the conditions of Theorem 3.26. ∎

Proof of Theorem 3.29. By Proposition 3.16 and Theorem 3.28, the conditions (1)–(4) are necessary. We have seen in Lemma 3.19 that the positive maximum principle implies that A is dissipative, so if A satisfies (1)–(4) then by Theorem 3.28 $\overline{A}$ generates a strongly continuous contraction semigroup on $\mathcal{C}(E)$. If $1 \in \mathcal{D}(\overline{A})$ and $\overline{A}1 = 0$ then $u_t := 1$ solves the Cauchy equation $\frac{\partial}{\partial t} u_t = \overline{A} u_t$ so by Proposition 3.15 and Lemma 3.22 $P_t 1 = 1$ for all $t \geq 0$.

To finish the proof, we must show that $P_t f \geq 0$ for all $f \geq 0$. This would be easy using the Cauchy equation if we would know that $\overline{A}$ satisfies the positive maximum principle; unfortunately it is not straightforward to show the latter. Therefore we use a different approach. We know that $\mathcal{R}(1 - \varepsilon A)$ is dense for all $\varepsilon > 0$ and that $(1 - \varepsilon A)^{-1} : \mathcal{R}(1 - \varepsilon A) \to \mathcal{D}(A)$ is a bounded operator. We claim that $(1 - \varepsilon A)^{-1}$ maps nonnegative functions into nonnegative functions. Indeed, if $f \in \mathcal{D}(A)$ does *not* satisfy $f \geq 0$, then f assumes a negative minimum over E in some point x, and therefore by the positive maximum principle applied to $-f$,

$$(1 - \varepsilon A) f(x) \leq f(x) < 0, \tag{3.104}$$

which shows that *not* $(1 - \varepsilon A) f \geq 0$. Thus $(1 - \varepsilon A) f \geq 0$ implies $f \geq 0$, i.e., $(1 - \varepsilon A)^{-1}$ maps nonnegative functions into nonnegative functions. By approximation it follows that also $(1 - \varepsilon \overline{A})^{-1}$ maps nonnegative functions into nonnegative functions.[12] Let $\overline{A}_\varepsilon = \varepsilon^{-1}((1 - \varepsilon \overline{A})^{-1} - 1)$ be the Yosida approximation of $\overline{A}$. Then $f \geq 0$ implies

$$e^{\overline{A}_\varepsilon t} f = e^{-\varepsilon t} e^{\varepsilon^{-1}(1 - \varepsilon \overline{A})^{-1} t} f = e^{-\varepsilon t} \sum_{n=0}^{\infty} \frac{\varepsilon^{-n}}{n!} (1 - \varepsilon \overline{A})^{-n} f \geq 0. \tag{3.105}$$

Letting $\varepsilon \to 0$ we conclude that $P_t f \geq 0$. ∎

[12] Since the set $\mathcal{P} := \{f \in \mathcal{C}(E) : f \geq 0\}$ is the closure of its interior and $\mathcal{R}(1 - \varepsilon A)$ is dense in $\mathcal{C}(E)$, it follows that $\mathcal{R}(1 - \varepsilon A) \cap \mathcal{P}$ is dense in $\mathcal{P}$.

4. Feller processes

4.1. **Markov processes.** Let E be a Polish space. In Proposition 2.12 we have seen that for a given initial law $\mathcal{L}(X_0)$ and transition function (or, equivalently, Markov semigroup) $(P_t)_{t\geq 0}$ on E, there exists a Markov process X, which is unique in finite dimensional distributions. We are not satisfied with this result, however, since do not know in general if X has a version with cadlag sample paths. This motivates us to change our definition of a Markov process. *From now on, we work with the following definition.*

Definition 4.1 (Markov process). *By definition, a* Markov process *with transition function $(P_t)_{t\geq 0}$, is a collection $(\mathbb{P}^x)^{x\in E}$ of probability laws on $(\mathcal{D}_E[0,\infty),\mathcal{B}(\mathcal{D}_E[0,\infty)))$ such that under the law $\mathbb{P}^x$ the stochastic process $X=(X_t)_{t\geq 0}$ given by the coordinate projections*

$$X_t(w) := \xi_t(w) = w(t) \qquad (w\in\mathcal{D}_E[0,\infty),\ t\geq 0), \tag{4.1}$$

satisfies the equivalent conditions (a)–(c) from Proposition 2.11 and one has $\mathbb{P}^x\{X_0=x\}=1$.

Sometimes we denote a Markov process by a pair $(X,(\mathbb{P}^x)^{x\in E})$, since we want to indicate which symbol we use for the coordinate projections. Note that a Markov process $(\mathbb{P}^x)^{x\in E}$ is uniquely determined by its transition function $(P_t)_{t\geq 0}$. We do not know if to each transition function $(P_t)_{t\geq 0}$ there exists a corresponding Markov process $(\mathbb{P}^x)^{x\in E}$. The problem is to show cadlag sample paths. Indeed, by Proposition 2.12, there exists for each $x\in E$ an stochastic process $X^x=(X^x_t)_{t\geq 0}$ such that $X^x_0=x$ and X^x satisfies the equivalent conditions (a)–(c) from Proposition 2.11. If for each $x\in E$ we can find a version of X^x with cadlag sample paths, then the laws $\mathbb{P}^x:=\mathcal{L}(X^x)$, considered as probability measures on $\mathcal{D}_E[0,\infty)$, form a Markov process in the sense of Definition 4.1.

We postpone the proof of the next theorem till later.

Theorem 4.2 (Feller processes). *Let E be compact and metrizable and let $(P_t)_{t\geq 0}$ be a Feller semigroup on $\mathcal{C}(E)$. Then there exists a Markov process $(\mathbb{P}^x)^{x\in E}$ with transition function $(P_t)_{t\geq 0}$.*

Thus, each Feller semigroup $(P_t)_{t\geq 0}$ defines a unique Markov process $(\mathbb{P}^x)^{x\in E}$. We call this the *Feller process* with Feller semigroup $(P_t)_{t\geq 0}$. If $(\mathcal{D}(G),G)$ is the generator of $(P_t)_{t\geq 0}$, then we also say that $(\mathbb{P}^x)^{x\in E}$ is the Feller process with generator G.

We develop some notation and terminology for general Markov processes in Polish spaces.

Lemma 4.3 (Measurability). *Let E be Polish and let $(\mathbb{P}^x)^{x\in E}$ be a Markov process with transition function $(P_t)_{t\geq 0}$. Then $(x,A)\mapsto\mathbb{P}^x(A)$ is a probability kernel from E to $\mathcal{D}_E[0,\infty)$.*

Proof. By definition, $\mathbb{P}^x$ is a probability measure on $\mathcal{D}_E[0,\infty)$ for each fixed $x\in E$. Formula (4.1) shows that for fixed A of the form $A=\{w\in$

$\mathcal{D}_E[0,\infty) : w_{t_1} \in A_1, \ldots, w_{t_n} \in A_n\}$ with $A_1, \ldots, A_n \in \mathcal{B}(E)$, the function $x \mapsto \mathbb{P}^x(A)$ is measurable. Since

$$\mathcal{D} := \{A \in \mathcal{B}(\mathcal{D}_E[0,\infty)) : x \mapsto \mathbb{P}^x(A) \text{ is measurable}\} \tag{4.2}$$

is a Dynkin system and since the coordinate projections generate the Borel-σ-field on $\mathcal{D}_E[0,\infty)$, the same is true for all $A \in \mathcal{B}(\mathcal{D}_E[0,\infty))$. ∎

If $(\mathbb{P}^x)^{x\in E}$ is a Markov process with transition function $(P_t)_{t\geq 0}$ and μ is a probability measure on E, then using Lemma 4.3 we define a probability measure $\mathbb{P}^\mu$ on $(\mathcal{D}_E[0,\infty), \mathcal{B}(\mathcal{D}_E[0,\infty)))$ by

$$\mathbb{P}^\mu(A) := \int_E \mu(\mathrm{d}x)\, \mathbb{P}^x(A) \qquad (A \in \mathcal{B}(\mathcal{D}_E[0,\infty))). \tag{4.3}$$

Under $\mathbb{P}^\mu$, the stochastic process given by the coordinate projections $X = (X_t)_{t\geq 0}$ satisfies the equivalent conditions (a)–(c) from Proposition 2.11 and $\mathbb{P}^x\{X_0 \in \cdot\,\} = \mu$. We call $(X, \mathbb{P}^\mu)$ the Markov process with transition function $(P_t)_{t\geq 0}$ started in the initial law μ. We let $\mathbb{E}^x$, $\mathbb{E}^\mu$ denote expectation with respect $\mathbb{P}^x, \mathbb{P}^\mu$, respectively.

Recall that two stochastic processes with the same finite dimensional distributions are called *versions* of each other. Thus, if $(X, (\mathbb{P}^x)^{x\in E})$ is a Markov process with transition function $(P_t)_{t\geq 0}$, then a stochastic process X', defined on any probability space, which has the same finite dimensional distributions as X under the law $\mathbb{P}^\mu$, is called a *version* of the Markov process with semigroup $(P_t)_{t\geq 0}$ and initial law μ. This is equivalent to the statement that X' satisfies the equivalent conditions (a)–(c) from Proposition 2.11 and $\mathcal{L}(X'_0) = \mu$. If X' has moreover cadlag sample paths then this is equivalent to $\mathcal{L}(X') = \mathbb{P}^\mu$, where we view X' as a random variable with values in $\mathcal{D}_E[0,\infty)$. We are usually only interested in versions with cadlag sample paths.

4.2. **Jump processes.** Jump processes are the simplest Markov processes. We have already met them in Exercise 3.66.

Proposition 4.4 (Jump processes)**.** *Let E be Polish, let K be a probability kernel on E, and let $r \geq 0$ be a constant. Define $G : B(E) \to B(E)$ by*

$$Gf := r(Kf - f) \qquad (f \in B(E)), \tag{4.4}$$

and put

$$P_t f := e^{Gt} f := \sum_{n=0}^{\infty} \frac{1}{n!}(Gt)^n f \qquad (f \in B(E),\ t \geq 0). \tag{4.5}$$

Then $(P_t)_{t\geq 0}$ is a Markov semigroup and there exists a Markov process $(\mathbb{P}^x)^{x\in E}$ corresponding to $(P_t)_{t\geq 0}$. If E is compact and K is a continuous probability kernel then $(\mathbb{P}^x)^{x\in E}$ is a Feller process.

Note that the infinite sum in (4.5) converges uniformly since $\|r(Kf - f)\| \leq 2r\|f\|$ for each $f \in B(E)$. Before we prove Proposition 4.4 we first look at a special case.

Example: (Poisson process with rate r). Let $E := \mathbb{N}$, $K(x,\{y\}) := 1_{\{y=x+1\}}$, and $r > 0$. Hence

$$Gf(x) = r\big(f(x+1) - f(x)\big) \qquad (f \in B(\mathbb{N})). \tag{4.6}$$

Then the Markov semigroup in (4.5) is given by

$$\begin{aligned} P_t f(x) &= e^{Gt} = e^{rt(K-1)} = e^{-rt}e^{rtK} \\ &= e^{-rt}\sum_{n=0}^{\infty}\frac{(rt)^n}{n!}K^n f(x) = e^{-rt}\sum_{n=0}^{\infty}\frac{(rt)^n}{n!}f(x+n), \end{aligned}$$

hence

$$P_t(x,\{x+n\}) = 1_{\{n\geq 0\}}e^{-rt}\frac{(rt)^n}{n!}. \tag{4.7}$$

We call the associated Markov process $(N,(\mathbb{P}^x)^{x\in\mathbb{N}})$ the *Poisson process* with intensity r. By condition (a) from Proposition 2.11,

$$\mathbb{P}^x(\{N_t - N_s = n\}|\mathcal{F}^N_s) = 1_{\{n\geq 0\}}e^{-r(t-s)}\frac{(r(t-s))^n}{n!}. \tag{4.8}$$

This says that $N_t - N_s$ is Poisson distributed with mean $r(t-s)$. Since the right-hand side of (4.8) does not depend on N_s, the random variable $N_t - N_s$ is independent of $(N_u)_{u\leq s}$. It follows that if $(N_t)_{t\geq 0}$ is a version of the Poisson process started in any initial law, then for any $0 \leq t_1 \leq \cdots \leq t_n$, the random variables

$$N_{t_1} - N_0, \ldots, N_{t_n} - N_{t_{n-1}}$$

are independent and Poisson distributed with means $r(t_1 - 0), \ldots, r(t_n - t_{n-1})$. Recall that if P, Q are Poisson distributed random variables with means p and q, then $P + Q$ is Poisson distributed with mean $p + q$.

Poisson processes describe the statistics of rare events. For each $n \geq 1$, let $(X^{(n)}_i)_{i\in\mathbb{N}}$ be a Markov chain in $\mathbb{N}$ with $X^{(n)}_0 = 0$ and transition probabilities

$$\mathbf{P}(\{X^{(n)}_{i+1} = y\}|X^{(n)}_0, \ldots, X^{(n)}_i) = p^{(n)}(X^{(n)}_i, y),$$

where

$$p^{(n)}(x,y) := \begin{cases} 1 - \frac{1}{n} & \text{if } y = x, \\ \frac{1}{n} & \text{if } y = x+1, \\ 0 & \text{otherwise.} \end{cases}$$

Fix $r > 0$ and define processes $(N^{(n)}_t)_{t\geq 0}$ by

$$N^{(n)}_t := X^{(n)}_{\lfloor nrt\rfloor} \qquad (t \geq 0).$$

Then

$$\mathbf{P}\{N^{(n)}_t = k\} = \binom{m}{k}\frac{1}{n}^k\left(1 - \frac{1}{n}\right)^{m-k}, \quad \text{where } m := \lfloor nrt\rfloor.$$

It follows that $N_t^{(n)}$ converges as $n \to \infty$ to a Poisson distributed random variable with mean rt. With a bit more work it is easy to see that the stochastic process $N^{(n)}$ converges in finite dimensional distributions to a Poisson process with intensity r, started in $N_0 = 0$.

The next exercise show how to construct versions of the Poisson process with cadlag sample paths. ∎

Exercise 4.5. *Let $(\sigma_k)_{k\geq 1}$ be independent exponentially distributed random variables with mean r^{-1} and set $\tau_n := \sum_{k=1}^n \sigma_k$. Show that*

$$N_t := \max\{n \geq 0 : \tau_n \leq t\} \qquad (t \geq 0) \tag{4.9}$$

defines a version $N = (N_t)_{t\geq 0}$ of the Poisson process with rate r started in $N_0 = 0$.

Proof of Proposition 4.4. The case $r = 0$ is trivial so assume $r > 0$. For each $x \in E$, let $(Y_n^x)_{n\geq 0}$ be a Markov chain started in $Y_0^x = x$ with transition kernel K i.e.,

$$\mathbf{P}(\{Y_n^x \in A\}|Y_0^x, \ldots, Y_{n-1}^x) = K(Y_{n-1}^x, A) \quad \text{a.s.} \qquad (A \in \mathcal{B}(E)). \tag{4.10}$$

Let $(\sigma_k)_{k\geq 1}$ be independent exponentially distributed random variables with mean r^{-1}, independent of Y^x, and set $\tau_n := \sum_{k=1}^n \sigma_k$. Define a process $X^x = (X_t^x)_{t\geq 0}$ by

$$X_t^x := Y_n^x \qquad \text{if} \quad \tau_n \leq t < \tau_{n+1}. \tag{4.11}$$

We claim that X^x satisfies the equivalent conditions (a)–(c) from Proposition 2.11. Since X^x obviously has cadlag sample paths this then implies that $\mathbb{P}^x := \mathcal{L}(X^x)$ defines a Markov process with semigroup $(P_t)_{t\geq 0}$. If E is compact and K is a continuous probability kernel then we have already seen in Exercise 3.66 that $(P_t)_{t\geq 0}$ is a Feller semigroup.

To see that X^x defined in (4.11) satisfies condition (c) from Proposition 2.11, let $N = (N_t)_{t\geq 0}$ be a Poisson process with intensity r, started in $N_0 = 0$, independent of Y^x. Then (4.11) says that

$$X_t^x := Y_{N_t}^x \qquad (t \geq 0), \tag{4.12}$$

i.e., X^x jumps according to the kernel K at random times that are given by a Poisson process with intensity r. It follows that for any $f \in B(E)$,

$$\begin{aligned}\mathbf{E}[f(Y_{N_t}^x)] &= \sum_{k=0}^{\infty} \mathbf{P}\{N_t = k\}\mathbf{E}[f(Y_k^x)] = \sum_{k=0}^{\infty} e^{-rt}\frac{(rt)^k}{k!}K^k f(x) \\ &= e^{-rt}e^{rtK}f(x) = e^{tr(K-1)}f(x) = e^{tG}f(x).\end{aligned}$$

The proof that X^x satisfies condition (c) from Proposition 2.11 goes basically the same. Let $0 \leq t_1 \leq \cdots \leq t_n$ and $f_1, \ldots, f_n \in B(E)$. Then, since a

Poisson process has independent increments,

$$\begin{aligned}
&\mathbf{E}\big[f_1(Y^x_{N_{t_1}})\cdots f_n(Y^x_{N_{t_n}})\big] \\
&= \sum_{k_1=1}^{\infty}\cdots\sum_{k_n=1}^{\infty}\mathbf{P}\{N_{t_1}=k_1\}\cdots\mathbf{P}\{N_{t_n}-N_{t_{n-1}}=k_n\} \\
&\qquad\qquad\qquad\cdot\mathbf{E}\big[f_1(Y^x_{k_1})\cdots f_n(Y^x_{k_1+\cdots+k_n})\big] \\
&= \sum_{k_1=1}^{\infty}\cdots\sum_{k_n=1}^{\infty}e^{-rt_1}\frac{(rt_1)^{k_1}}{k_1!}\cdots e^{-r(t_n-t_{n-1})}\frac{(r(t_n-t_{n-1}))^{k_n}}{k_n!} \\
&\qquad\qquad\qquad\cdot K^{k_1}f_1K^{k_2}f_2\cdots K^{k_n}f_n(x) \\
&= e^{t_tG}f_1e^{(t_2-t_1)G}f_2\cdots e^{(t_n-t_{n-1})G}f_n(x).
\end{aligned}$$

Inserting $f_i = 1_{A_i}$ we see that condition (c) from Proposition 2.11 is satisfied. ∎

Remark. Jump processes can be approximated with Markov chains. Let E be Polish, K a probability kernel on E, $x \in E$, and $r > 0$. For each $n \geq 1$, let $(Y_i^{(n)})_{i\geq 0}$ be a Markov chain with $Y_0^{(n)} = x$ and transition probabilities

$$\mathbf{P}(\{Y_{i+1}^{(n)} \in \cdot\}|Y_0^{(n)},\ldots,Y_i^{(n)}) = K^{(n)}(Y_i^{(n)},\,\cdot\,),$$

where

$$K^{(n)}(y,\mathrm{d}z) = \tfrac{1}{n}K(y,\mathrm{d}z) + (1-\tfrac{1}{n})\delta_y(\mathrm{d}z).$$

Then the processes $(X_t^{(n)})_{t\geq 0}$ given by

$$X_t^{(n)} := Y_{\lfloor nrt\rfloor}^{(n)}$$

converge as $n \to \infty$ in finite dimensional distributions to the jump process X with jump kernel K, jump rate r, and initial condition $X_0 = x$. ∎

A well-known example of a jump process is continuous-time random walk.

Example: (Random walk). Let $d \geq 1$ and let $p : \mathbb{Z}^d \to \mathbb{R}$ be a probability distribution on $\mathbb{Z}^d$, i.e., $p(x) \geq 0$ for all $x \in \mathbb{Z}$ and $\sum_x p(x) = 1$. Let $E := \mathbb{Z}^d$ (equipped with the discrete topology), $K(x,\{y\}) := p(y-x)$, $r > 0$, and define G as in (4.4). The jump process $(X,(\mathbb{P}^x)^{x\in\mathbb{Z}^d})$ with generator G is called the (continuous-time) random walk that jumps from x to y with rate $rp(y-x)$.

Let $(Z_k)_{k\geq 1}$ be independent random variables with distribution $\mathbf{P}\{Z_k = x\} = p(x)$ $(x \in \mathbb{Z}^d)$ and let $(\sigma_k)_{k\geq 1}$ be independent exponentially distributed random variables with mean r^{-1}, independent of $(Z_k)_{k\geq 1}$. Set $Y_n^x := x + \sum_{k=1}^n Z_k$, $\tau_n := \sum_{k=1}^n \sigma_k$, and put

$$X_t^x := Y_n^x \qquad \text{if} \quad \tau_n \leq t < \tau_{n+1} \tag{4.13}$$

Then $X^x = (X_t^x)_{t\geq 0}$ is a version of the random walk that jumps from x to y with rate $rp(y-x)$, started in $X_0^x = x$. ∎

Often, one wants to consider jump processes in which the jump rate r is a function of the position of the process. As long as r is a bounded function, such jump processes exist by a trivial extension of Proposition 4.4.

Proposition 4.6 (Jump processes with non-constant rate)**.** *Let E be Polish, let K be a probability kernel on E, and let $r \in B(E)$ be nonnegative. Define $G : B(E) \to B(E)$ by*

$$Gf := r(Kf - f) \qquad (f \in B(E)). \tag{4.14}$$

Then $P_t f := e^{Gt} f$ $(f \in B(E),\ t \geq 0)$ defines a Markov semigroup and there exists a Markov process $(\mathbb{P}^x)^{x \in E}$ corresponding to $(P_t)_{t \geq 0}$. If E is compact and K and r are continuous then $(\mathbb{P}^x)^{x \in E}$ is a Feller process.

Proof. Set $R := \sup_{x \in E} r(x)$ and define

$$K'(x, \mathrm{d}y) := \frac{r(x)}{R} K(x, \mathrm{d}y) + \left(1 - \frac{r(x)}{R}\right) \delta_x(\mathrm{d}y). \tag{4.15}$$

Then $Gf = R(K'f - f)$ so we are back at the situation in Proposition 4.4. ∎

Example: (Moran model). Fix $n \geq 1$, put $E := \{0, 1, \ldots, n\}$, and

$$G_X f(x) := \tfrac{1}{2} x(n - x)\big(f(x + 1) + f(x - 1) - 2f(x)\big), \tag{4.16}$$

which corresponds to setting $r(x) = x(n-x)$ and $K(x, \{y\}) := \frac{1}{2} 1_{\{y=x+1\}} + \frac{1}{2} 1_{\{y=x-1\}}$ for $x = \{1, \ldots, n-1\}$. Observe that since $r(0) = r(n) = 0$ it is irrelevant how we define $K(0,\,\cdot\,)$ and $K(n,\,\cdot\,)$. The jump process $(X, (\mathbb{P}^x)^{x \in E})$ with generator G_X is called the Moran model with population size n.

The Moran model arises in the following way. Consider n organisms that are divided into two types, denoted by $0, 1$. (For example, 0 might represent a white flower and 1 a red one.) Let $S_n := \{0,1\}^n = \{y = (y(1), \ldots, y(n)) : y(i) \in \{0, 1\}\ \forall i\}$ be the set of all different ways in which we can assign types to these n organisms. Put $v_{ij}(y)(i) := y(j)$ and $v_{ij}(y)(k) := y(k)$ if $k \neq i$. Then $v_{ij}(y)$ is the configuration in which the i-th organism has adopted the type of the j-th organism. Let $(Y, \mathbb{P}^{y \in S_n})$ be the Markov process with generator

$$G_Y f(y) := \tfrac{1}{2} \sum_{ij} \big(f(v_{ij}(y)) - f(y)\big). \tag{4.17}$$

This means that each unordered pair $\{i, j\}$ of organisms is selected with rate 1, and then one of these organisms, chosen with equal probabilities, takes over the type of the other one. (Note that there is no harm in including $i = j$ in the sum in (4.17) since $v_{ii}(y) = y$.) Now if Y^y is a version of the Markov process with generator G_Y started in $Y^y_0 = y$, then

$$X^x_t := \sum_{i=1}^{n} Y^y_t(i) \qquad (t \geq 0) \tag{4.18}$$

is a version of the Moran model started in $x := \sum_{i=1}^n y(i)$. To see this, at least intuitively, note that if $x = \sum_{i=1}^n y(i)$ then $x(n-x)$ is the number of unordered pairs $\{i,j\}$ of organisms such that i and j have different types, and therefore $\frac{1}{2}x(n-x)$ is the total rate of 1's changing to 0's, which equals the total rate of 0's changing to 1's. ∎

4.3. **Feller processes with compact state space.** We will now take a look at some examples of Markov processes that are not jump processes. All processes that we will look at are processes on compact subsets of $\mathbb{R}^d$ with continuous sample paths (although we will not prove the latter here). One should keep in mind that there are many more possibilities for a Markov process not to be a jump process. For example, there are processes that have a combination of continuous and jump dynamics or that make infinitely many (small) jumps in each open time interval.

Let $d \geq 1$, let $D \subset \mathbb{R}^d$ be a bounded open set and let $\overline{D}$ denote its closure. Let $f|_{\overline{D}}$ denote the restriction of a function $f : \mathbb{R}^d \to \mathbb{R}$ to $\overline{D}$. By definition:

$$\mathcal{C}^2(\overline{D}) := \{f|_{\overline{D}} : f : \mathbb{R}^d \to \mathbb{R} \text{ twice continuously differentiable}\}. \tag{4.19}$$

Let M_+^d denote the space of real $d \times d$ matrices m that are *symmetric*, i.e., $m_{ij} = m_{ji}$, and *nonnegative definite*, i.e.,

$$\sum_{ij} v_i m_{ij} v_j \geq 0 \qquad \forall v \in \mathbb{R}^d. \tag{4.20}$$

Let $a : \overline{D} \to M_+^d$ and $b : \overline{D} \to \mathbb{R}^d$ be continuous functions and let $(\mathcal{D}(A), A)$ be the linear operator on $\mathcal{C}(\overline{D})$ defined by

$$Af(x) := \tfrac{1}{2}\sum_{ij} a_{ij}(x)\tfrac{\partial^2}{\partial x_i \partial x_j} f(x) + \sum_i b_i(x)\tfrac{\partial}{\partial x_i} f(x) \qquad (x \in \overline{D}). \tag{4.21}$$

For $x \in D$ these derivatives are defined in the obvious way, that is, $\frac{\partial}{\partial x_i} f(x) = \lim_{\varepsilon \to 0} \varepsilon^{-1}\big(f(x + \varepsilon\delta_i) - f(x)\big)$. For x in the boundary $\partial D := \overline{D} \backslash D$ we have to be a bit careful since it may happen that $x + \varepsilon\delta_i \notin \overline{D}$ for all $\varepsilon \neq 0$. By definition, each $f \in \mathcal{C}^2(\overline{D})$ can be extended to a continuously differentiable function $\overline{f}$ on all of $\mathbb{R}^d$. Therefore, we define

$$\tfrac{\partial}{\partial x_i} f := (\tfrac{\partial}{\partial x_i}\overline{f})\big|_{\overline{D}}. \tag{4.22}$$

To see that this definition does not depend on the choice of the extension $\overline{f}$, note that if $\hat{f}$ is another extension, then $\frac{\partial}{\partial x_i}\overline{f} = \frac{\partial}{\partial x_i}\hat{f}$ on D. By continuity, $\frac{\partial}{\partial x_i}\overline{f} = \frac{\partial}{\partial x_i}\hat{f}$ on $\overline{D}$.[13]

[13] Alternatively, we might have defined $\mathcal{C}^2(\overline{D})$ as the space of all functions $f : \overline{D} \to \mathbb{R}$ whose partial derivatives up to second order exist on D and can be extended to continuous functions on $\overline{D}$. For 'nice' (for example convex) domains D this definition coincides with the definition in (4.19). This is a consequence of Whitney's extension theorem, see [EK86, Appendix 6].

We ask ourselves when the closure of A generates a Feller process in $\overline{D}$, i.e., satisfies the conditions (1)–(3) from Theorem 3.28. By the Stone-Weierstrass theorem (Theorem 3.30), $\mathcal{C}^2(\overline{D})$ is dense in $\mathcal{C}(\overline{D})$, so condition (1) is always satisfied.

If $f \in \mathcal{C}^2(\overline{D})$ assumes its maximum in a point $x \in D$, then $Af(x) \leq 0$. This is a consequence of the fact that $a(x)$ is nonnegative definite. In fact, since $a(x)$ is symmetric, it can be diagonalized. Therefore, for each x there exist orthonormal vectors $e^1(x), \ldots, e^d(x) \in \mathbb{R}^d$ and constants $a^1(x), \ldots, a^k(x)$ such that

$$\sum_{ij} a_{ij}(x)\tfrac{\partial^2}{\partial x_i \partial x_j} f(x) = \sum_k a^k(x)\tfrac{\partial^2}{\partial \varepsilon^2} f(x + \varepsilon e^k(x))\big|_{\varepsilon=0}. \tag{4.23}$$

Since a is nonnegative definite the constants $a^k(x)$ are all nonnegative, and if f assumes its maximum in x then $\frac{\partial^2}{\partial \varepsilon^2} f(x + \varepsilon e^k(x))|_{\varepsilon=0} \leq 0$ for each k.

Exercise 4.7. *Let $D := \{x \in \mathbb{R}^2 : |x| < 1\}$ be the open unit ball in $\mathbb{R}^2$ and put*

$$\begin{pmatrix} a_{11}(x) & a_{12}(x) \\ a_{21}(x) & a_{22}(x) \end{pmatrix} := \begin{pmatrix} x_2^2 & -x_1 x_2 \\ -x_1 x_2 & x_1^2 \end{pmatrix} \tag{4.24}$$

and

$$(b_1(x), b_2(x)) := c\,(x_1, x_2). \tag{4.25}$$

For which values of c does the operator A in (4.21) satisfy the positive maximum principle?

The preceding exercise shows that it is not always easy to see when A satisfies the positive maximum principle also for $x \in \partial D$. If this is the case, however, and by some means one can also check Condition (4) from Theorem 3.29, then $\overline{A}$ generates a Feller process $(X, (\mathbb{P}^x)^{x \in \overline{D}})$ in $\overline{D}$. We will later see that under $\mathbb{P}^x$, X has a.s. continuous sample paths. We call $(X, (\mathbb{P}^x)^{x \in \overline{D}})$ the *diffusion* with *drift* b and *local diffusion rate* (or *diffusion matrix*) a. The next lemma explains the meaning of the functions a and b.

Lemma 4.8 (Drift and diffusion rate)**.** *Assume that the closure of the operator A in (4.21) generates a Feller semigroup $(P_t)_{t \geq 0}$. Then, as $t \to 0$,*

$$\begin{aligned} &\text{(i)} \quad \int_{\overline{D}} P_t(x, \mathrm{d}y)(y_i - x_i) &&= b_i(x)t + o(t), \\ &\text{(ii)} \quad \int_{\overline{D}} P_t(x, \mathrm{d}y)(y_i - x_i)(y_j - x_j) &&= a_{ij}(x)t + o(t), \end{aligned} \tag{4.26}$$

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

Proof. For any $f \in \mathcal{C}^2(\overline{D})$ we have by (3.32)

$$\int_{\overline{D}} P_t(x, \mathrm{d}y) f(y) = P_t f(x) = f(x) + tAf(x) + o(t) \quad \text{as } t \to 0. \tag{4.27}$$

Fix $x \in \overline{D}$ and set $f_i(y) := (y_i - x_i)$. Then $f(x) = 0$ and $Af_i(x) = b_i(x)$, and therefore (4.27) yields (4.26) (i). Likewise, inserting $f_{ij}(y) := (y_i - x_i)(y_j - x_j)$ into (4.27) yields (4.26) (ii). ∎

By condition (a) from Proposition 2.11, Lemma 4.8 says that if X is a version of the Markov process with generator $\overline{A}$, started in any initial law, then

(4.28)
$$\begin{aligned} &\text{(i)} && \mathbf{E}[(X_{t+\varepsilon}(i) - X_t(i))|\mathcal{F}_t^X] &&= b_i(X_t)\varepsilon + o(\varepsilon),\\ &\text{(ii)} && \mathbf{E}[(X_{t+\varepsilon}(i) - X_t(i))(X_{t+\varepsilon}(j) - X_t(j))|\mathcal{F}_t^X] &&= a_{ij}(X_t)\varepsilon + o(\varepsilon). \end{aligned}$$

Therefore, the functions b and a describe the mean and the covariance matrix of small increments of the diffusion process X.[14] If $a \equiv 0$ then

$$\mathbb{P}^x\{X_t = x_t \ \forall t \geq 0\} = 1, \tag{4.29}$$

where $t \mapsto x_t$ solves the differential equation

$$\tfrac{\partial}{\partial t} x_t = b(x_t) \quad \text{with} \quad x_0 = x. \tag{4.30}$$

In general, diffusions can be obtained as solutions to *stochastic differential equations* of the form

$$\mathrm{d}X_t = \sigma(X_t)\mathrm{d}B_t + b(X_t)\mathrm{d}t \tag{4.31}$$

where $\sigma(x)$ is a matrix such that $\sum_k \sigma_{ik}(x)\sigma_{jk}(x) = a_{ij}(x)$ and B is d-dimensional Brownian motion, but this falls outside the scope of this section.

Example: (Wright-Fisher diffusion). By Exercise 3.32, the operator A_{WF} from Exercise 3.20 generates a diffusion process $(Y, (\mathbb{P}^y)^{y\in[0,1]})$ in $[0,1]$. This diffusion is known as the *Wright-Fisher diffusion.*

For each $n \geq 1$, let X^n be a Moran model in $\{0, \ldots, n\}$ (see (4.16)) with some initial law $\mathcal{L}(X_0^n)$. Define

$$Y_t^n := \tfrac{1}{n} X_t^n, \tag{4.32}$$

and assume that $\mathcal{L}(Y_0^n) \Rightarrow \mu$ for some probability measure μ on $[0,1]$. We claim that $\mathcal{L}(Y^n) \Rightarrow \mathcal{L}(Y)$ where Y is a version of the Wright-Fisher diffusion with initial law $\mathcal{L}(Y_0) = \mu$. The proof of this fact relies on some deep results that we have not seen yet, so we will only give a heuristic argument. Let $(P_t^n)_{t\geq 0}$ be the transition function of Y^n. Then by (4.16),

$$P_t^n(y, \cdot\,) = \delta_y + tn^2 y(1-y)\big(\tfrac{1}{2}\delta_{y+\frac{1}{n}} + \tfrac{1}{2}\delta_{y-\frac{1}{n}} - \delta_y\big) + o(t), \tag{4.33}$$

and therefore

$$\begin{aligned} &\text{(i)} && \int_{\overline{D}} P_t^n(y, \mathrm{d}z)(z-y) = o(t),\\ &\text{(ii)} && \int_{\overline{D}} P_t^n(y, \mathrm{d}z)(z-y)^2 = ty(1-y) + o(t). \end{aligned} \tag{4.34}$$

[14] Indeed, since the mean of $P_t(x, \cdot\,)$ is x plus a term of order t, the covariance matrix of $P_t(x, \cdot\,)$ is equal to $\int_{\overline{D}} P_t(x, \mathrm{d}y)(y_i - x_i)(y_j - x_j)$ up to an error term of order t^2.

Thus, at least the first and second moments of small increments of the process Y^n converge to those of Y. ∎

The next example shows that the domain of an operator is not only of technical interest, but can significantly contribute to the behavior of the corresponding Markov process.

Example: (Brownian motion with absorption and reflection). Define linear operators $(\mathcal{D}(A_{\rm ab}), A_{\rm ab})$ and $(\mathcal{D}(A_{\rm re}), A_{\rm re})$ on $\mathcal{C}[0,1]$ by

$$\begin{aligned} \mathcal{D}(A_{\rm ab}) &:= \{f \in \mathcal{C}^2[0,1] : f''(0) = 0 = f''(1)\}, \\ A_{\rm ab} f(x) &:= \tfrac{1}{2} f''(x) \qquad (x \in [0,1]), \end{aligned} \tag{4.35}$$

and

$$\begin{aligned} \mathcal{D}(A_{\rm re}) &:= \{f \in \mathcal{C}^2[0,1] : f'(0) = 0 = f'(1)\}, \\ A_{\rm re} f(x) &:= \tfrac{1}{2} f''(x) \qquad (x \in [0,1]). \end{aligned} \tag{4.36}$$

Then the closures of $A_{\rm ab}$ and $A_{\rm ab}$ generate a Feller processes in $[0,1]$. The operator $\overline{A}_{\rm ab}$ generates Brownian motion absorbed at the boundary and $\overline{A}_{\rm re}$ generates Brownian motion reflected at the boundary.

To see that $A_{\rm ab}$ and $A_{\rm re}$ satisfy the positive maximum principle, note that if $f \in \mathcal{D}(A_{\rm ab})$ or $f \in \mathcal{D}(A_{\rm re})$ assumes its maximum in a point $x \in (0,1)$ then $\frac{1}{2}f''(0) \leq 0$. If $f \in \mathcal{D}(A_{\rm ab})$ assumes its maximum in a point $x \in \{0,1\}$ then $A_{\rm ab}f(x) = \frac{1}{2}f''(x) = 0$ by the definition of $\mathcal{D}(A_{\rm ab})$! Similarly, if $f \in \mathcal{D}(A_{\rm re})$ assumes its maximum in a point $x \in \{0,1\}$ then $A_{\rm re}f(x) = \frac{1}{2}f''(x) \leq 0$ because of the fact that $f'(x) = 0$ by the definition of $\mathcal{D}(A_{\rm re})$.

The fact that $A_{\rm ab}$ and $A_{\rm re}$ satisfy Condition (4) from Theorem 3.29 follows from the theory of partial differential equations, see for example [Fri64]. ∎

4.4. **Feller processes with locally compact state space.** So far, we have only been able to treat Feller processes with compact state spaces. We will now show how to deal with processes with locally compact state spaces. We start with an example.

Example: (Brownian motion). Fix $d \geq 1$ and define

$$P_t(x, \mathrm{d}y) := \tfrac{1}{(2\pi t)^{d/2}}\, e^{-\frac{1}{2t}|y-x|^2} \mathrm{d}y. \tag{4.37}$$

Then $(P_t)_{t\geq 0}$ is a transition function on $\mathbb{R}^d$ and there exists a Markov process $(B, (\mathbb{P}^x)^{x\in\mathbb{R}^d})$ with continuous sample paths associated with $(P_t)_{t\geq 0}$.

Let $\overline{\mathbb{R}^d} := \mathbb{R}^d \cup \{\infty\}$ be the one-point compactification of $\mathbb{R}^d$ (compare with (1.66)) and define a Markov process $(\overline{\mathbb{P}}^x)^{x\in\overline{\mathbb{R}^d}}$ by

$$\overline{\mathbb{P}}^x := \begin{cases} \mathbb{P}^x & \text{if } x \in \mathbb{R}^d, \\ \delta_\infty & \text{if } x = \infty, \end{cases} \tag{4.38}$$

where δ_∞ denotes the delta-measure on the constant function $w(t) := \infty$ for all $t \geq 0$. Note that $\overline{\mathbb{P}}^x$ is a measure on $\mathcal{D}_{\overline{\mathbb{R}^d}}[0,\infty)$ while $\mathbb{P}^x$ is a measure on

$\mathcal{D}_{\mathbb{R}^d}[0,\infty)$, so when we say that $\overline{\mathbb{P}}^x = \mathbb{P}^x$ for $x \in \mathbb{R}^d$ we mean that $\overline{\mathbb{P}}^x$ is the image of $\mathbb{P}^x$ under the embedding map $\mathcal{D}_{\mathbb{R}^d}[0,\infty) \subset \mathcal{D}_{\overline{\mathbb{R}^d}}[0,\infty)$.

We claim that $(\overline{\mathbb{P}}^x)^{x\in\overline{\mathbb{R}^d}}$ is a Feller process with compact state space $\overline{\mathbb{R}^d}$. It is not hard to see that this is a Markov process with transition function

$$\overline{P}_t(x,\,\cdot\,) := \begin{cases} P_t(x,\,\cdot\,) & \text{if } x \in \mathbb{R}^d, \\ \delta_\infty & \text{if } x = \infty. \end{cases} \tag{4.39}$$

We must show that this transition function is continuous. This means that we must show that $(t,x) \mapsto \overline{P}_t f(x)$ is continuous for each $f \in \mathcal{C}(\overline{\mathbb{R}^d})$. Since $\overline{P}_t 1(x) = 1$, by subtracting a constant it suffices to show that $(t,x) \mapsto \overline{P}_t f(x)$ is continuous for each $f \in \mathcal{C}_0(\overline{\mathbb{R}^d}) := \{f \in \mathcal{C}(\overline{\mathbb{R}^d}) : f(\infty) := 0\}$. Assume that $(t_n,x_n) \to (t,x) \in [0,\infty) \times \overline{\mathbb{R}^d}$. Without loss of generality we may assume that $x_n \neq \infty$ for all n. We distinguish two cases. 1. If $x \neq \infty$, then by uniform convergence

$$\begin{aligned} P_{t_n} f(x_n) &= \int_{\mathbb{R}^d} \tfrac{1}{(2\pi t_n)^{d/2}}\, e^{-\frac{1}{2t_n}|y-x_n|^2} f(y)\,\mathrm{d}y \\ &\underset{n\to\infty}{\longrightarrow} \int_{\mathbb{R}^d} \tfrac{1}{(2\pi t)^{d/2}}\, e^{-\frac{1}{2t}|y-x|^2} f(y)\,\mathrm{d}y = P_t f(x). \end{aligned} \tag{4.40}$$

2. If $x = \infty$, then for each compact set $C \subset \mathbb{R}^d$

$$|P_{t_n} f(x_n)| \leq \|f\| \tfrac{1}{(2\pi t_n)^{d/2}} \sup_{y\in C} e^{-\frac{1}{2t_n}|y-x_n|^2} + \sup_{x\in\overline{\mathbb{R}^d}\setminus C} \|f\|. \tag{4.41}$$

Since for each $\varepsilon > 0$ we can find a compact set C such that $\sup_{x\in\overline{\mathbb{R}^d}\setminus C} \|f\| \leq \varepsilon$, taking the limit $n \to \infty$ in (4.41) we find that $\limsup_{n\to\infty} |P_{t_n} f(x_n)| \leq \varepsilon$ for each $\varepsilon > 0$, and therefore, by (4.39)

$$\lim_{n\to\infty} P_{t_n} f(x_n) = 0 = \overline{P}_t f(\infty). \tag{4.42}$$

∎

We can use the compactification trick from the previous example more generally. We start with a simple observation. Let E be locally compact but not compact, separable, and metrizable, and let $\overline{E} := E \cup \{\infty\}$ be its one-point compactification. Let $\mathcal{C}_0(E) := \{f \in \mathcal{C}_\mathrm{b}(E) : \lim_{x\to\infty} f(x) = 0\}$ denote the separable Banach space of continuous real functions on E vanishing at infinity, equipped with the supremumnorm.

Lemma 4.9 (Compactification of Markov process)**.** *Assume that* $(\overline{\mathbb{P}}^x)^{x\in\overline{E}}$ *is a Markov process in* $\overline{E}$ *with transition function* $(\overline{P}_t)_{t\geq 0}$*, and that*

(1) *(non-explosion)* $\overline{\mathbb{P}}^x\{X_t, X_{t-} \neq \infty\ \forall t \geq 0\} = 1\ \forall x \neq \infty.$

Let $\mathbb{P}^x$ *and* $P_t(x,\,\cdot\,)$ *denote the restrictions of* $\overline{\mathbb{P}}^x$ *and* $\overline{P}_t(x,\,\cdot\,)$ *to* $\mathcal{D}_E[0,\infty)$ *and* E*, respectively. Then* $(\mathbb{P}^x)^{x\in E}$ *is a Markov process in* E *with transition function* $(P_t)_{t\geq 0}$*. If moreover*

(2) *(non-implosion)* $\overline{\mathbb{P}}^\infty\{X_t = \infty\ \forall t \geq 0\} = 1,$

then for each $t \geq 0$, P_t maps $\mathcal{C}_0(E)$ into itself and $(P_t)_{t\geq 0}$ is a strongly continuous contraction semigroup on $\mathcal{C}_0(E)$.

Proof. The fact that $(\mathbb{P}^x)^{x\in E}$ is a Markov process in E with transition function $(P_t)_{t\geq 0}$ if the Feller process in $\overline{E}$ is non-explosive is almost trivial. We must only show that the event in condition (1) is actually measurable. Since $X = (X_t)_{t\geq 0}$ can be viewed as a random variable with values in $\mathcal{D}_{\overline{E}}[0,\infty)$, it suffices to show that $\mathcal{D}_E[0,\infty)$ is a measurable subset of $\mathcal{D}_{\overline{E}}[0,\infty)$. This follows from the fact that $\mathcal{D}_E[0,\infty)$ is Polish in the induced topology, so that by Proposition 1.24, $\mathcal{D}_E[0,\infty)$ is a countable intersection of open sets in $\mathcal{D}_{\overline{E}}[0,\infty)$.

Observe that there is a natural identification between the space $\mathcal{C}_0(E)$ and the closed subspace of $\mathcal{C}(\overline{E})$ given by $\mathcal{C}_0(\overline{E}) := \{f \in \mathcal{C}(E) : f(\infty) = 0\}$. If the Feller process in $\overline{E}$ is not only non-explosive but also non-implosive, then

$$\overline{P}_t f(\infty) = f(\infty) = 0 \tag{4.43}$$

for each $f \in \mathcal{C}_0(\overline{E}) := \{f \in \mathcal{C}(E) : f(\infty) = 0\}$, which shows that P_t maps $\mathcal{C}_0(E)$ into itself. Since $(\overline{P}_t)_{t\geq 0}$ is a strongly continuous contraction semigroup on $\mathcal{C}(E)$ its restriction to the closed subspace $\mathcal{C}_0(\overline{E})$ is also a strongly continuous contraction semigroup. ∎

The next Proposition gives sufficient conditions for non-explosion and non-implosion in terms of the generator of a process. The function f in condition (1) is an example of a *Lyapunov function.*

Proposition 4.10 (Non-explosion and non-implosion)**.** *Let $\overline{E}$ be the one-point compactification of a locally compact separable metrizable space E and let $(\overline{\mathbb{P}}^x)^{x\in\overline{E}}$ be a Feller process in $\overline{E}$ with generator $(\mathcal{D}(G), G)$. If*

(1) *There exist functions $f, g : E \to \mathbb{R}$ such that $f \geq 0$, $\lim_{x\to\infty} f(x) = \infty$, and $\sup_{x\in E} g(x) < \infty$, and functions $f_n \in \mathcal{D}(G)$ such that $0 \leq f_n(x) \uparrow f(x)$ for all $x \in E$, $f_n(\infty) \uparrow \infty$, and $Gf_n \to g$ uniformly on compacta in E.*

then $(\overline{\mathbb{P}}^x)^{x\in E}$ is non-explosive. If

(2) *$Gf(\infty) = 0$ for all $f \in \mathcal{D}(G)$,*

then $(\overline{\mathbb{P}}^x)^{x\in E}$ is non-implosive.

Proof. The proof that (1) implies that $(\overline{\mathbb{P}}^x)^{x\in E}$ is non-explosive will be postponed to Section 5.6.

If (2) holds then any solution to the Cauchy equation $\frac{\partial}{\partial t}u(t) = Gu(t)$ satisfies $\frac{\partial}{\partial t}u(t)(\infty) = 0$. Therefore, by Proposition 3.15, $\overline{P}_t f(\infty) = f(\infty)$ for all $f \in \mathcal{D}(G)$ and $t \geq 0$, where $(\overline{P}_t)_{t\geq 0}$ is the semigroup of $(\overline{\mathbb{P}}^x)^{x\in E}$. Since $\mathcal{D}(G)$ is dense it follows that $\overline{P}_t f(\infty) = f(\infty)$ for all $f \in \mathcal{C}(\overline{E})$, $t \geq 0$. This means that $\overline{P}_t(\infty, \cdot\,) = \delta_\infty$ for each $t \geq 0$. It follows that $\overline{\mathbb{P}}^\infty\{X_t = \infty\} = 1$ for all $t \geq 0$, which implies that $\overline{\mathbb{P}}^\infty\{X_t = \infty\ \forall t \in \mathbb{Q} \cap [0,\infty)\} = 1$, and therefore $(\overline{\mathbb{P}}^x)^{x\in E}$ is non-implosive by the right-continuity of sample paths. ∎

Example: (Feller diffusion). Identify $\mathcal{C}[0,\infty]$ with the space $\{f\in\mathcal{C}[0,\infty) : \lim_{x\to\infty} f(x) \text{ exists}\}$ and define an operator $(\mathcal{D}(A_{\rm Fel}), A_{\rm Fel})$ on $\mathcal{C}[0,\infty]$ by

(4.44)
$$\begin{aligned}\mathcal{D}(A_{\rm Fel}) &:= \{f\in\mathcal{C}^2[0,\infty) : \lim_{x\to\infty} f(x) \text{ exists and } \lim_{x\to\infty} x\tfrac{\partial^2}{\partial x^2}f(x)=0\},\\ A_{\rm Fel}f(x) &:= x\tfrac{\partial^2}{\partial x^2}f(x) \qquad (x\in[0,\infty)).\end{aligned}$$

We claim that the closure of $A_{\rm Fel}$ generates the semigroup of a non-explosive and non-implosive Feller process in $[0,\infty]$. It is not hard to check that $A_{\rm Fel}$ satisfies the positive maximum principle. Consider the class of *Laplace functions* $\{f_\lambda\}_{\lambda\geq 0}$ defined by

$$f_\lambda(x) := e^{-\lambda x} \qquad (x\in[0,\infty)). \tag{4.45}$$

We calculate

$$x\tfrac{\partial^2}{\partial x^2}f_\lambda(x) = \lambda^2 x e^{-\lambda x} \underset{x\to\infty}{\longrightarrow} 0, \tag{4.46}$$

which shows that $f_\lambda\in\mathcal{D}(A_{\rm Fel})$ for all $\lambda\geq 0$. By the Stone-Weierstrass theorem (Theorem 3.30), the linear span of $\{f_\lambda\}_{\lambda\geq 0}$ is dense in $\mathcal{C}[0,\infty]$. We claim that for each $\lambda\geq 0$ there exists a solution u to the Cauchy equation

$$\begin{cases}\tfrac{\partial}{\partial t}u(t) = A_{\rm Fel}u(t) \qquad (t\geq 0),\\ \quad u(0) = f_\lambda.\end{cases} \tag{4.47}$$

Indeed, it is easy to see that

$$\tfrac{\partial}{\partial t}e^{-x/t} = x\tfrac{\partial^2}{\partial x^2}\, e^{-x/t} \qquad (t>0,\ x\in[0,\infty)) \tag{4.48}$$

so the solution to (3.59) is given by

$$u(t) = f_{\lambda_t} \quad \text{where} \quad \lambda_t := (\lambda^{-1}+t)^{-1} \qquad (t\geq 0) \tag{4.49}$$

if $\lambda>0$ and $\lambda_t := 0$ $(t\geq 0)$ if $\lambda = 0$. It therefore follows from Theorem 3.29 and Lemma 3.31 that $\overline{A}_{\rm Fel}$ generates a Feller semigroup on $\mathcal{C}[0,\infty]$.

Let $(\overline{\mathbb{P}}^x)^{x\in[0,\infty]}$ denote the corresponding Feller process. It is easy to see that the function $f(x)=x$ (with $g(x)=0$) satisfies condition (1) from Proposition 4.10, so $(\overline{\mathbb{P}}^x)^{x\in[0,\infty]}$ is non-explosive. Since $\lim_{x\to\infty} A_{\rm Fel}f(x)=0$ for all $f\in\mathcal{D}(A_{\rm Fel})$ it follows that $(\overline{\mathbb{P}}^x)^{x\in[0,\infty]}$ is also non-implosive.

Let $P_t(x,\cdot\,)$ denote the restriction of $\overline{P}_t(x,\cdot\,)$ to $[0,\infty)$. By what we have just proved, $(P_t)_{t\geq 0}$ is the transition function of a Markov process $(X,(\mathbb{P}^x)^{x\in[0,\infty)})$ on $[0,\infty)$, which is called the *Feller diffusion*.

Formula (4.49) tells us that the semigroup $(P_t)_{t\geq 0}$ maps Laplace functions into itself. Indeed,

$$P_t f_\lambda = f_{\lambda_t} \qquad (t,\lambda\geq 0) \tag{4.50}$$

with λ_t as in (4.49). This is closely related to the *branching property* of the Feller diffusion. By this, we mean that if X^x and X^y are independent versions of the Feller diffusion started in $X_0^x = x$ and $X_0^y = y$, respectively, and X^{x+y} is a version of the Feller diffusion started in $X_0^{x+y} := x + y$, then

$$\mathcal{L}(X_t^x + X_t^y) = \mathcal{L}(X_t^{x+y}) \qquad (t \geq 0). \tag{4.51}$$

To see this, note that (4.50) says that

$$\mathbb{E}^x\big[e^{-\lambda X_t}\big] = e^{-\lambda_t x} \qquad (t, \lambda \geq 0). \tag{4.52}$$

By independence

$$\begin{aligned} \mathbf{E}\big[e^{-\lambda(X_t^x + X_t^y)}\big] &= \mathbf{E}\big[e^{-\lambda X_t^x}\big]\mathbf{E}\big[e^{-\lambda X_t^y}\big] \\ &= e^{-\lambda_t x}e^{-\lambda_t y} = e^{-\lambda_t(x+y)} = \mathbf{E}\big[e^{-\lambda X_t^{x+y}}\big]. \end{aligned} \tag{4.53}$$

Since this holds for all $\lambda \geq 0$ and the linear span of the Laplace functionals is dense, (4.51) follows. ∎

Exercise 4.11. *Let* $(X, (\mathbb{P}^x)^{x \in [0,\infty)})$ *be the Feller diffusion. Calculate the extinction probability* $\mathbb{P}^x[X_t = 0]$ *for each* $t, x \geq 0$.

5. Harmonic functions and martingales

5.1. **Harmonic functions.** Let $(\mathbb{P}^x)^{x\in E}$ be a Feller process with a compact metrizable state space E, Feller semigroup $(P_t)_{t\geq 0}$ and generator $(\mathcal{D}(G), G)$.

Lemma 5.1 (Harmonic functions)**.** *The following conditions on a function $h\in\mathcal{C}(E)$ are equivalent:*

(a) $P_t h = h \qquad \forall t\geq 0.$
(b) $h\in\mathcal{D}(G)$ *and* $Gh = 0.$

Proof. If $P_t h = h$ for all $t\geq 0$ then $\lim_{t\to 0} t^{-1}(P_t h - h) = 0$, so $h\in\mathcal{D}(G)$ and $Gh=0$. Conversely, if $h\in\mathcal{D}(G)$ and $Gh=0$, then the function $u_t := h$ $(t\geq 0)$ solves the Cauchy equation $\frac{\partial}{\partial t}u_t = Gu_t$ with initial condition $u_0 = h$, so by Propositions 3.15 and 3.22 it follows that $P_t h = u_t = h$ for all $t\geq 0$. ∎

A function $h\in\mathcal{C}(E)$ satisfying the equivalent conditions (a) and (b) from Lemma 5.1 is called a *harmonic* function for the Feller process $(\mathbb{P}^x)^{x\in E}$.

Example 5.2 (Harmonic function for Wright-Fisher diffusion)**.** *Let $A_{\rm WF}$ be as in Exercises 3.20 and 3.32 and let $(Y,(\mathbb{P}^y)^{y\in[0,1]})$ be the Wright-Fisher diffusion, i.e., the Feller process with generator $\overline{A}_{\rm WF}$. Then the function $h:[0,1]\to[0,1]$ given by $h(x):=x$ is harmonic for X. As a consequence, the Wright-Fisher diffusion satisfies*

$$\mathbf{E}^x[X_t] = x \qquad (t\geq 0). \tag{5.1}$$

Proof. Since $\frac{1}{2}x(1-x)\frac{\partial^2}{\partial x^2}x = 0$, h satisfies condition (b) from Lemma 5.1. As a consequence, $\mathbf{E}^x[X_t] = P_t h(x) = h(x) = x$ for all $t\geq 0$. ∎

Let $(\mathbb{P}^x)^{x\in E})$ is a Feller process with a compact metrizable state space E and let $h\in\mathcal{C}(E)$ be harmonic. Then, if X is a version of $(\mathbb{P}^x)^{x\in E}$ started in an arbitrary initial law, by condition (a) from Proposition 2.11,

$$\mathbf{E}[h(X_t)|\mathcal{F}^X_s] = P_t h(X_s) = h(X_s) \quad \text{a.s.} \qquad (0\leq s\leq t). \tag{5.2}$$

This motivates the following definitions. By definition, a *filtration* is a family $(\mathcal{F}_t)_{t\geq 0}$ of σ-fields such that $\mathcal{F}_s\subset\mathcal{F}_t$ for all $0\leq s\leq t$. An *$\mathcal{F}_t$-martingale* is a stochastic process M such that M_t is $\mathcal{F}_t$-measurable, $E[|M_t|]<\infty$, and $\mathbf{E}[M_t|\mathcal{F}_s] = M_s$ for all $0\leq s\leq t$. In the next sections we will study filtrations and martingales in more detail.

5.2. **Filtrations.** By definition, a *filtered probability space* is a quadruple $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\geq 0},\mathbf{P})$ such that $(\Omega,\mathcal{F},\mathbf{P})$ is a probability space and $(\mathcal{F}_t)_{t\geq 0}$ is a filtration on Ω with $\mathcal{F}_t\subset\mathcal{F}\ \forall t\geq 0$. For example, if X is a stochastic process, then $(\mathcal{F}^X_t)_{t\geq 0}$, defined in (2.5), is the *filtration generated by X*.

We say that a stochastic process X is *adapted* to a filtration $(\mathcal{F}_t)_{t\geq 0}$, or simply *$\mathcal{F}_t$-adapted*, if X_t is $\mathcal{F}_t$-measurable for each $t\geq 0$.

Definition 5.3 (Progressive processes)**.** *A stochastic process X on $(\Omega, \mathcal{F}, \mathbf{P})$ is said to be* progressively measurable *with respect to $(\mathcal{F}_t)_{t\geq 0}$, or simply $\mathcal{F}_t$-*progressive*, if the map $(s,\omega) \mapsto X_s(\omega)$ from $[0,t]\times\Omega$ into E is $\mathcal{B}[0,t]\times\mathcal{F}_t$-measurable for each $t \geq 0$.*

Exercise 5.4. *Let X be a stochastic process and $(\mathcal{F}_t)_{t\geq 0}$ a filtration. Assume that X is $\mathcal{F}_t$-adapted and that X has right continuous sample paths. Show that X is $\mathcal{F}_t$-progressive. (Hint: adapt the proof of Lemma 1.4.)*

If $(\mathcal{F}_t)_{t\geq 0}$ is a filtration, then

$$\mathcal{F}_{t+} := \bigcap_{s>t} \mathcal{F}_s \qquad (t \geq 0) \tag{5.3}$$

defines a new, larger filtration $(\mathcal{F}_{t+})_{t\geq 0}$. If $\mathcal{F}_{t+} = \mathcal{F}_t \ \forall t \geq 0$ then we say that the filtration $(\mathcal{F}_t)_{t\geq 0}$ is *right continuous*. It is not hard to see that $(\mathcal{F}_{t+})_{t\geq 0}$ is right continuous.

Recall that the *completion* of a σ-field $\mathcal{F}$ with respect to a probability measure $\mathbf{P}$ is the σ-field

$$\overline{\mathcal{F}} := \{A \subset \Omega : \exists B \in \mathcal{F} \text{ s.t. } 1_A = 1_B \text{ a.s.}\}. \tag{5.4}$$

There is a unique extension of the probability measure $\mathbf{P}$ to a probability measure on $\overline{\mathcal{F}}$. If $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbf{P})$ is a filtered probability space then

$$\overline{\mathcal{F}}_t := \{A \subset \Omega : \exists B \in \mathcal{F}_t \text{ s.t. } 1_A = 1_B \text{ a.s.}\} \qquad (t \geq 0) \tag{5.5}$$

defines a new filtration $(\overline{\mathcal{F}}_t)_{t\geq 0}$. If $\overline{\mathcal{F}}_t = \mathcal{F}_t \ \forall t \geq 0$ then we say that the filtration $(\mathcal{F}_t)_{t\geq 0}$ is *complete*.[15] A random variable X with values in a Polish space is $\overline{\mathcal{F}}_t$-measurable if and only if there exists an $\mathcal{F}_t$-measurable random variable Y such that $X = Y$ a.s.

Lemma 5.5 (Usual conditions)**.** *If $(\mathcal{F}_t)_{t\geq 0}$ is a filtration, then*

$$\overline{\mathcal{F}}_{t+} := \bigcap_{s>t} \overline{\mathcal{F}}_s = \{A \subset \Omega : \exists B \in \mathcal{F}_{t+} \text{ s.t. } 1_A = 1_B \text{ a.s.}\} \qquad (t \geq 0) \tag{5.6}$$

defines a complete, right-continuous filtration.

Proof. It is easy to see that $\bigcap_{s>t}\overline{\mathcal{F}}_s$ is right continuous and that $\{A \subset \Omega : \exists B \in \mathcal{F}_{t+} \text{ s.t. } 1_A = 1_B \text{ a.s.}\}$ is complete. To see that the two formulas for $\overline{\mathcal{F}}_{t+}$ in (5.6) are equivalent, observe that $A \in \bigcap_{s>t}\overline{\mathcal{F}}_s \ \Rightarrow \ \forall n \ \exists B_n \in \mathcal{F}_{t+\frac{1}{n}}$ s.t. $1_A = 1_{B_n}$ a.s. Put $1_{B_\infty} := \liminf_m 1_{B_m}$. Then $1_A = 1_{B_\infty}$ a.s. and since $1_{B_\infty} = \liminf_{m\geq n} 1_{B_m}$ we have $B_\infty \in \mathcal{F}_{t+\frac{1}{n}} \ \forall n \ \Rightarrow \ B_\infty \in \mathcal{F}_{t+}$. This shows that $A \in \{A \subset \Omega : \exists B \in \mathcal{F}_{t+} \text{ s.t. } 1_A = 1_B \text{ a.s.}\}$. Conversely, if $\exists B \in \mathcal{F}_{t+}$ s.t. $1_A = 1_B$ a.s., then obviously $A \in \overline{\mathcal{F}}_s$ for all $s > t$ so $A \in \bigcap_{s>t}\overline{\mathcal{F}}_s$. ∎

[15] Warning: $\overline{\mathcal{F}}_t$ is not the same as the completion of the σ-field $\mathcal{F}_t$ with respect to the restriction of $\mathbf{P}$ to $\mathcal{F}_t$. The reason is the class of null sets of the restriction of $\mathbf{P}$ to $\mathcal{F}_t$ is smaller than the class of null sets of $\mathbf{P}$. Because of this fact, some authors prefer to call $(\overline{\mathcal{F}}_t)_{t\geq 0}$ the *augmentation*, rather than the completion, of $(\mathcal{F}_t)_{t\geq 0}$.

A filtration that is complete and right-continuous is said to fulfill the *usual conditions.*

5.3. **Martingales.**

Definition 5.6 (Martingale)**.** *An* $\mathcal{F}_t$-submartingale *is a real-valued stochastic process* M*, adapted to a filtration* $(\mathcal{F}_t)_{t\geq 0}$*, such that* $\mathbf{E}[|M_t|] < \infty \ \forall t \geq 0$ *and*

$$\mathbf{E}[M_t|\mathcal{F}_s] \geq M_s \quad \text{a.s.} \qquad (0 \leq s \leq t). \tag{5.7}$$

A stochastic process M *is called an* $\mathcal{F}_t$-supermartingale *if* $-M$ *is an* $\mathcal{F}_t$*-submartingale, and an* $\mathcal{F}_t$-martingale *if* M *is an* $\mathcal{F}_t$*-submartingale and an* $\mathcal{F}_t$*-supermartingale.*

We can think of an $\mathcal{F}_t$-martingale as a model for a fair game of chance, where M_t is the capital that a player holds at time $t \geq 0$ and $\mathcal{F}_t$ is the information available to the player at that time. Then (5.7) says that if the player holds a capital M_s at time s, then the expected capital that the player will hold at a later time t, given the information at time s, is precisely M_s.

Lemma 5.7 (Martingale filtration)**.** *Let* $(\mathcal{F}_t)_{t\geq 0}$ *and* $(\mathcal{G}_t)_{t\geq 0}$ *be filtrations such that* $\mathcal{F}_t \subset \mathcal{G}_t$ *for all* $t \geq 0$*. Then every* $\mathcal{G}_t$*-submartingale that is* $\mathcal{F}_t$*-adapted is an* $\mathcal{F}_t$*-submartingale.*

Proof. Since $\mathcal{F}_t \subset \mathcal{G}_t$, since M is a $\mathcal{G}_t$-submartingale, and since M is $\mathcal{F}_t$-adapted:

$$\mathbf{E}[M_t|\mathcal{F}_s] = \mathbf{E}\big[\mathbf{E}[M_t|\mathcal{G}_s]\big|\mathcal{F}_s\big] \geq \mathbf{E}[M_s|\mathcal{F}_s] = M_s \quad \text{a.s.} \qquad (0 \leq s \leq t).$$

∎

In particular, if follows that every $\mathcal{F}_t$-submartingale is also an $\mathcal{F}^M_t$-submartingale. If a stochastic process M is an $\mathcal{F}^M_t$-submartingale, $\mathcal{F}^M_t$-supermartingale, or $\mathcal{F}^M_t$-martingale, then we simply say that M is a submartingale, supermartingale, or martingale, respectively.

Note that if $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbf{P})$ is a filtered probability space and M_∞ is a real random variable such that $\mathbf{E}[|M_\infty|] < \infty$, then

$$M_t := \mathbf{E}[M_\infty|\mathcal{F}_t] \qquad (t \geq 0) \tag{5.8}$$

defines an $\mathcal{F}_t$-martingale. This follows from the facts that $\mathbf{E}\big[\big|\mathbf{E}[M_\infty|\mathcal{F}_t]\big|\big] \leq \mathbf{E}\big[\mathbf{E}[|M_\infty||\mathcal{F}_t]\big] = \mathbf{E}[|M_\infty|] < \infty$ and $\mathbf{E}\big[\mathbf{E}[M_\infty|\mathcal{F}_t]\big|\mathcal{F}_s\big] = \mathbf{E}[M_\infty|\mathcal{F}_s]$ for all $0 \leq s \leq t$. Formula (5.8) defines the stochastic process M uniquely up to modifications, since for each fixed t the conditional expectation is unique up to a.s. equality.

These observations raise a number of questions. Do all $\mathcal{F}_t$-martingales have a last element M_∞ as in (5.8)? Can we find modifications of M with cadlag sample paths? Before we address these questions we first pose another one: We know that the conditional expectation $\mathbf{E}[X|\mathcal{F}]$ of a random variable X with respect to a σ-field $\mathcal{F}$ is continuous in X. For example, if $X_n \to X$

in L^p-norm for some $0 \le p < \infty$, then $\mathbf{E}[X_n|\mathcal{F}] \to \mathbf{E}[X|\mathcal{F}]$ in L^p-norm. But how about the continuity of $\mathbf{E}[X|\mathcal{F}]$ in the σ-field $\mathcal{F}$?

Let $(\mathcal{F}_n)_{n\in\mathbb{N}}$ be a sequence of σ-fields. We say that the σ-fields $\mathcal{F}_n$ *decrease* to a limit $\mathcal{F}_\infty$, denoted as $\mathcal{F}_n \downarrow \mathcal{F}_\infty$, if $\mathcal{F}_0 \supset \mathcal{F}_1 \supset \cdots$ and

$$\mathcal{F}_\infty := \bigcap \mathcal{F}_n.$$

Likewise, we say that the σ-fields $\mathcal{F}_n$ *increase* to a limit $\mathcal{F}_\infty$, denoted as $\mathcal{F}_n \uparrow \mathcal{F}_\infty$, if $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots$ and

$$\mathcal{F}_\infty := \sigma\Big(\bigcup \mathcal{F}_n\Big).$$

One has the following theorem. (See [Chu74, Theorem 9.4.8], or [Bil86, Theorems 35.5 and 35.7].)

Theorem 5.8 (Continuity of conditional expectation in the σ-field)**.** *Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and let $(\mathcal{F}_n)_{n\in\mathbb{N}}$ be a sequence of sub-σ-fields of $\mathcal{F}$. Assume that $\mathbf{E}[|X|] < \infty$ and that $\mathcal{F}_n \downarrow \mathcal{F}_\infty$ or $\mathcal{F}_n \uparrow \mathcal{F}_\infty$. Then*

$$\mathbf{E}[X|\mathcal{F}_n] \underset{n\to\infty}{\longrightarrow} \mathbf{E}[X|\mathcal{F}_\infty] \quad \textit{a.s. and in } L^1\textit{-norm.}$$

Corollary 5.9 (Filtration enlargement)**.** *Let $(\mathcal{F}_t)_{t\ge0}$ be a filtration. Then every $\mathcal{F}_t$-submartingale with right continuous sample paths is also an $\overline{\mathcal{F}}_{t+}$-submartingale.*

Proof. By Theorem 5.8 and the right continuity of sample paths, we have $\mathbf{E}[M_t|\overline{\mathcal{F}}_{s+}] = \mathbf{E}[M_t|\mathcal{F}_{s+}] = \lim_{n\to\infty} \mathbf{E}[M_t|\mathcal{F}_{s+\frac{1}{n}}] = \lim_{n\to\infty} M_{s+\frac{1}{n}} = M_s$ a.s. ∎

Coming back to our earlier questions about martingales, here are two answers.

Theorem 5.10 (Modification with cadlag sample paths)**.** *Let $(\mathcal{F}_t)_{t\ge0}$ be a filtration and let M be an $\overline{\mathcal{F}}_{t+}$-submartingale. Assume that $t \mapsto \mathbf{E}[M_t]$ is right continuous. Then M has a modification with cadlag sample paths.*

This result can be found in [KS91, Theorem 1.3.13]. Note that if M is a martingale, then $E[M_t]$ does not depend on t so that in this case $t \mapsto \mathbf{E}[M_t]$ is trivially right continuous. The next result can be found in [KS91, Theorem 1.3.15].

Theorem 5.11 (Submartingale convergence)**.** *Let M be a submartingale with right continuous sample paths, and assume that $\sup_{t\ge0} \mathbf{E}[M_t \vee 0] < \infty$. Then there exists a random variable M_∞ such that $\mathbf{E}[|M_\infty|] < \infty$ and $M_t \underset{t\to\infty}{\longrightarrow} M_\infty$ a.s.*

5.4. **Stopping times.** There is one more result about martingales that is of central importance. Think of a martingale as a fair game of chance. Then formula (5.7) says that the expected gain of a player who stops playing at a fixed time t is zero. But how about players who stop playing at a random time? It turns out that the answer depends on what we mean by a random time. If the information available to the player at time t is $\mathcal{F}_t$, then the decision whether to stop playing should be made on the basis of this information only. This leads to the definition of stopping times.

Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration. By definition, an *$\mathcal{F}_t$-stopping time* is a function $\tau:\Omega\to[0,\infty]$ such that the stochastic process $(1_{\{\tau\leq t\}})_{t\geq 0}$ is $\mathcal{F}_t$-adapted. Obviously, this is equivalent to the statement that the event $\{\tau\leq t\}$ (i.e., the set $\{\omega:\tau(\omega)\leq t\}$) is $\mathcal{F}_t$-measurable for each $t\geq 0$. We interpret τ as a random time with the property that, if $\mathcal{F}_t$ is the information that is available to us at time t, then we can at any time t decide whether the stopping time τ has already occurred.

Lemma 5.12 (Optional times). *Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration on Ω and let $\tau:\Omega\to[0,\infty]$ be a function. Then τ is an $\mathcal{F}_{t+}$-stopping time if and only if $\{\tau<t\}\in\mathcal{F}_t$ $\forall t\geq 0$.*

Proof. If τ is an $\mathcal{F}_{t+}$-stopping time then $\{\tau\leq s\}\in\bigcap_{t>s}\mathcal{F}_t$ $\forall s\geq 0$, hence $\{\tau\leq s\}\in\mathcal{F}_t$ $\forall t>s\geq 0$. Therefore, for each $t\geq 0$ we can choose $s_n\uparrow t$ to see that $\{\tau<t\}=\bigcup_n\{\tau\leq s_n\}\in\mathcal{F}_t$ $\forall t\geq 0$. Conversely, if $\{\tau<t\}\in\mathcal{F}_t$ $\forall t\geq 0$, then for each $t>s\geq 0$ we can choose $t>u_n\downarrow s$ to see that $\{\tau\leq s\}=\bigcap_n\{\tau<u_n\}\in\mathcal{F}_t$, hence $\{\tau\leq s\}\in\bigcap_{t>s}\mathcal{F}_t=:\mathcal{F}_{s+}$ $\forall s\geq 0$. ∎

$\mathcal{F}_{t+}$-stopping times are also called *optional times*.

Lemma 5.13 (Stopped process). *Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration, let τ be an $\mathcal{F}_{t+}$-stopping time, and let X be an $\mathcal{F}_t$-progressive stochastic process. Then $(X_{t\wedge\tau})_{t\geq 0}$ is $\mathcal{F}_t$-progressive. If $\tau<\infty$ then X_τ is measurable.*

Proof. The fact that X is progressive means that for each $t\geq 0$, the map $(s,\omega)\mapsto X_s(\omega)$ from $[0,t]\times\Omega$ to E is $\mathcal{B}[0,t]\times\mathcal{F}_t$-measurable. We need to show that $(s,\omega)\mapsto X_{s\wedge\tau(\omega)}(\omega)$ is $\mathcal{B}[0,t]\times\mathcal{F}_t$-measurable. It suffices to show that $(s,\omega)\mapsto s\wedge\tau(\omega)$ is measurable with respect to $\mathcal{B}[0,t]\times\mathcal{F}_t$ and $\mathcal{B}[0,t]$. Then $(s,\omega)\mapsto(s\wedge\tau(\omega),\omega)\mapsto X_{s\wedge\tau(\omega)}(\omega)$ from $[0,t]\times\Omega\to[0,t]\times\Omega\to E$ is measurable with respect to $\mathcal{B}[0,t]\times\mathcal{F}_t$, $\mathcal{B}[0,t]\times\mathcal{F}_t$, and $\mathcal{B}(E)$. Now, for any $0<s<t$ one has $\{(s,\omega):s\wedge\tau(\omega)<u\}=\{(s,\omega):s<u\}\cap\{(s,\omega):\tau(\omega)<u\}=([0,u)\times\Omega)\cap([0,t]\times\{\omega:\tau(\omega)<u\})\in\mathcal{B}[0,t]\times\mathcal{F}_t$, which proves that $(s,\omega)\mapsto s\wedge\tau(\omega)$ is measurable with respect to $\mathcal{B}[0,t]\times\mathcal{F}_t$ and $\mathcal{B}[0,t]$.

If $\tau<\infty$ (i.e., $\tau(\omega)<\infty$ for all $\omega\in\Omega$), then it follows that $X_\tau=\lim_{n\to\infty}X_{n\wedge\tau}$ is measurable. ∎

Lemma 5.14 (Operations with stopping times). *Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration.*

(1) *If τ, σ are $\mathcal{F}_t$-stopping times, then $\tau \wedge \sigma$ is an $\mathcal{F}_t$-stopping time.*
(2) *If τ_n are $\mathcal{F}_t$-stopping times, then $\sup_n \tau_n$ is an $\mathcal{F}_t$-stopping time.*
(3) *If τ_n are $\mathcal{F}_{t+}$-stopping times such that $\tau_n \uparrow \tau$ and $\tau_n < \tau$ $\forall n$, then τ is an $\mathcal{F}_t$-stopping time.*

Proof. To prove (1), note that $\{\tau \wedge \sigma \leq t\} = \{\tau \leq t\} \cup \{\sigma \leq t\} \in \mathcal{F}_t$ $\forall t \geq 0$. To prove (2), note that $\{\sup_n \tau_n \leq t\} = \bigcap_n \{\tau_n \leq t\} \in \mathcal{F}_t$ $\forall t \geq 0$. To prove (3), finally, note that in this case $\{\tau \leq t\} = \bigcap_n \{\tau_n < t\} \in \mathcal{F}_t$ $\forall t \geq 0$. ∎

A typical example of a stopping time is a first entrance time. Let E be a Polish space and let X be an E-valued stochastic process. For any $\Delta \subset E$, the *first entrance time* of X into Δ is defined as

$$\tau_\Delta := \inf\{t \geq 0 : X_t \in \Delta\}, \tag{5.9}$$

where $\tau(\omega) := \infty$ if $\{t \geq 0 : X_t(\omega) \in \Delta\} = \emptyset$. Note that $X_{\tau_\Delta} \in \Delta$ if $\tau_\Delta < \infty$, X has right continuous sample paths, and Δ is closed.

Proposition 5.15 (First entrance times)**.** *Let X have cadlag sample paths. If Δ is closed, then τ_Δ is an $\mathcal{F}^X_t$-stopping time.*

Proof. For each $t \geq 0$, define a map $S_t : \mathcal{D}_E[0,\infty) \to \mathcal{D}_E[0,\infty)$ by

$$(S_t(w))_s := w_{s \wedge t} \qquad (s \geq 0).$$

Then $S_t(X)$ is the process X stopped at time t. We claim that $S_t(X) : \Omega \to \mathcal{D}_E[0,\infty)$ is $\mathcal{F}^X_t$-measurable. This follows from the facts that the Borel-σ-field on $\mathcal{D}_E[0,\infty)$ is generated by the coordinate projections $(\pi_s)_{s\geq 0}$ (Proposition 1.16) and that $S_t(X)^{-1}(\pi_s^{-1}(A)) = (S_t(w))_s^{-1}(A) = X^{-1}_{s\wedge t}(A) \in \mathcal{F}^X_t$ for each $s \geq 0$ and $A \in \mathcal{B}(E)$. Since $E\backslash\Delta$ is an open subset of E it is Polish, hence the space $\mathcal{D}_{E\backslash\Delta}[0,\infty)$ is a Polish subspace of $\mathcal{D}_E[0,\infty)$, and therefore, by Proposition 1.24 (b), a countable intersection of open subsets of $\mathcal{D}_E[0,\infty)$. In particular, $\mathcal{D}_{E\backslash\Delta}[0,\infty)$ is a measurable subset of $\mathcal{D}_E[0,\infty)$, and therefore $\{\tau \leq t\} = \{S_t(X) \notin \mathcal{D}_{E\backslash\Delta}[0,\infty)\} \in \mathcal{F}^X_t$ for each $t \geq 0$. ∎

The next theorem shows what happens to a player who stops playing at a stopping time τ. For a proof, see for example [KS91, Theorem 1.3.22].

Theorem 5.16 (Optional sampling)**.** *Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration, let M be an $\mathcal{F}_t$-submartingale with right continuous sample paths, and let τ be an $\mathcal{F}_t$-stopping time such that $\tau \leq T$ for some $T < \infty$. Then*

$$\mathbf{E}[M_\tau] \geq M_0 \quad \text{a.s.} \tag{5.10}$$

5.5. **Applications.** The next example gives an application of Theorem 5.11.

Example 5.17 (Convergence of the Wright-Fisher diffusion)**.** *Let X^x be a version of the Wright-Fisher diffusion started in $X^x_0 = x \in [0,1]$. Then there exists a random variable X^x_∞ such that $\mathbf{E}[X^x_\infty] = x$ and*

$$\lim_{t\to\infty} X^x_t = X^x_\infty \quad \text{a.s.}$$

Proof. It follows from Example 5.2 and (5.2) that X is a nonnegative martingale. Therefore, by Theorem 5.11, there exists a random variable X_∞ such that $X_t \to X_\infty$ a.s. It follows from (5.1) and bounded convergence that $\mathbf{E}[X_\infty] = x$. ∎

Exercise 5.17 leaves a number of questions open. It is not hard to see that the boundary points $\{0,1\}$ are traps for the Wright-Fisher diffusion, in the sense that $\mathbf{P}^x[X_t = x \ \forall t \geq 0] = 1$ if $x \in \{0,1\}$. Therefore, we ask: is it true that the random variable X_∞ from Example 5.17 takes values in $\{0,1\}$? Does the Wright-Fisher diffusion reach the traps in finite time? In order to answer these questions, we need one more piece of general theory.

Let $(X, (\mathbb{P}^x)^{x\in E})$ be a Feller process on a compact metrizable space E and with generator G. If $h \in \mathcal{D}(G)$ satisfies $Gh = 0$ then it follows from Lemma 5.1 and formula (5.2) that $(h(X_t))_{t\geq 0}$ is an $\mathcal{F}^X_t$-martingale. Even if a function $f \in \mathcal{D}(G)$ does not satisfy $Gf = 0$, we can still associate a martingale with f.

Proposition 5.18 (Martingale problem)**.** *Let X be a version of a Feller process with generator $(\mathcal{D}(G), G)$, started in any initial law. Then, for every $f \in \mathcal{D}(G)$, the process M^f given by*

$$M^f_t := f(X_t) - \int_0^t Gf(X_s)\mathrm{d}s \qquad (t \geq 0) \tag{5.11}$$

is an $\mathcal{F}^X_t$-martingale.

Proof.

$$\begin{aligned}
\mathbf{E}[M^f_t|\mathcal{F}_u] &= \mathbf{E}[f(X_t)|\mathcal{F}_u] - \int_0^t \mathbf{E}[Gf(X_s)|\mathcal{F}_u]\mathrm{d}s \\
&= P_{t-u}f(X_u) - \int_0^u Gf(X_s)\mathrm{d}s - \int_u^t P_{s-u}Gf(X_u)\mathrm{d}s \\
&= f(X_u) - \int_0^u Gf(X_s)\mathrm{d}s = M^f_u,
\end{aligned}$$

where we have used that

$$\int_0^t P_sGf = \int_0^t \tfrac{\partial}{\partial s}P_sf = P_tf - f$$

by Proposition 3.15. ∎

The next two examples give applications of Proposition 5.18.

Example 5.19 (Wright-Fisher diffusion converges to traps)**.** *Let X^x be a version of the Wright-Fisher diffusion started in $X^x_0 = x \in [0,1]$. Then the random variable X^x_∞ from Example 5.17 is $\{0,1\}$-valued.*

Proof. Denote the Wright-Fisher diffusion by $(X, (\mathbb{P}^x)^{x\in E})$. The function $f(x) := x^2$ satisfies $f \in \mathcal{D}(A_{\rm WF})$ and $A_{\rm WF} f(x) = x(1-x)$. Therefore, by Proposition 5.18,

$$\mathbb{E}^x[X_t^2] = \int_0^t \mathbb{E}^x[X_s(1-X_s)]\,\mathrm{d}s \qquad (t \geq 0).$$

Since $X_t \in [0,1]$ it follows, letting $t \to \infty$, that

$$\mathbb{E}^x\Big[\int_0^\infty X_s(1-X_s)\,\mathrm{d}s\Big] \leq 1.$$

In particular, $\int_0^\infty X_s(1-X_s)\mathrm{d}s$ is finite a.s., which is possible only if $X_\infty \in \{0,1\}$ a.s. ∎

Example 5.20 (Wright-Fisher diffusion gets trapped in finite time)**.** *Let X^x be a version of the Wright-Fisher diffusion started in $X^x_0 = x \in [0,1]$. Define (using Proposition 5.15) an $\mathcal{F}^X_t$-stopping time τ by*

$$\tau := \inf\big\{t \geq 0 : X^x_t \in \{0,1\}\big\}.$$

Then $\mathbf{E}[\tau] < \infty$.

Proof. Let $(X, (\mathbb{P}^x)^{x\in E})$ be the Wright-Fisher diffusion. The idea of the proof is to show that there exists a continuous function $f : [0,1] \to [0,\infty)$ such $f(0) = f(1) = 0$ and the process

$$M_t := f(X_t) + \int_0^t 1_{(0,1)}(X_s)\mathrm{d}s \qquad (t \geq 0) \tag{5.12}$$

is an $\mathcal{F}^X_t$-martingale. Let us first explain why we are interested in such a function. If the process in (5.12) is a martingale, then by optional sampling (Theorem 5.16), $\mathbb{E}^x[M_{\tau\wedge t}] = \mathbb{E}^x[M_0]$, hence

$$\mathbb{E}^x[\tau\wedge t] = \mathbb{E}^x\Big[\int_0^{\tau\wedge t} 1_{(0,1)}(X_s)\mathrm{d}s\Big] = f(x) - \mathbb{E}^x[f(X_{\tau\wedge t})] \qquad (t \geq 0).$$

Letting $t \uparrow \infty$ we see that $\mathbb{E}^x[\tau] \leq f(x)$, so $\tau < \infty$ a.s. Since f is zero on $\{0,1\}$ it follows that $\mathbb{E}^x[f(X_{\tau\wedge t})] \to 0$ as $t \uparrow \infty$, so we find that

$$\mathbb{E}^x[\tau] = f(x) \qquad (x \in [0,1]). \tag{5.13}$$

To get a function f such that (5.12) holds, we choose $0 < \varepsilon_n < \frac{1}{2}$ such that $\varepsilon_n \downarrow 0$, we define

$$h_n(x) := \begin{cases} -\dfrac{2}{x(1-x)} & (x \in (\varepsilon_n, 1-\varepsilon_n)), \\ -\dfrac{2\varepsilon_n}{\varepsilon_n(1-\varepsilon_n)} & (x \in [0,\varepsilon_n] \cup [1-\varepsilon_n, 1]), \end{cases}$$

and we put

$$f_n(x) := \int_0^x \mathrm{d}y \int_{\frac{1}{2}}^y \mathrm{d}z\, h_n(z) \qquad (x \in [0,1]).$$

Then the functions $f_n : [0,1] \to \mathbb{R}$ are continuous, symmetric in the sense that $f_n(x) = f_n(1-x)$, and satisfy $f_n(0) = f_n(1) = 0$. Moreover, we have $f_n \uparrow f$, where

$$f(x) := \int_0^x \mathrm{d}y \int_{\frac{1}{2}}^y \mathrm{d}z\, \frac{-2}{z(1-z)} \qquad (x \in [0,1]).$$

To see that this is finite, note that for $y \le \frac{1}{2}$,

$$\int_{\frac{1}{2}}^y \mathrm{d}z\, \frac{-2}{z(1-z)} = \int_y^{\frac{1}{2}} \mathrm{d}z\, \frac{2}{z(1-z)} \le 4 \int_y^{\frac{1}{2}} \frac{\mathrm{d}z}{z} = 4\big(\log(\tfrac{1}{2}) - \log(y)\big),$$

which is integrable at zero. The functions f_n satisfy

$$A_{\mathrm{WF}} f_n(x) = \tfrac{1}{2} x(1-x) h_n(x) \downarrow_{n\to\infty} -1_{(0,1)}(x) \qquad (x \in [0,1]).$$

The fact that the process M in (5.12) is a martingale now follows from Proposition 5.18 and Lemma 5.21 below. ∎

We say that a sequence of bounded real functions f_n, defined on a measurable space, converges to a *bounded pointwise limit* f, if $f_n \to f$ pointwise while $\sup_n \|f_n\| < \infty$. We denote this as

$$f = \mathrm{bp}\lim_{n\to\infty} f_n.$$

Recall that the integral is continuous with respect to bounded pointwise convergence. So, if X_n are real-valued random variables and $X_n \to X$, then $\mathbf{E}[X_n] \to \mathbf{E}[X]$.

Lemma 5.21 (Bounded pointwise limits)**.** *Let X be a Feller process on a compact metrizable space E and let G be its generator. Let $f_n \in \mathcal{D}(G)$, $f \in \mathcal{C}(E)$, and $g \in B(E)$ be functions such that*

$$f = \mathrm{bp}\lim_{n\to\infty} f_n \quad \textit{and} \quad g = \mathrm{bp}\lim_{n\to\infty} G f_n.$$

Then the process M given by

$$M_t := f(X_t) - \int_0^t g(X_s)\mathrm{d}s \qquad (t \ge 0)$$

is an $\mathcal{F}_t^X$-martingale.

Proof. We know that the processes

$$M_t^{(n)} := f_n(X_t) - \int_0^t G f_n(X_s)\mathrm{d}s \qquad (t \ge 0)$$

are $\mathcal{F}_t^X$-martingales. In particular, $\mathbf{E}[M_t^{(n)} | \mathcal{F}_s^X] = M_s^{(n)}$ a.s. for all $0 \le s \le t$. By the definition of the conditional expectation, this is equivalent to the fact that

(1) $M^{(n)}_s$ is $\mathcal{F}^X_s$-measurable,
(2) $\mathbf{E}[M^{(n)}_t 1_A] = \mathbf{E}[M^{(n)}_s 1_A]\ \forall A \in \mathcal{F}^X_s$,

for all $0 \leq s \leq t$. For each fixed $t \geq 0$, we observe that

$$M_t = \operatorname{bp}\lim_{n\to\infty} M^{(n)}_t.$$

It follows that

(1) M_s is $\mathcal{F}^X_s$-measurable,
(2) $\mathbf{E}[M_t 1_A] = \mathbf{E}[M_s 1_A]\ \forall A \in \mathcal{F}^X_s$,

which proves that $\mathbf{E}[M_t|\mathcal{F}^X_s] = M_s$ a.s. for all $0 \leq s \leq t$. ∎

5.6. **Non-explosion.** Using martingales and stopping times, we can complete the proof of Proposition 4.10 started in Section 4.4. Let $\overline{E}$ be the one-point compactification of a locally compact separable metrizable space E and let $(X, (\mathbb{P}^x)^{x\in E})$ be a Feller process in $\overline{E}$ with generator $(\mathcal{D}(G), G)$. Recall that $(\mathbb{P}^x)^{x\in E}$ is called non-explosive if

$$\overline{\mathbb{P}}^x\{X_t, X_{t-} \neq \infty\ \forall t \geq 0\} = 1 \quad \forall x \neq \infty.$$

Proof of Proposition 4.10 (continued). The fact that condition (2) from Proposition 4.10 implies non-implosion has already been proved in Section 4.4. Assume that condition (1) from Proposition 4.10 holds. For each $R > 0$, put

$$O_R := \{x \in E : f(x) < R\} \tag{5.14}$$

and define stopping times τ_R by

$$\tau_R := \inf\{t \geq 0 : X_t \in E\backslash O_R\} \qquad (R > 0). \tag{5.15}$$

Fix $x \in E$. By Proposition 5.18 and optional stopping, for each $t > 0$,

$$\begin{aligned}
&\mathbb{P}^x\{\tau_R \leq t\} \inf_{x\in E\backslash O_R} f_n(x) \\
&\leq \mathbb{E}^x[f_n(X_{t\wedge\tau_R})] = f(x) + \mathbb{E}^x\Big[\int_0^{t\wedge\tau_R} Gf_n(X_s)\mathrm{d}s\Big] \\
&\leq f(x) + t \sup_{x\in O_R} Gf_n(x),
\end{aligned} \tag{5.16}$$

where in the last inequality we have used that $f(X_s) < R$ for all $s < \tau_R$. Since $\overline{O}_R$ is compact and Gf_n converges uniformly on compacta to g,

$$\limsup_{n\to\infty} \sup_{x\in O_R} Gf_n(x) \leq \sup_{x\in E} g(x). \tag{5.17}$$

We claim that moreover

$$\inf_{x\in E\backslash O_R} f_n(x) \underset{n\to\infty}{\longrightarrow} R. \tag{5.18}$$

Indeed, by our assumptions, the sets $\{x \in \overline{E}\backslash O_R : f_n(x) \leq R - \varepsilon\}$ are compact subsets of $\overline{E}$, decreasing to the empty set. Therefore, for each $\varepsilon > 0$ there exists an n with $\{x \in \overline{E}\backslash O_R : f_n(x) \leq R - \varepsilon\} = \emptyset$.

Inserting (5.17) and (5.18) into (5.16), we find that

$$\mathbb{P}^x\{\tau_R \leq t\} \leq R^{-1}\big(f(x) + t\sup_{x\in E} g(x)\big). \tag{5.19}$$

Letting $R \uparrow \infty$ shows that

$$\mathbb{P}^x\{X_s, X_{s-} \neq \infty \ \forall s \leq t\} = 1 \tag{5.20}$$

for each fixed $t > 0$. Letting $t \uparrow \infty$ shows that $(\mathbb{P}^x)^{x\in E}$ is non-explosive. ∎

6. Convergence of Markov processes

6.1. **Convergence in path space.** In this section, we discuss the convergence of a sequence of Feller processes to a limiting Feller process. The martingale problem from Proposition 5.18 will play an important role in the proofs. As an application of our main result, we will complete the proof of Theorem 4.2.

Let $\mathcal{G}_n$ be multivalued linear operators on a Banach space V, i.e., the $\mathcal{G}_n$ are linear subspaces of $V \times V$. We define the *extended limit* $\mathrm{ex}\lim_{n\to\infty} \mathcal{G}_n$ as

$$\mathrm{ex}\lim_{n\to\infty} \mathcal{G}_n := \{(f,g) : \exists (f_n, g_n) \in \mathcal{G}_n \text{ s.t. } (f_n, g_n) \underset{n\to\infty}{\longrightarrow} (f,g)\}. \tag{6.1}$$

If the $\mathcal{G}_n$ are single-valued, and therefore the graphs of some linear operators $(\mathcal{D}(A_n), A_n)$, and moreover $\mathrm{ex}\lim_{n\to\infty} \mathcal{G}_n$ is single-valued and the graph of $(\mathcal{D}(A), A)$, then we also write $\mathrm{ex}\lim_{n\to\infty} A_n = A$.

Exercise 6.1. *Show that* $\mathrm{ex}\lim_{n\to\infty} \mathcal{G}_n$ *is always a closed linear operator. Show that*

(i) $\mathrm{ex}\lim\limits_{n\to\infty} (\lambda_1 + \lambda_2 \mathcal{G}_n) = \lambda_1 + \lambda_2 \mathrm{ex}\lim\limits_{n\to\infty} \mathcal{G}_n$ *for all* $\lambda_1, \lambda_2 \in \mathbb{R},\ \lambda_2 \neq 0.$

(ii) $\mathrm{ex}\lim\limits_{n\to\infty} \mathcal{G}_n^{-1} = (\mathrm{ex}\lim\limits_{n\to\infty} \mathcal{G}_n)^{-1}.$

(iii) *If* $\mathcal{G}_n$ *is dissipative for each* n *then* $\mathrm{ex}\lim_{n\to\infty} \mathcal{G}_n$ *is dissipative.*

Exercise 6.2. *Let* A_n, A *be bounded linear operators. Show that* $Af = \lim_{n\to\infty} A_n f$ *for all* $f \in V$ *implies* $A = \mathrm{ex}\lim_{n\to\infty} A_n$. *Hint: Lemma 3.43.*

The main result of this section is:

Theorem 6.3. (Convergence of Feller processes) *Let* E *be a compact metrizable space and let* $(\mathbb{P}^{(n),x})^{x\in E}$ *and* $(\mathbb{P}^x)^{x\in E}$ *be Feller processes in* E *with Feller semigroups* $(P_t^{(n)})_{t\geq 0}$ *and* $(P_t)_{t\geq 0}$ *and generators* G_n *and* G, *respectively. Then the following statements are equivalent:*

(a) $\mathrm{ex}\lim\limits_{n\to\infty} G_n \supset G.$

(b) $\mathrm{ex}\lim\limits_{n\to\infty} G_n = G.$

(c) $P_t^{(n)} f \underset{n\to\infty}{\longrightarrow} P_t f$ *for all* $f \in \mathcal{C}(E)$ *and* $t \geq 0$.

(d) $\mathbb{P}^{(n),\mu_n}\{(X_{t_1}, \ldots, X_{t_m}) \in \cdot\} \underset{n\to\infty}{\Longrightarrow} \mathbb{P}^{(n),\mu}\{(X_{t_1}, \ldots, X_{t_m}) \in \cdot\}$ *whenever* $\mu_n \underset{n\to\infty}{\Longrightarrow} \mu$.

(e) $\mathbb{P}^{(n),\mu_n} \underset{n\to\infty}{\Longrightarrow} \mathbb{P}^{\mu}$ *whenever* $\mu_n \underset{n\to\infty}{\Longrightarrow} \mu$.

Condition (a) means that $\mathrm{ex}\lim_{n\to\infty} G_n$, considered as a multivalued operator, contains G. Thus, (a) says that for all $f \in \mathcal{D}(G)$ there exist $f_n \in \mathcal{D}(G_n)$ such that $f_n \to f$ and $G_n f_n \to Gf$. We can reformulate conditions (d) and (e) as follows. Let $X^{(n)}$ and X be random variables with laws $\mathbb{P}^{(n),\mu_n}$ and $\mathbb{P}^{\mu}$, respectively, i.e., $X^{(n)}$ is a version of the Markov process with semigroup $(P_t^{(n)})_{t\geq 0}$, started in the initial law $\mathcal{L}(X_0^{(n)}) = \mu_n$ and X is a version of the Markov process with semigroup $(P_t)_{t\geq 0}$, started in the initial law $\mathcal{L}(X_0) = \mu$.

Then condition (d) says that $\mu_n \Rightarrow \mu$ implies that $X^{(n)}$ converges to X in finite dimensional distributions, and (e) says that $\mu_n \Rightarrow \mu$ implies that $\mathcal{L}(X^{(n)}) \Rightarrow \mathcal{L}(X)$, where $\mathcal{L}(X^{(n)})$ and $\mathcal{L}(X)$ are probability measures on the 'path space' $\mathcal{D}_E[0,\infty)$. In this case we say that $X^{(n)}$ converges to X in the sense of weak convergence in path space.

Under weak additional assumptions, weak convergence in path space implies convergence in finite dimensional distributions.

Lemma 6.4. (Converge of finite dimensional distributions) *Let $Y^{(n)}$ and Y be $\mathcal{D}_E[0,\infty)$-valued random variables. Assume that $P\{Y_{t-} = Y_t\} = 1$ for all $t \geq 0$. Then $\mathcal{L}(Y^{(n)}) \Rightarrow \mathcal{L}(Y)$ implies that $\mathcal{L}(Y^{(n)}_{t_1}, \ldots, Y^{(n)}_{t_k}) \Rightarrow \mathcal{L}(Y_{t_1}, \ldots, Y_{t_k})$ for all $0 \leq t_1 \leq \cdots \leq t_k$.*

Proof. See [EK86, Theorem 3.7.8 (a)]. ∎

Exercise 6.5. *Assume that $Y^{(n)}$ and Y are stochastic processes with sample paths in $\mathcal{C}_E[0,\infty)$. Show that weak convergence in path space (of the $Y^{(n)}$ to Y) implies convergence in finite dimensional distributions.*

Weak convergence in paths space is usually a more powerful statement that convergence in finite dimensional distributions (and more difficult to prove). The next example shows that weak convergence in path space is not implied by convergence of finite dimensional distributions.

(Counter-)Example. Let for $n \geq 1$, X^n be the $\{0,1\}$-valued Markov process with infinitesimal matrix (generator)

$$A^{(n)} := \begin{pmatrix} -1 & 1 \\ n & -n \end{pmatrix} \tag{6.2}$$

and initial law $\mathbb{P}\{X^n_0 = 0\} = 1$.

Recall that then the corresponding semigroup is given by

$$\begin{aligned} T^{(n)}_t f &= e^{A^{(n)}t} f \\ &= \big(\mathrm{Id} - \frac{1}{(n+1)} \sum_{k \geq 1} \frac{(-(n+1)t)^k}{k!} A^{(n)}\big) f \\ &= \big(\mathrm{Id} - \frac{1}{(n+1)} (e^{-(n+1)t} - 1) A^{(n)}\big) f \end{aligned} \tag{6.3}$$

where we have used that $(A^{(n)})^k = (-(n+1))^{k-1} A^{(n)}$.

Put $f(0) := 1$ and $f(1) := 0$ then

$$\mathbb{P}^{0,(n)}\{X_t = 0\} = T^{(n)}_t f(0) = 1 - \frac{1}{(n+1)} (1 - e^{-(n+1)t}) \underset{n \to \infty}{\longrightarrow} 1. \tag{6.4}$$

One can iterate the argument to show that finite dimensional distributions of X under $\mathbb{P}^{0,(n)}$ converge to those of a process that is identical 1. One the other hand, $\mathbb{P}^{0,(n)}$ is supported on the set of paths which $\mathbb{E}^{0,(n)}[\tau] = 1$, where $\tau := \inf\{t \geq 0 : X_t \in \{1\}\} < \infty$. Hence, the sequence $\mathcal{L}(X^n)$ does not converge in the sense of weak convergence on $\mathcal{D}_{\{0,1\}}[0,\infty)$. ∎

Exercise 6.6. *Let Y be a Poisson process with parameter λ, and define*

$$X^n_t := \frac{1}{n}\big(Y_{n^2 t} - \lambda n^2 t\big). \tag{6.5}$$

apply Theorem 6.3 to show that $\{X^n\}$ converges in distribution and identify its limit.

The fact that conditions (a), (b) and (c) from Theorem 6.3 are equivalent follows from abstract semigroup theory. We will only prove the easy implication (c)⇒(b). For a full proof, see [EK86, Theorem 1.6.1].

Proposition 6.7. (Convergence of semigroups) *Assume that $(S^{(n)}_t)_{t\geq 0}$ and $(S_t)_{t\geq 0}$ are strongly continuous contraction semigroups on a Banach space V, with generators G_n and G, respectively. Then the following statements are equivalent:*

(a) $\operatorname*{ex\,lim}_{n\to\infty} G_n \supset G$.
(b) $\operatorname*{ex\,lim}_{n\to\infty} G_n = G$.
(c) $S^{(n)}_t f \underset{n\to\infty}{\longrightarrow} S_t f$ *for all* $f \in V$ *and* $t \geq 0$.

Proof. (c)⇒(b): Fix $\lambda > 0$. By Lemma 3.21, $(\lambda - G)^{-1}$ is a bounded linear operator which is given by

$$(\lambda - G)^{-1} f = \int_0^\infty S_t f e^{-\lambda t} \mathrm{d}t \qquad (f \in V).$$

A similar formula holds for $(\lambda - G_n)^{-1}$. Since S_t and $S^{(n)}_t$ are contractions, $\|S_t f - S^{(n)}_t f\| \leq 2\|f\|$, so using bounded convergence

$$\|(\lambda - G)^{-1} f - (\lambda - G_n)^{-1} f\| \leq \int_0^\infty \|S_t f - S^{(n)}_t f\| e^{-\lambda t} \mathrm{d}t \underset{n\to\infty}{\longrightarrow} 0.$$

By Exercise 6.2 this proves that

$$\operatorname*{ex\,lim}_{n\to\infty} (\lambda - G_n)^{-1} = (\lambda - G)^{-1}.$$

By Exercise 6.1, it follows that $\operatorname{ex\,lim}_{n\to\infty} G_n = G$.

Since (b)⇒(a) is trivial, to complete the proof it suffices to prove that (a)⇒(c). This implication is more difficult. One proves that the Yosida approximations G_ε and $G_{n,\varepsilon}$ of G and G_n satisfy $G_{n,\varepsilon} f \to G_\varepsilon f$ for each $f \in V$ and $\varepsilon > 0$, uses this to derive estimates that are uniform in ε, and then lets $\varepsilon \to 0$. ∎

The main technical tool in the proof of Theorem 6.3 is a tightness criterion for sequences of probability laws on $\mathcal{D}_E[0,\infty)$, which we will not prove. Recall the concept of tightness from Proposition 3.2. To stress the importance of tightness, we note the following fact.

Lemma 6.8. (Application of tightness) *Let $Y^{(n)}$ be a sequence of processes with sample paths in $\mathcal{D}_E[0,\infty)$. Assume that the finite dimensional distributions of $Y^{(n)}$ converge and that the laws $\mathcal{L}(Y^{(n)})$ are tight. Then there exists a process Y with sample paths in $\mathcal{D}_E[0,\infty)$ such that $\mathcal{L}(Y^{(n)}) \Rightarrow \mathcal{L}(Y)$.*

Proof. The weak limits $\lim_{n\to\infty} \mathcal{L}(Y^{(n)}_{t_1},\ldots,Y^{(n)}_{t_n})$ form a consistent family in the sense of Kolmogorov's extension theorem, so by the latter there exists an E-valued process Y' such that the $Y^{(n)}$ converge to Y' in finite dimensional distributions. Since the laws $\mathcal{L}(Y^{(n)})$ are tight, we can select a convergent subsequence $\mathcal{L}(Y^{(n_m)}) \Rightarrow \mathcal{L}(Y)$. If we can show that all convergent subsequences have the same limit $\mathcal{L}(Y)$, then by the exercise below, the laws $\mathcal{L}(Y^{(n)})$ converge to $\mathcal{L}(Y)$.

For any function $f \in \mathcal{C}(E)$ and $0 \le t < u$, the map $w \mapsto \int_t^u f(w(s))\mathrm{d}s$ from $\mathcal{D}_E[0,\infty)$ to $\mathbb{R}$ is bounded and continuous. (Note that the coordinate projections are not continuous!) Therefore, $\mathcal{L}(Y^{(n_m)}) \Rightarrow \mathcal{L}(Y)$ implies that $E[\int_t^u f(Y^{(n_m)}_s)\mathrm{d}s] \to E[\int_t^u f(Y_s)\mathrm{d}s]$ for each $t \ge 0$ and $\varepsilon > 0$. Moreover $E[\int_t^u f(Y^{(n_m)}_s)\mathrm{d}s] = \int_t^u E[f(Y^{(n_m)}_s)]\mathrm{d}s \to \int_t^u E[f(Y'_s)]\mathrm{d}s$ by bounded convergence, so by the right-continuity of sample paths

$$E[f(Y_t)] = \lim_{\varepsilon\to 0} E\big[\varepsilon^{-1}\int_0^\varepsilon f(Y_{t+s})\mathrm{d}s\big] = \lim_{\varepsilon\to 0} \varepsilon^{-1}\int_0^\varepsilon E[f(Y'_{t+s})]\mathrm{d}s.$$

A similar argument shows that

$$E[f_1(Y_{t_1})\cdots f_k(Y_{t_k})] = \lim_{\varepsilon\to 0}\varepsilon^{-1}\int_0^\varepsilon E[f_1(Y'_{t_1+s})\cdots f_k(Y'_{t_k+s})]\mathrm{d}s. \tag{6.6}$$

for any $f_1,\ldots,f_k \in \mathcal{C}(E)$ and $0 \le t_1 \le \cdots \le t_k$. This clearly determines the finite dimensional distributions of Y, and therefore $\mathcal{L}(Y)$, uniquely. (Warning: the finite dimensional distributions of Y and Y' need in general not be the same!) ∎

Exercise 6.9. *Let M be a metrizable space and let $(x_n)_{n\ge 1}$ be a sequence in M. Assume that the closure of the set $\{x_n : n \ge 1\}$ is compact and that the sequence $(x_n)_{n\ge1}$ has only one cluster point x. Show that $x_n \to x$.*

The next theorem relates tightness of probability measures on $\mathcal{D}_E[0,\infty)$ to martingales in the spirit of Proposition 5.18. Below, for any measurable function $h : [0,\infty) \to \mathbb{R}$, $T > 0$, and $p \in [1,\infty]$ we define:

$$\|h\|_{p,T} := \begin{cases} \big(\int_0^T |h(t)|^p \mathrm{d}t\big)^{\frac{1}{p}} & \text{if } p < \infty, \\ \operatorname{ess\,sup}_{t\in[0,T]} |h(t)| & \text{if } p = \infty. \end{cases} \tag{6.7}$$

Here the essential supremum is defined as:

$$\operatorname{ess\,sup}_{t\in[0,T]} |h(t)| := \inf\{H \geq 0 : |h(t)| \leq H \text{ a.s.}\},$$

where a.s. means almost surely with respect to Lebesgue measure. Thus, $\|h\|_{p,T}$ is just the L_p-norm of the function $[0,T] \ni t \mapsto h(t)$ with respect to Lebesgue measure.

Theorem 6.10. (Tightness criterion) *Let E be compact and metrizable and let $\{X^{(n)} : n \geq 1\}$ be a sequence of processes with sample paths in $\mathcal{D}_E[0,\infty)$, defined on probability spaces $(\Omega^{(n)}, \mathbf{P}^{(n)}, \mathcal{F}^{(n)})$ and adapted to filtrations $(\mathcal{F}^{(n)}_t)_{t\geq 0}$. Let $\mathcal{D} \subset \mathcal{C}(E)$ be dense and assume that for all $f \in \mathcal{D}$ and $n \geq 1$ there exist $\mathcal{F}^{(n)}_t$-adapted real processes $F^{(n)}$ and $G^{(n)}$ with cadlag sample paths, such that*

$$M_t := F^{(n)}_t - \int_0^t G^{(n)}_s \mathrm{d}s$$

is an $\mathcal{F}^{(n)}_t$-martingale, and such that for each $T > 0$,

$$\sup_n \mathbf{E}^{(n)}\big[\sup_{t\in[0,T]\cap\mathbb{Q}} |F^{(n)}_t - f(X^{(n)}_t)|\big] < \infty \tag{6.8}$$

and

$$\sup_n \mathbf{E}^{(n)}\big[\|G^{(n)}\|_{p,T}\big] < \infty \quad \textit{for some } p \in (1,\infty]. \tag{6.9}$$

Then the laws $\{\mathcal{L}(X^{(n)}) : n \geq 1\}$ are tight.

Proof. This is a much simplified version of Theorems 3.9.1 and 3.9.4 in [EK86]. ∎

Remark. For example, if $X^{(n)}$ is a Feller process with generator G_n and $f_n \in \mathcal{D}(G_n)$, then by Proposition 5.18, $M_t := f_n(X^{(n)}_t) - \int_0^t G_n f_n(X^{(n)}_s)\mathrm{d}s$ is an $\mathcal{F}^{X^{(n)}}_t$-martingale. Thus, a typical application of Theorem 6.10 is to take $F^{(n)}_t := f_n(X^{(n)}_t)$ and $G^{(n)}_t := G_n f_n(X^{(n)}_t)$. ∎

Counterexample. Taking $p = 1$ in (6.9) is not sufficient. To see this, for $n \geq 1$ let $X^{(n)}$ be the Markov process with generator (6.2) and initial law $\mathbb{P}^{(n)}\{X^{(n)}_0 = 0\} = 1$. Take for $\mathcal{D}$ the space of all real functions f on $\{0,1\}$ and for such a function put $F^{(n)}_t := f(X^{(n)}_t)$ and $G^{(n)}_t := A_n f(X^{(n)}_t)$. Then by Proposition 5.18, $F^{(n)}_t - \int_0^t G^{(n)}_s \mathrm{d}s$ is an $\mathcal{F}^{X^{(n)}}_t$-martingale, (6.8) is satisfies, and by (6.4)

$$\begin{aligned}
&\mathbf{E}\big[|g_n(X^{(n)}_t)|\big] = \mathbf{E}\big[|A^{(n)} f(X^{(n)}_t)|\big] \\
&\quad = n|f(0) - f(1)|\mathbf{P}\{X^{(n)}_t = 1\} + |f(0) - f(1)|\mathbf{P}\{X^{(n)}_t = 0\} \\
&\quad = |f(0) - f(1)|\Big(1 + \frac{n-1}{n+1}(1 - e^{-(n+1)t})\Big) \leq 2|f(0) - f(1)|.
\end{aligned}$$

This shows that

$$\sup_n \mathbf{E}\big[\|g_n(X^{(n)})\|_{1,T}\big] = \sup_n \mathbf{E}\big[\int_0^T |g_n(X_t^{(n)})|\mathrm{d}s\big] \le 2T|f(0)-f(1)| < \infty,$$

so (6.9) is satisfied for $p = 1$. Since the $X^{(n)}$ converge in finite dimensional distributions, if the laws $\{\mathcal{L}(X^{(n)}) : n \ge 1\}$ were tight, then $X^{(n)}$ would also converge weakly in path space. We have already seen that this is not the case. ∎

Proof of Theorem 6.3. Conditions (a), (b) and (c) are equivalent by Proposition 6.7. Our next step is to show that (c) is equivalent to (d). Indeed, if (c) holds, then for any $f_1, \ldots, f_n \in \mathcal{C}(E)$ and $0 = t_0 \le \cdots \le t_k$,

$$\begin{aligned}\mathbb{E}^{(n),\mu_n}\big[f_1(X_{t_1})\cdots f_k(X_{t_k})\big] &= \mu_n P^{(n)}_{t_1-t_0} f_1 \cdots P^{(n)}_{t_k-t_{k-1}} f_k \\ &\underset{n\to\infty}{\longrightarrow} \mu P_{t_1-t_0} f_1 \cdots P_{t_k-t_{k-1}} f_k = \mathbb{E}^{\mu}\big[f_1(X_{t_1})\cdots f_k(X_{t_n})\big],\end{aligned}$$

where we have used Lemma 3.43. This implies (d). Conversely, if (d) holds, then for any $f \in \mathcal{C}(E)$, $x_n \to x$, and $t \ge 0$,

$$P_t^{(n)} f(x_n) = \mathbb{E}^{(n),x_n}[f(X_t)] \underset{n\to\infty}{\longrightarrow} \mathbb{E}^x[f(X_t)] = P_t f(x),$$

which proves that $P_t^{(n)} f$ converges uniformly to $P_t f$ (compare the proof of Proposition 3.7).

To complete the proof, it suffices to show that (a) and (d) imply (e) and that (e) implies (b). (Warning: it is not immediately obvious that (e) implies (d) since weak convergence in path space does not in general imply convergence in finite dimensional distributions.)

(a) & (d)⇒(e): Let $X^{(n)}$ be random variables with laws $\mathbb{P}^{(n),\mu_n}$. We start by showing that the laws $\mathcal{L}(X^{(n)})$ are tight. This is a straightforward application of Theorem 6.10. We choose $\mathcal{D} := \mathcal{D}(G)$, which is dense in $\mathcal{C}(E)$. By (a), for each $f \in \mathcal{D}$ there exist $f_n \in \mathcal{D}(G_n)$ such that $f_n \to f$ and $G_n f_n \to Gf$. Setting $F_t^{(n)} := f_n(X_t^{(n)})$ and $G_t^{(n)} := G_n f_n(X_t^{(n)})$, using Proposition 5.18, we see that (6.8) and (6.8) are satisfied, where in the latter we can take $p = \infty$.

Since the laws $\mathcal{L}(X^{(n)})$ are tight, we can select a convergent subsequence $\mathcal{L}(X^{(n_m)}) \Rightarrow \mathcal{L}(X)$. We are done if we can show that $\mathcal{L}(X) = \mathbb{P}^{\mu}$ (and hence all weak cluster points are the same). In the same way as in the proof of Lemma 6.8 (see in particular (6.6)), we find that

$$\begin{aligned}E[f_1(X_{t_1})\cdots f_k(X_{t_k})] &= \lim_{\varepsilon\to 0} \varepsilon^{-1}\int_0^{\varepsilon} \mathrm{d}s\, \mu P_{t_1-t_0+s} f_1 \cdots P_{t_k-t_{k-1}+s} f_k \\ &\quad \mu P_{t_1-t_0} f_1 \cdots P_{t_k-t_{k-1}} f_k\end{aligned}$$

for any $f_1, \ldots, f_k \in \mathcal{C}(E)$ and $0 \le t_1 \le \cdots \le t_k$. This proves that X is a version of the Markov process with semigroup $(P_t)_{t\ge 0}$ started in the initial law μ.

(e)⇒(b): This is similar to the proof of the implication (c)⇒(b) in Proposition 6.7. Fix $\lambda > 0$. Then

$$(\lambda - G)^{-1}f(x) = \mathbb{E}^x\big[\int_0^\infty f(X_t)e^{-\lambda t}\mathrm{d}t\big] \qquad (x \in E,\ f \in \mathcal{C}(E)).$$

A similar formula holds for $(\lambda - G_n)^{-1}$. Since $w \mapsto \int_0^\infty f(w(t))e^{-\lambda t}\mathrm{d}t$ from $\mathcal{D}_E[0,\infty)$ to $\mathbb{R}$ is bounded and continuous, $\mathbb{P}^{(n),x_n} \Rightarrow \mathbb{P}^x$ implies that

$$(\lambda - G_n)^{-1}f(x_n) \underset{n\to\infty}{\longrightarrow} (\lambda - G)^{-1}f(x) \qquad (f \in \mathcal{C}(E),\ x_n, x \in E,\ x_n \to x).$$

This shows that $\|(\lambda - G_n)^{-1}f - (\lambda - G)^{-1}f\| \to 0$. Just as in the proof of Proposition 6.7, this implies that $\mathrm{ex}\!\lim_{n\to\infty} G_n = G$. ∎

6.2. **Proof of the main result (Theorem 4.2).** The proof of Theorem 6.3 has an important corollary.

Corollary 6.11. (Existence of limiting process) *Let E be compact and metrizable and let $(P^{(n)}_t)_{t\geq 0}$ and $(P_t)_{t\geq 0}$ be Feller semigroups on $\mathcal{C}(E)$ with generators G_n and G, respectively. Assume that $\mathrm{ex}\!\lim_{n\to\infty} G_n \supset G$ and that for each n there exists a Markov process $(\mathbb{P}^{(n),x})^{x\in E}$ with semigroup $(P^{(n)}_t)_{t\geq 0}$. Then there exists a Markov process $(\mathbb{P}^x)^{x\in E}$ with semigroup $(P_t)_{t\geq 0}$.*

Proof. By Proposition 2.12, there exists for each $x \in E$ an E-valued stochastic process $X^x = (X^x_t)_{t\geq 0}$ such that $X^x_0 = x$ and X^x satisfies the equivalent conditions (a)–(c) from Proposition 2.11. We need to show that X^x has a version with cadlag sample paths. Let $X^{(n),x}$ be $\mathcal{D}_E[0,\infty)$-valued random variables with laws $P^{(n),x}$. Our proof of Theorem 6.3 shows that the laws $\mathcal{L}(X^{(n),x})$ are tight and that each cluster point has the same finite dimensional distributions as X^x. It follows that the $X^{(n),x}$ converge weakly in path space and that their limit is a version of X^x with cadlag sample paths. ∎

We will use Corollary 6.11 to complete the proof of Theorem 4.2. All we need to do is to show that a general Feller semigroup can be approximated by 'easy' semigroups, for which we know that they correspond to a Markov process.

Proof of Theorem 4.2. Let E be compact and metrizable and let $(P_t)_{t\geq 0}$ be a Feller semigroup on E with generator G. For each $\varepsilon > 0$, let G_ε denote the Yosida approximation to G, defined in (3.91). We claim that G_ε is the generator of a jump process in the sense of Proposition 4.4 (and hence there exists a Markov process associated with the semigroup generated by G_ε). Indeed, by Lemma 3.21,

$$(1 - \varepsilon G)^{-1}f = \int_0^\infty P_t f\, \varepsilon^{-1}e^{-t/\varepsilon}\mathrm{d}t,$$

so if we define continuous probability kernels K_ε on E by

$$K_\varepsilon(x,A) := \int_0^\infty P_t(x,A)\,\varepsilon^{-1}e^{-t/\varepsilon}\mathrm{d}t \qquad (x \in E,\ A \in \mathcal{B}(E)),$$

then $G_\varepsilon f = \varepsilon^{-1}(K_\varepsilon f - f)$, which shows that G_ε is the generator of a jump process. Choose $\varepsilon_n \to 0$. Then formula (3.92) implies that $\mathrm{ex}\lim_{n\to\infty} G_{\varepsilon_n} \supset G$, which by Corollary 6.11 shows that there exists a Markov process $(\mathbb{P}^x)^{x\in E}$ with semigroup $(P_t)_{t\geq 0}$. ∎

7. Strong Markov property

Let $X := (X_t)_{t\geq 0}$, defined on $(\Omega, \mathcal{F}, \mathbb{P})$, be an E-valued Markov process with respect to a filtration $(\mathcal{F}_t)_{t\geq 0}$ such that X is $(\mathcal{F}_t)$-progressive (recall Definition 5.3).

Recall that the Markov property says that given the "present", the future is independent of the past. In this section we want to replace the deterministic notion of "present" by a stopping time.

Recall the intuitive description of $\mathcal{F}_t$ as the information known to an observer at time t. For an $(\mathcal{F}_t)$-stopping time τ, the σ-algebra should have the same intuitive meaning.

Definition 7.1 (σ-algebra generated by a stopping time)**.** *For an $\mathcal{F}_t$-stopping time τ, put*

$$\mathcal{F}_S := \big\{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t, \, \forall \, t \geq 0\big\}. \tag{7.1}$$

Similarly, $\mathcal{F}_{\tau+}$ is defined by replacing in (7.1) $\mathcal{F}_t$ by $\mathcal{F}_{t+}$.

Exercise 7.2. *Fix $t \geq 0$. Show that if $\mathbf{P}\{\tau = t\}$, then $\mathcal{F}_t = \mathcal{F}_\tau$ upto $\mathbf{P}$-zero sets.*

We immediately get the following useful properties.

Lemma 7.3. *Let σ and τ be $(\mathcal{F}_t)$-stopping times, let X be an $(\mathcal{F}_t)$-progressive E-valued process. Then the following hold:*

- (i) *$\mathcal{F}_\tau$ is a σ-algebra.*
- (ii) *$\tau \wedge \sigma$ is $\mathcal{F}_\tau$-measurable.*
- (iii) *If $\sigma \subseteq \tau$ then $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$.*
- (vi) *X_τ is $\mathcal{F}_\tau$-measurable.*

Proof. (i) Obviously, $\emptyset, \Omega \in \mathcal{F}_\tau$. If $A \in \mathcal{F}_\tau$ then

$$A^c \cap \{\tau \leq t\} = \{\tau \leq t\} \setminus (A \cap \{\tau \leq t\}) \in \mathcal{F}_t \tag{7.2}$$

for all $t \geq 0$, and therefore $A^c \in \mathcal{F}_\tau$. Similarly, if $(A_n)_{n\in\mathbb{N}} \in \mathcal{F}_\tau$ then

$$\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) \cap \{\tau \leq t\} = \bigcup_{n\in\mathbb{N}} (A_n \cap \{\tau \leq t\}) \in \mathcal{F}_t \tag{7.3}$$

for all $t \geq 0$, and hence $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}_\tau$.

(ii) For each $c \geq 0$ and $t \geq 0$,

$$\{\sigma \wedge \tau \leq c\} \cap \{\tau \leq t\} = \{\sigma \wedge \tau \leq c \wedge t\} \cap \{\tau \leq t\} \in \mathcal{F}_t. \tag{7.4}$$

Hence $\{\sigma \wedge \tau \leq c\} \in \mathcal{F}_\tau$ and $\sigma \wedge \tau$ is $\mathcal{F}_\tau$-measurable.

(iii) Let $A \in \mathcal{F}_\sigma$. Then for all $t \geq 0$,

$$A \cap \{\tau \leq t\} = A \cap \{\tau \leq t\} \cap \{\sigma \leq t\} \in \mathcal{F}_t. \tag{7.5}$$

Hence $A \in \mathcal{F}_\tau$.

(iv) Fix $t \geq 0$, and apply (ii) to $\tau := t$ to the effect that $\sigma \wedge t \in \mathcal{F}_\sigma$. $X_{\sigma\wedge t}$ is the composition of the $(\Omega, \mathcal{F}_t)$-$([0,t] \times \Omega, \mathcal{B}([0,t]) \times \mathcal{F}_t)$ measurable mapping which sends ω to $(\sigma(\omega) \wedge t, \omega)$ with the $([0,t] \times \Omega, \mathcal{B}([0,t]) \times \mathcal{F}_t)$-$(E, \mathcal{B}(E))$-measurable mapping which sends (s,ω) to $X_s(\omega)$. Notice that for the measurability of the second mapping one uses that X is $(\mathcal{F}_t)$-progressive. As a consequence $X_{\sigma\wedge t}$ is $\mathcal{F}_t$-measurable. Therefore, for all $t \geq 0$ and $\Gamma \in B(E)$,

$$\{X_\sigma \in \Gamma\} \cap \{\sigma \leq t\} = \{X_{\sigma\wedge t} \in \Gamma\} \cap \{\sigma \leq t\} \in \mathcal{F}_t. \tag{7.6}$$

Hence $\{X_\sigma \in \Gamma\} \in \mathcal{F}_\sigma$ for all $\Gamma \in B(E)$, or equivalently, X_σ is $\mathcal{F}_\sigma$-measurable. ∎

We next define the strong Markov property of a Markov process.

Definition 7.4 (Strong Markov property)**.** *Let $X := (X_t)_{t\geq 0}$, defined on $(\Omega, \mathcal{F}, \mathbb{P})$, be an E-valued Markov process with respect to a filtration $(\mathcal{F}_t)_{t\geq 0}$ such that X is $(\mathcal{F}_t)$-progressive (recall Definition 5.3). Suppose $P_t(x, A)$ is a transition function for X, and let τ be a $(\mathcal{F}_t)$-stopping time with $\tau < \infty$, almost surely.*

- *X is said to be* strong Markov at τ *if*
$$\mathbb{P}\big\{X_{\tau+t} \in A | \mathcal{G}_\tau\big\} = P_t(X_\tau, A) \tag{7.7}$$
for all $t \geq 0$ and $A \in \mathcal{B}(E)$.
- *X is said to be a* strong Markov process with respect to $(\mathcal{F}_t)$ *if X is strong Markov at τ for all $(\mathcal{F}_t)$-stopping times τ with $\tau < \infty$, almost surely.*

(Counter-)Example. A typical counterexample appears once we mix deterministic evolution with random evolution. Consider the $\mathbb{R}$-valued process which has the following dynamics.

- If $x \neq 0$, then X grows (deterministically) with unit speed,
- while if X reaches $x = 0$, then it spends their an exponential time with unit parameter.

In formulae, its semigroup is given as

$$T_t f(x) := \begin{cases} f(x+t), & \text{if } x \leq 0, x+t \leq 0 \\ e^{-(t+x)} f(0) + \int_{-x}^{t} \mathrm{d}u\, e^{-u} f(t-u), & \text{if } x \leq 0, x+t > 0, \\ f(x+t), & \text{if } x > 0. \end{cases} \tag{7.8}$$

It is easy to check that (7.8) indeed gives a Markovian semigroup. To see that the corresponding Markov process does not has the strong Markov property, put

$$\sigma := \inf\{t \geq 0 :\, X_t > 0\}, \tag{7.9}$$

and start the process in $x < 0$ (and thereby ensure that $\sigma < \infty$, a.s.). Since $\{\sigma \geq t\} = \bigcup_{s\in[0,t]\cap\mathbb{Q}}\{X_s \leq 0\} \in \mathcal{F}^X_t$ for all $t \geq 0$, σ is a $(\mathcal{F}^X_{t+})$-stopping time. Moreover, since X has right continuous paths, $X_{\sigma+} = X_\sigma = 0$. Hence $\mathbf{E}[X_{\sigma+t}|X_{\sigma+}] < t$, a.s., while $\mathbf{E}[X_{\sigma+t}|\mathcal{F}^X_{\sigma+}] = t$ which contradicts (7.7) (with $(\mathcal{F}_t)$ replaced by $(\mathcal{F}^X_{t+})$). ∎

The following result says that progressive Markov processes are strong Markov at discrete stopping times.

Proposition 7.5. *Let X be E-valued, $(\mathcal{F}_t)$-progressive, and $(\mathcal{F}_t)$-Markov, and let $P_t(x,A)$ be a transition function for X. Let τ be a discrete $(\mathcal{F}_t)$-stopping time with $\tau < \infty$, almost surely. Then X is strong Markov at τ.*

Proof. Let τ be a discrete $(\mathcal{F}_t)$-stopping time with $\tau < \infty$ a.s. We need to show that for all $B \in \mathcal{G}_\tau$,

$$\mathbb{E}\big[f(X_{t+\tau}); B\big] = \mathbb{E}\big[\int P_t(X_\tau, \mathrm{d}y) f(y); B\big]. \tag{7.10}$$

By assumption, there are $t_1, t_2, \ldots$ such that $\tau \in \{t_1, t_2, \ldots\}$. Furthermore, if $B \in \mathcal{G}_\tau$ then $B \cap \{\tau = t_k\} \in \mathcal{G}_{t_k}$ for all $k \in \mathbb{N}$, and hence for all $f \in B(E)$ and $t \geq 0$,

$$\begin{aligned}\mathbb{E}\big[f(X_{t+\tau}); B \cap \{\tau = t_k\}\big] &= \mathbb{E}\big[f(X_{t_k+\tau}); B \cap \{\tau = t_k\}\big] \\ &= \mathbb{E}\big[\int P_t(X_{t_i}, \mathrm{d}y) f(y); B \cap \{\tau = t_k\}\big] \\ &= \mathbb{E}\big[\int P_t(X_\tau, \mathrm{d}y) f(y); B \cap \{\tau = t_k\}\big].\end{aligned} \tag{7.11}$$

Summing over all k yields (7.10). ∎

The next result states that each stopping time is the limit of a decreasing sequence of discrete stopping times.

Lemma 7.6. *Let $(\mathcal{F}_t)_{t\geq 0}$ be a filtration, and τ be a $(\mathcal{F}_{t+})$-stopping time. Then there exists a decreasing sequence $(\tau_n)_{n\in\mathbb{N}}$ of discrete $(\mathcal{F}_t)$-stopping times such that $\tau = \lim_{n\to\infty} \tau_n$.*

Proof. Choose for each $n \in \mathbb{N}$, $0 = t_0^n < t_1^n < \cdots$ such that $\lim_{n\to\infty} t_n = \infty$ and $\lim_{n\to\infty} \sup_{k\in\mathbb{N}} (t^n_{k+1} - t^n_k) = 0$. Then put

$$\tau_n := \begin{cases} t^n_{k+1}, & \text{if } t^n_k \leq t < t^n_{k+1}, \\ \infty, & \tau = \infty. \end{cases} \tag{7.12}$$

Obviously, $\lim_{n\to\infty} \tau_n = \tau$, while $(\tau_n)_{n\in\mathbb{N}}$ is decreasing if $(t^{n+1}_k)_{k\in\mathbb{N}}$ is finer than $(\tau^n_k)_{k\in\mathbb{N}}$. ∎

We will exploit the latter to state that Feller semigroups define strong Markov processes.

Theorem 7.7 (Feller semigroups give strong Markov processes)**.** *Let E be locally compact and separable, and let $(P_t)_{t\geq 0}$ be a Feller semigroup on $C_b(E)$. Then for each probability law ν on E there exists a Markov process X corresponding to $(P_t)_{t\geq 0}$ with initial law ν and sample paths in $\mathcal{D}_E([0,\infty))$ which is strong Markov with respect to the filtration $\mathcal{F}_t := \mathcal{F}^X_{t+}$.*

Proof. We already know from Theorem 4.2 (combined with the considerations for locally compact state spaces discussed in Subsection 4.4) that under the above assumptions there is a Markov process X with (cadlag paths) corresponding to $(P_t)_{t\geq 0}$ with initial law ν. It remains to verify the strong Markov property..

Assume for the moment that τ is a discrete $(\mathcal{F}_t)$-stopping time with $\tau < \infty$, i.e., τ can be written as

$$\tau := \sum_{n\geq 1} t_n 1\{\tau = t_n\} \tag{7.13}$$

for suitable $(t_n)_{n\in\mathbb{N}}$ in $[0,\infty)$. Let $A \in \mathcal{F}_t$, $s > 0$, and $f \in \hat{C}(E)$. Then $\{\tau = t_n\} \in \mathcal{F}_{t_n+\varepsilon}$ for all $\varepsilon > 0$ and $n \in \mathbb{N}$, so

$$\begin{aligned} \int_{A\cap\{\tau=t_n\}} \mathrm{d}\mathbb{P}\, f(X_{\tau+s}) &= \int_{A\cap\{\tau=t_n\}} \mathrm{d}\mathbb{P}\, f(X_{t_n+s}) \\ &= \int_{A\cap\{\tau=t_n\}} P_{s-\varepsilon} f(X_{t_n+\varepsilon}) \end{aligned} \tag{7.14}$$

for all $\varepsilon \in [0,s]$. Since $(P_t)_{t\geq 0}$ is strongly continuous, $T_s f$ is continuous on E for all $s \geq 0$. Moreover, since X has right continuous sample paths, we can let $\varepsilon \downarrow 0$ in (7.14) to the effect that it holds with for $\varepsilon = 0$ as well. This gives

$$\mathbf{E}\big[X_{\tau+s}\big|\mathcal{F}_\tau\big] = P_s f(X_\tau) \tag{7.15}$$

for discrete τ.

If τ is an arbitrary $(\mathcal{F}_t)$-stopping time, with $\tau < \infty$, a.s., we know from Lemma 7.6 that τ can be written as the decreasing limit of discrete stopping times $(\tau_n)_{n\in\mathbb{N}}$. It follows then from continuity of $P_s f$ and the right continuous sample paths that (7.15) holds. ∎

References

[Bil86] P. Billingsley. *Probability and Measure*. John Wiley & Sons, New York, 1986.

[Bou58] N. Bourbaki. *Éléments de Mathématique, 2nd ed., Book 3, Chap. 9*. Hermann & Cie, Paris, 1958.

[Bou64] N. Bourbaki. *Éléments de Mathématique, 2nd ed., Book 3, Fascicule de Résultats*. Hermann & Cie, Paris, 1964.

[Cho69] G. Choquet. *Lectures on Analysis, Vol. 1*. W.A. Benjamin, New York, 1969.

[Chu74] K.L. Chung. *A Course in Probability Theory, 2nd ed.* Academic Press, Orlando, 1974.

[Dan19] P.J. Daniell. Integrals in an infinite number of dimensions. *Annals of Mathematics*, 20:281–288, 1919.

[EK86] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes: Characterization and convergence*. John Wiley and Sons, 1986.

[Fri64] A. Friedman. *Partial Differential Equations of Parabolic Type*. Prentice-Hall, Englewood Cliffs, 1964.

[Kel55] J.L. Kelley. *General Topology*. Van Nostrand, New York, 1955.

[Kol33] A.N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitstheorie*, volume 2(3) of *Ergeb. Math.* Springer, Berlin, 1933.

[Kol56] A.N. Kolmogorov. On skorohod convergence. *Theory Probab. Appl.*, 1:213–222, 1956.

[KS88] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, 1988.

[KS91] I. Karatzas and E.S. Shreve. *Brownian Motion and Stochastic Calculus, 2nd ed.* Springer, New York, 1991.

[RS80] Michael Reed and Barry Simon. *Functional Analysis*, volume I. Academic Press, Inc., 1980.

[Sch73] L. Schwartz. *Radon Measures on Arbitrary Topological Spaces and Cylindical Measures*. Tata Institute, Oxford University Press, London, 1973.

Jan Swart, Mathematisches Institut, Universität Erlangen–Nürnberg, Bismarckstrasse $1\frac{1}{2}$, 91054 Erlangen, GERMANY

E-mail address: swart@mi.uni-erlangen.de

Anita Winter, Mathematisches Institut, Universität Erlangen–Nürnberg, Bismarckstrasse $1\frac{1}{2}$, 91054 Erlangen, GERMANY

E-mail address: winter@mi.uni-erlangen.de