Ergodic theory for Stochastic PDEs

July 10, 2008

M. Hairer

Mathematics Institute, The University of Warwick
Email: M.Hairer@Warwick.co.uk

Contents

1 Introduction

These lecture notes cover the material presented at the LMS-EPSRC Short Course on Stochastic Partial Differential Equations held at Imperial College London in July 2008. They extend the material presented at the '12$^{\text{th}}$ National Summer School in Mathematics for Graduate Students' that took place at the University of Wuhan in July 2007.

They are by no means meant to be exhaustive and a solid background in probability theory and PDE theory is required to follow them. The structure of these notes is as follows. In Sections 2 to 4, we introduce the concepts of a time homogeneous Markov process, its set of invariant measures We then proceed in Section 5 to show a very general structure theorem that gives us a feeling of what the set of all invariant probability measures for a given Markov process can look like. Sections 6 and 7 are then devoted to the presentation of a few criteria that yield existence and uniqueness of the invariant measure(s). The main philosophy that we try to convey there is that:

- In order to have *existence* of an invariant measure, the Markov process should satisfy some *compactness* property, together with some *regularity*.

- In order to have *uniqueness* of the invariant measure, the Markov process should satisfy some *irreducibility* property, together with some *regularity*.

These two claims illustrate that the interplay between measure-theoretic notions (existence and uniqueness of an invariant measure) and topological concepts (compactness, irreducibility) is a fundamental aspect of the ergodic theory of Markov processes.

Section 8 is devoted to an explanation (rather than a complete proof) of Hörmander's famous 'sums of squares' theorem and how it can be used to check whether a diffusion has transition probabilities that are continuous in the total variation distance, thus satisfying the regularity condition required for showing the uniqueness of an invariant measure. In Section 9, we then show in which ways the proof of Hörmander's theorem breaks down for infinite-dimensional diffusions.

For situation where the forcing noise is sufficiently rough, we see however in Section 10 that not all is lost. In particular, we give a proof of the Bismut-Elworthy-Li formula that allows to show the strong Feller property for a rather large class of semilinear parabolic stochastic PDEs. In cases where the noise is very weak, this has no chance of being applicable. It therefore motivates the introduction in Section 11 of the asymptotic strong Feller property, which is the weakest type of regularity condition so far, still ensuring uniqueness of the invariant measure when combined with topological irreducibility. We also show that the asymptotic strong Feller property is satisfied by a class of stochastic reaction-diffusion equations, provided that the noise acts on sufficiently many Fourier modes with small wave number.

2 Definition of a Markov process

Loosely speaking, a Markov process is a stochastic process such that its future and its past are conditionally independent, given its present. The precise mathematical formulation of this sentence is given by:

Definition 2.1 A stochastic process $\{X_t\}_{t\in\mathbf{T}}$ taking values in a state space $\mathcal{X}$ is called a *Markov process* if, for any $N > 0$, any ordered collection $t_{-N} < \ldots < t_0 < \ldots < t_N$ of times, and any two functions $f, g\colon \mathcal{X}^N \to \mathbf{R}$, the equality

$$\begin{aligned}\mathbf{E}(f(X_{t_1},\ldots,X_{t_N})\, g(X_{t_{-1}},\ldots,X_{t_{-N}}) \mid X_{t_0}) \\ = \mathbf{E}(f(X_{t_1},\ldots,X_{t_N}) \mid X_{t_0})\mathbf{E}(g(X_{t_{-1}},\ldots,X_{t_{-N}}) \mid X_{t_0})\end{aligned}$$

holds almost surely.

In other words, the 'future' $(X_{t_1},\ldots,X_{t_N})$ and the 'past' $(X_{t_{-1}},\ldots,X_{t_{-N}})$ are independent, given the 'present' X_{t_0}.

In most situations, Markov processes are constructed from their *transition probabilities*, that is the specifications that give the probability $\mathcal{P}_{s,t}(x, A)$ that the process is in the set A at time t, given that it was located at the point x at time $s < t$. This motivates the following definitions:

Definition 2.2 A *Markov transition kernel* over a Polish space $\mathcal{X}$ is a map $\mathcal{P}\colon \mathcal{X} \times \mathcal{B}(\mathcal{X}) \to \mathbf{R}_+$ such that:

- For every set $A \in \mathcal{B}(\mathcal{X})$, the map $x \mapsto \mathcal{P}(x, A)$ is measurable.
- For every $x \in \mathcal{X}$, the map $A \mapsto \mathcal{P}(x, A)$ is a probability measure.

Definition 2.3 A *Markov operator* over a Polish space $\mathcal{X}$ is a bounded linear operator $\mathcal{P}\colon \mathcal{B}_b(\mathcal{X}) \to \mathcal{B}_b(\mathcal{X})$ such that:

- $\mathcal{P}1 = 1$.
- $\mathcal{P}\varphi$ is positive whenever φ is positive.
- If a sequence $\{\varphi_n\} \subset \mathcal{B}_b(\mathcal{X})$ converges pointwise to an element $\varphi \in \mathcal{B}_b(\mathcal{X})$, then $\mathcal{P}\varphi_n$ converges pointwise to $\mathcal{P}\varphi$.

It is possible to check that the two definitions actually define one and the same object in the following sense:

Lemma 2.4 *Given a Polish space $\mathcal{X}$, there is a one-to-one correspondence between Markov transition kernels over $\mathcal{X}$ and Markov operators over $\mathcal{X}$ given by $\mathcal{P}(x, A) = (\mathcal{P}\mathbf{1}_A)(x)$.*

Proof. It is straightforward to check that a Markov transition kernel defines a Markov operator, the past property being an immediate consequence of Lebegue's dominated convergence theorem.

Conversely, if we define a candidate for a Markov transition kernel by $\mathcal{P}(x, A) = (\mathcal{P}\mathbf{1}_A)(x)$, we only need to check that $A \mapsto \mathcal{P}(x, A)$ is a probability measure for every x. The only non-trivial assertion is countable additivity, which however follows at once from the last property of a Markov operator. □

We will therefore use the terminologies 'Markov transition kernel' and 'Markov operator' interchangeably. We will also use the symbol $\mathcal{P}$ for both a Markov operator acting on bounded measurable functions and the corresponding transition probabilities. In order to streamline the notation, we will make a further abuse of notations and use the same symbol $\mathcal{P}$ for the operator acting on signed measures by

$$(\mathcal{P}\mu)(A) = \int_{\mathcal{X}} \mathcal{P}(x, A)\mu(dx) \; .$$

It will hopefully always be clear from the context which is the object that we are currently talking about.

When considering Markov processes with continuous time, it is natural to consider a family of Markov operators indexed by time. We call such a family a *Markov semi-group*, provided that it satisfies the relation $\mathcal{P}_{t+s} = \mathcal{P}_t \circ \mathcal{P}_s$, for any $s, t > 0$. We call a Markov process X a *time-homogeneous Markov process* with semigroup $\{\mathcal{P}_t\}$ if, for any two times $s < t$, we have

$$\mathbf{P}(X_t \in A \,|\, X_s) = \mathcal{P}_{t-s}(X_s, A) \; ,$$

almost surely.

A probability measure μ on $\mathcal{X}$ is *invariant* for the Markov operator $\mathcal{P}$ if the equality

$$\int_{\mathcal{X}} (\mathcal{P}\varphi)(x)\,\mu(dx) = \int_{\mathcal{X}} \varphi(x)\,\mu(dx)$$

holds for every function $\varphi \in \mathcal{B}_b(\mathcal{X})$. In other words, one has $\mathcal{P}_t\mu = \mu$ for every positive time t.

2.1 Examples

Example 2.5 (Finite state space) Take $\mathcal{X} = \{1, \ldots, n\}$ for some $n \in \mathbf{N}$. In this case, both spaces $\mathcal{B}_b(\mathcal{X})$ and $\mathcal{M}(\mathcal{X})$ are canonically isomorphic to $\mathbf{R}^n$ via the identifications $\varphi_i = \varphi(i)$ for functions and $\mu_i = \mu(\{i\})$ for measures. The pairing between functions and measures then corresponds to the Euclidean scalar product on $\mathbf{R}^n$.

A Markov operator over $\mathcal{X}$ acting on *measures* is therefore given by an $n \times n$ matrix P with the properties that $P_{ij} \geq 0$ for any pair (i, j) and that $\sum_j P_{ij} = 1$ for every i. The number P_{ij} represents the probability of jumping to the state i, given that the current state is j. The corresponding operator acting on *functions* is given by the transpose matrix P^T, since $\langle P\mu, \varphi\rangle = \langle \mu, P^T\varphi\rangle$.

Example 2.6 (i.i.d. random variables) Take an arbitrary state space $\mathcal{X}$ and a probability measure $\mu \in \mathcal{M}_1(\mathcal{X})$. A sequence $\{X_i\}$ of independent, identically distributed, $\mathcal{X}$-valued random variables with law μ is a Markov process. The corresponding Markov operator is given by $(\mathcal{P}\varphi)(x) = \int_{\mathcal{X}} \varphi(y)\,\mu(dy)$, which is always a constant function.

Example 2.7 (Random walk) Let $\{\xi_n\}$ be a sequence of i.i.d. real-valued random variables with law μ and define $X_0 = 0$, $X_n = \sum_{k=1}^n \xi_k$. The process X is called a *random walk* with increments ξ. The most prominent example is the *simple random walk* which coresponds to the case where the ξ_n are Bernoulli random variables taking the values ± 1 with equal probabilities. More information about the behaviour of random walks can be found in [Spi76].

Example 2.8 (Brownian motion) This is probably the most studied Markov process in continuous time, see for example the monograph [RY99]. Its state space is $\mathbf{R}$ and, for a given $\sigma > 0$, its Markov semigroup is given by

$$(\mathcal{P}_t\varphi)(x) = \frac{1}{\sigma\sqrt{2\pi t}} \int_{\mathbf{R}} e^{-\frac{(x-y)^2}{2\sigma^2 t}} \varphi(y)\,dy \;. \tag{2.1}$$

In terms of transition probabilities, we can also write this as $\mathcal{P}_t(x, \cdot\,) = \mathcal{N}(x, \sigma^2 t)$, the Gaussian law with mean x and variance $\sigma^2 t$. The Brownian motion with variance $\sigma^2 = 1$ is also called the *standard Wiener process*. Brownian motion is named after 19th century botanist Robert Brown who studied the motion of grains of pollen in suspension in a fluid [Bro28]. It can be obtained as a scaling limit

$$B_t \sim \lim_{N\to\infty} \frac{1}{\sqrt{N}} X_{[Nt]} \;,$$

where X_n denotes the random walk from the previous example, provided that $\sigma^2 = \int_{\mathbf{R}} x^2 \mu(dx) < \infty$. The modeling idea is that the grain of pollen is constantly bombarded by water molecules which push it into a random direction[1].

It is interesting to note that the kernel appearing in the right hand side of (2.1) is the fundamental solution of the heat equation, which implies that $\psi(x, t) = (\mathcal{P}_t\varphi)(x)$ solves the partial differential equation

$$\frac{\partial\psi}{\partial t} = \frac{1}{2}\frac{\partial^2\psi}{\partial x^2} \;, \quad \psi(x, 0) = \varphi(x) \;.$$

[1] A more realistic model has a parameter $\gamma > 0$ taking into account the inertia of the grain of pollen is given by the process $B_\gamma(t) = \gamma\int_0^t e^{-\gamma(t-s)}B(s)\,ds$, where B is the *mathematical Brownian motion* that we just described. The process B_γ is sometimes referred to as the *physical Brownian motion* and converges to the mathematical Brownian motion in the limit $\gamma \to \infty$.

This link between Brownian motion and the heat equation was discovered by Einstein in [Ein05] and is still permeates much of stochastic analysis. It allows to give probabilistic proof of analytical questions and vice-versa.

Example 2.9 (Differential equations) Markov processes do not have to be random. A perfectly valid example of a continuous-time Markov process is the solution of the ordinary differential equation on $\mathbf{R}^n$

$$\frac{dx}{dt} = f(x)\;, \qquad x(0) = x_0\;, \tag{2.2}$$

where $f\colon \mathbf{R}^n \to \mathbf{R}^n$. If we assume for example that f is globally Lipschitz continuous (that is $|f(x) - f(y)| \le L|x - y|$ for some constant $L > 0$), then (2.2) has a unique solution for all times and for all initial conditions. Denote by $\Phi_t\colon \mathbf{R}^n \to \mathbf{R}^n$ the solution map for (2.2), that is $\Phi_t(x_0)$ is the solution of (2.2) at time t. Obviously, uniqueness of the solutions implies that $\Phi_s \circ \Phi_t = \Phi_{s+t}$, so that the family of operators $\mathcal{P}_t$ defined by $\mathcal{P}_t\varphi = \varphi \circ \Phi_t$ is indeed a Markov semigroup. The corresponding transition probabilities are given by Dirac measures: $\mathcal{P}_t(x, \cdot\,) = \delta_{\Phi_t(x)}$.

Example 2.10 (Autoregressive process) Let $\{\xi_n\}$ be a sequence of i.i.d. random variables on $\mathbf{R}^d$ with law μ and let $\alpha \in \mathbf{R}$. Given an $\mathbf{R}^d$-valued random variable X_0 independent of $\{\xi_n\}$, we construct a sequence $\{X_n\}$ by the recursion formula:

$$X_{n+1} = \alpha X_n + \xi_n\;.$$

This is a Markov process in discrete time and the corresponding Markov operator is given by

$$(\mathcal{P}\varphi)(x) = \int_{\mathbf{R}^d} \varphi(\alpha x + y)\,\mu(dy)\;.$$

Actually, we can look at more general recursions of the form

$$X_{n+1} = F(X_n, \xi_n)\;,$$

which still yields a Markov process with Markov operator

$$(\mathcal{P}\varphi)(x) = \int_{\mathbf{R}^d} \varphi(F(x,y))\,\mu(dy)\;.$$

The autoregressive process and the random walk are two examples of Markov processes with this structure.

Example 2.11 (Stochastic differential equations) Let $f\colon \mathbf{R}^d \to \mathbf{R}^d$ and $\sigma\colon \mathbf{R}^d \to \mathbf{R}^{d\times m}$ be two functions that are globally Lipschitz continuous and consider the stochastic differential equation

$$dx = f(x)\,dt + \sigma(x)\,dw\;, \qquad x(0) = x_0\;, \tag{2.3}$$

where w is a standard Wiener process. For a detailed discussion of the meaning of (2.3), we refer to [Øks03a]. The Markov semigroup $\mathcal{P}_t$ associated to the solutions of an SDE is, as in the case of Brownian motion, the solution of a partial differential equation:

$$\partial_t \mathcal{P}_t\varphi = \mathcal{L}\mathcal{P}_t\varphi\;,$$

where the differential operator $\mathcal{L}$ is given by

$$(\mathcal{L}\varphi)(x) = \sum_i f_i(x)\frac{\partial\varphi}{\partial x_i} + \sum_{i,j,k} \sigma_{ik}(x)\sigma_{jk}(x)\frac{\partial^2\varphi}{\partial x_i \partial x_j}\;.$$

3 Dynamical systems

In this section, we give a short survey of the basic notions and results from the theory of dynamical systems. For a much more exhaustive overview, we refer to the excellent monographs [Sin94, Wal82].

Definition 3.1 Let E be a Polish space and let $\mathbf{T}$ be either $\mathbf{N}$, $\mathbf{Z}$, $\mathbf{R}_+$, or $\mathbf{R}$. A dynamical system on E is a collection $\{\Theta_t\}_{t\in\mathbf{T}}$ of maps $\Theta_t \colon E \to E$ such that $\Theta_t \circ \Theta_s = \Theta_{s+t}$ for every $s, t \in \mathbf{T}$ and such that the map $(t, x) \mapsto \Theta_t(x)$ is jointly measurable. It is called *continuous* if each of the maps Θ_t is continuous.

Remark 3.2 The reason why we use the letter E to denote a Polish space in this section instead of $\mathcal{X}$ is that in the application we are interested in, we will take $E = \mathcal{X}^{\mathbf{Z}}$ and Θ_t the shift map.

Given a dynamical system, a natural object of interest is the set of probability measures that are invariant under the action of Θ_t. Denoting by $\Theta_t^*\mu$ the push-forward of μ under the map Θ_t, we define the set of *invariant measures* for $\{\Theta_t\}$ by

$$\mathcal{J}(\Theta) = \{\mu \in \mathscr{M}_1(E) \,:\, \Theta_t^*\mu = \mu \textit{ for all } t \in \mathbf{T}\} \;.$$

We can also define in a similar way the σ-algebra of all *invariant subsets* of E:

$$\mathcal{I} = \{A \subset \mathcal{X} \,:\, A \text{ Borel and } \Theta_t^{-1}(A) = A \text{ for every } t \in \mathbf{T}\} \;.$$

(This set depends obviously also on the choice of dynamical system, but we will omit this in order not to get our notations overcrowded.)

One of the most striking results of the theory of dynamical systems is that some kind of 'law of large numbers' can be shown to hold in great generality:

Theorem 3.3 (Birkhoff's Ergodic Theorem) *Let $(\Theta_t)_{t\in\mathbf{T}}$ be a measurable dynamical system over a Polish space E. Fix an invariant measure $\mu \in \mathcal{J}(\Theta)$ and let $f \in L^1(E, \mu)$. Then,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(\Theta^n(x)) = \mathbf{E}_\mu(f \,|\, \mathcal{I})(x)$$

for μ-almost every $x \in E$. In particular, the expression on the left converges to a limit for μ-almost every starting point x.

This theorem suggests strongly that an important class of invariant measures is given by those under which the invariant σ-algebra is trivial.

Definition 3.4 An invariant measure μ for a dynamical system $\{\Theta_t\}$ is *ergodic* if $\mu(A) \in \{0, 1\}$ for every $A \in \mathcal{I}$.

Corollary 3.5 *With the notations of Theorem 3.3, if μ is ergodic, then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(\Theta^n(x)) = \mathbf{E}_\mu f$$

μ - almost surely.

Proof of the corollary. By definition, the function $\bar{f} \equiv \mathbf{E}(f \,|\, \mathcal{I})$ is $\mathcal{I}$-measurable. Define the sets $A_+ = \{x \in E \,|\, \bar{f}(x) > \mathbf{E}\bar{f}\}$, $A_- = \{x \in E \,|\, \bar{f}(x) < \mathbf{E}\bar{f}\}$, and $A_0 = \{x \in E \,|\, \bar{f}(x) = \mathbf{E}\bar{f}\}$. All three sets belong to $\mathcal{I}$ and they form a partition of E. Therefore, by ergodicity of μ, exactly one of them has measure 1 and the other two must have measure 0. If it was A_+, one would have $\mathbf{E}\bar{f} = \int_{A_+} f(x)\,\mu(dx) > \mathbf{E}\bar{f}$, which is a contradiction and similarly for A_-. This implies that $\mu(A_0) = 1$, and so $\mu(\bar{f} = \mathbf{E}\bar{f}) = 1$ as requested. □

Before we trun to the proof of Theorem 3.3, we establish the following important result:

Theorem 3.6 (Maximal Ergodic Theorem) *With the notations of Theorem 3.3, define*

$$S_N(x) = \sum_{n=0}^{N-1} f(\theta^n x)\;, \quad M_N(x) = \max\{S_0(x), S_1(x), \ldots, S_N(x)\}\;,$$

with the convention $S_0 = 0$. Then, $\int_{\{M_N>0\}} f(x)\,\mu(dx) \geq 0$ for every $N \geq 1$.

Proof. For every $N \geq k \geq 0$ and every $x \in E$, one has $M_N(\Theta(x)) \geq S_k(\Theta(x))$ by definition, and so $f(x) + M_N(\Theta(x)) \geq f(x) + S_k(\Theta(x)) = S_{k+1}(x)$. Therefore

$$f(x) \geq \max\{S_1(x), S_2(x), \ldots, S_N(x)\} - M_N(\Theta(x))\;.$$

Furthermore, $\max\{S_1(x), \ldots, S_N(x)\} = M_N(x)$ on the set $\{M_N > 0\}$, so that

$$\begin{aligned}\int_{\{M_N>0\}} f(x)\,\mu(dx) &\geq \int_{\{M_N>0\}} (M_N(x) - M_N(\Theta(x)))\,\mu(dx)\\ &\geq \mathbf{E}M_N - \int_{A_N} M_N(x)\,\mu(dx)\;,\end{aligned}$$

where $A_N = \{\Theta(x) \,|\, M_N(x) > 0\}$. The first inequality follows from the fact that $M_N \geq 0$ and the second inequality follows from the fact that Θ is measure-preserving. Since $M_N \geq 0$, $\int_A M_N(x)\,\mu(dx) \leq \mathbf{E}M_N$ for every set A, so that the expression above is greater or equal to 0, which is the required result. □

We can now turn to the

Proof of Birkhoff's Ergodic Theorem. Replacing f by $f - \mathbf{E}(f \,|\, \mathcal{I})$, we can assume without loss of generality that $\mathbf{E}(f \,|\, \mathcal{I}) = 0$. Define $\bar{\eta} = \limsup_{n\to\infty} S_n/n$ and $\underline{\eta} = \liminf_{n\to\infty} S_n/n$. It is sufficient to show that $\bar{\eta} \leq 0$ almost surely, since this implies (by considering $-f$ instead of f) that $\underline{\eta} \geq 0$ and so $\bar{\eta} = \underline{\eta} = 0$.

It is clear that $\bar{\eta}(\Theta(x)) = \bar{\eta}(x)$ for every x, so that, for every $\varepsilon > 0$, one has $A^\varepsilon = \{\bar{\eta}(x) > \varepsilon\} \in \mathcal{I}$. Define

$$f^\varepsilon(x) = (f(x) - \varepsilon)\,\mathbf{1}_{A^\varepsilon}(x)\;,$$

and define S_N^ε and M_N^ε accordingly. It follows from The maximal ergodic theorem, Theorem 3.6, that $\int_{\{M_N^\varepsilon>0\}} f^\varepsilon(x)\,\mu(dx) \geq 0$ for every $N \geq 1$. Note that with these definitions we have that

$$\frac{S_N^\varepsilon(x)}{N} = \begin{cases} 0 & \text{if } \bar{\eta}(x) \leq \varepsilon \\ \frac{S_N(x)}{N} - \varepsilon & \text{otherwise.} \end{cases} \tag{3.1}$$

The sequence of sets $\{M_N^\varepsilon > 0\}$ increases to the set $B^\varepsilon \equiv \{\sup_N S_N^\varepsilon > 0\} = \{\sup_N \frac{S_N^\varepsilon}{N} > 0\}$. It follows from (3.1) that

$$B^\varepsilon = \{\bar\eta > \varepsilon\} \cap \Big\{\sup_N \frac{S_N}{N} > \varepsilon\Big\} = \{\bar\eta > \varepsilon\} = A^\varepsilon \;.$$

Since $\mathbf{E}|f^\varepsilon| \le \mathbf{E}|f| + \varepsilon < \infty$, the dominated convergence theorem implies that

$$\lim_{N\to\infty} \int_{\{M_N^\varepsilon > 0\}} f^\varepsilon(x)\,\mu(dx) = \int_{A^\varepsilon} f^\varepsilon(x)\,\mu(dx) \ge 0 \;,$$

and so

$$\begin{aligned} 0 &\le \int_{A^\varepsilon} f^\varepsilon(x)\,\mu(dx) = \int_{A^\varepsilon} (f(x) - \varepsilon)\,\mu(dx) = \int_{A^\varepsilon} f(x)\,\mu(dx) - \varepsilon\mu(A^\varepsilon) \\ &= \int_{A^\varepsilon} \mathbf{E}(f(x)\,|\,\mathcal{I})\,\mu(dx) - \varepsilon\mu(A^\varepsilon) = -\varepsilon\mu(A^\varepsilon) \;, \end{aligned}$$

where we used the fact that $A^\varepsilon \in \mathcal{I}$ to go from the first to the second line. Therefore, one must have $\mu(A^\varepsilon) = 0$ for every $\varepsilon > 0$, which implies that $\bar\eta \le 0$ almost surely. □

4 Stationary Markov processes

In this section, we show that, to every invariant measure for a given Markov semigroup, one can associate a dynamical system in a canonical way. This allows to take over all the definitions from the previous section and to apply them to the study of stationary Markov processes.

Given a Markov semigroup $\mathcal{P}_t$ over a Polish space $\mathcal{X}$ and an invariant probability measure μ for $\mathcal{P}_t$, we associate to it a probability measure $\mathbf{P}_\mu$ on $\mathcal{X}^{\mathbf{R}}$ in the following way. For any bounded measurable function $\varphi\colon \mathcal{X}^{\mathbf{R}} \to \mathbf{R}$ such that there exists a function $\tilde\varphi\colon \mathcal{X}^n \to \mathbf{R}$ and an n-tuple of times $t_1 < \ldots < t_n$, we write

$$(\mathbf{P}_\mu)\varphi = \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} \tilde\varphi(x_1,\ldots,x_n)\mathcal{P}_{t_n - t_{n-1}}(x_{n-1}, dx_n)\ldots\mathcal{P}_{t_2-t_1}(x_1,dx_2)\mu(x_1) \;. \tag{4.1}$$

It is straightforward to check that this set of specifications is consistent and that therefore, by Kolmogorov's extension theorem [RY99], there exists a unique measure $\mathbf{P}_\mu$ on $\mathcal{X}^{\mathbf{R}}$ such that (4.1) holds.

Since μ is invariant, the measure $\mathbf{P}_\mu$ is stationary, that is $\theta_t^*\mathbf{P}_\mu = \mathbf{P}_\mu$ for every $t \in \mathbf{R}$, where the shift map $\theta_t\colon \mathcal{X}^{\mathbf{R}} \to \mathcal{X}^{\mathbf{R}}$ is defined by

$$(\theta_t x)(s) = x(t+s) \;.$$

Therefore, the measure $\mathbf{P}_\mu$ is an invariant measure for the dynamical system θ_t over $\mathcal{X}^{\mathbf{R}}$. This allows to carry over in a natural way the following notions from the theory of dynamical systems:

Definition 4.1 An invariant measure μ for a Markov semigroup $\mathcal{P}_t$ is *ergodic* if $\mathbf{P}_\mu$ is ergodic for the shift map θ_t.

5 Main structure theorem

Recall the construction from Section 4 that associates to every invariant probability measure μ of a given transition operator a measure $\mathbf{P}_\mu$ on the space $\mathcal{X}^{\mathbf{Z}}$ of $\mathcal{X}$-valued processes. We furthermore defined the shifts θ_n on $\mathcal{X}^{\mathbf{Z}}$ by

$$(\theta_n x)(m) = x(n+m) ,$$

and we write $\theta = \theta_1$. By the definition of stationarity, one has:

In this section, we will often approximate sets belonging to one particular σ-algebra by sets belonging to another σ-algebra. In this context, it is convenient to introduce a notation for the *completion* of a σ-algebra under a given probability measure. Assuming that it is clear from the context what the probability measure $\mathbf{P}$ is, we define the completion $\bar{\mathcal{F}}$ of a σ-algebra $\mathcal{F}$ to be the smallest σ-algebra containing $\mathcal{F}$ with the additional property that if $A \in \bar{\mathcal{F}}$ with $\mathbf{P}(A) = 0$ and $B \subset A$ is any subset of A, then $B \in \bar{\mathcal{F}}$.

The main result of this section is the following characterisation of the set of all invariant measure for a given Markov semigroup:

Theorem 5.1 *The set $\mathcal{J}(\mathcal{P})$ of all invariant probability measures for a Markov semigroup $\{\mathcal{P}_t\}$ is convex and $\mu \in \mathcal{J}(\mathcal{P})$ is ergodic if and only if it is an extremal point of $\mathcal{J}(\mathcal{P})$ (that is it cannot be decomposed as $\mu = t\mu_1 + (1-t)\mu_2$ with $t \in (0,1)$ and $\mu_i \in \mathcal{J}(\mathcal{P})$).*

Furthermore, any two ergodic invariant probability measures are either identical or mutually singular and, for every invariant measure $\mu \in \mathcal{J}(\mathcal{P})$ there exists a probability measure ϱ_μ on the set $\mathcal{E}(\mathcal{P})$ of ergodic invariant measures for $\mathcal{P}$ such that $\mu(A) = \int_{\mathcal{J}(\mathcal{P})} \nu(A)\, \varrho_\mu(d\nu)$.

Before we turn to the proof of Theorem 5.1, we prove the following very important preliminary result:

Proposition 5.2 *Let $\mathbf{P}$ be the law of a stationary Markov process on $\mathcal{X}^{\mathbf{Z}}$. Then, the σ-algebra $\mathcal{I}$ of all subsets invariant under θ is contained in the completion $\bar{\mathcal{F}}_0^0$ of $\mathcal{F}_0^0$ under $\mathbf{P}$.*

Proof. We introduce the following notation. For any subset $A \subset \mathcal{X}^{\mathbf{Z}}$ and any subset $I \subset \mathbf{Z}$, denote by $\Pi_I A \subset \mathcal{X}^{\mathbf{Z}}$ the set[2]

$$\Pi_I A = \{y \in \mathcal{X}^{\mathbf{Z}} \,|\, \exists x \in A \text{ with } x_k = y_k \;\forall k \in I\} .$$

Note that one has $A \subset \Pi_I A$ for any I. Furthermore, we have $A = \bigcap_{n\ge 0} \Pi_{[-n,n]} A$, so that $\mathbf{P}(\Pi_{[-n,n]} A \setminus A) \to 0$ as $n \to \infty$. Note also that if $A = \Pi_{[k,\ell]} A$, then $A \in \mathcal{F}_k^\ell$.

Fix now $k > 0$. Since $A \in \mathcal{I}$, one has

$$x \in \Pi_{[-n,n]} A \quad \Leftrightarrow \quad \theta^{-(k+n)} x \in \Pi_{[k,2n+k]} A .$$

Since furthermore $(\theta^\ell)^* \mathbf{P} = \mathbf{P}$ for every $\ell \in \mathbf{Z}$, this implies that $\mathbf{P}(\Pi_{[-n,n]} A) = \mathbf{P}(\Pi_{[1,1+2n]} A)$ for every $n \ge 0$, so that $\mathbf{P}(\Pi_{[1,1+2n]} A \setminus A) \to 0$ as $n \to \infty$. This

[2] Note that in complete generality, the fact that A is a Borel set doesn't imply that $\Pi_I A$ is again a Borel set. However, a result from geometric measure theory [Fed69] states that such sets (*i.e.* images of Borel sets under a continuous map) are *universally measurable*, which means that they belong to the Lebesgue completion of the Borel sets under *any* finite Borel measure.

shows that $A \in \bar{\mathcal{F}}_1^\infty$. The same reasoning shows that $\mathbf{P}(\Pi_{[-1-2n,-1]}A \setminus A) \to 0$ as $n \to \infty$, so that $A \in \bar{\mathcal{F}}_{-\infty}^{-1}$.

We use from now on the notation $A \sim B$ to signify that A and B differ by a set of $\mathbf{P}$-measure 0. By the Definition 2.1 of a Markov process, if f and g are two functions that are respectively $\bar{\mathcal{F}}_1^\infty$ and $\bar{\mathcal{F}}_{-\infty}^{-1}$-measurable, then the equality

$$\mathbf{E}(fg \,|\, \mathcal{F}_0^0) = \mathbf{E}(f \,|\, \mathcal{F}_0^0)\, \mathbf{E}(g \,|\, \mathcal{F}_0^0)\;,$$

holds $\mathbf{P}$-almost surely. Applying this result with $f = g = \mathbf{1}_A$, we find that

$$\mathbf{E}(\mathbf{1}_A^2 \,|\, \mathcal{F}_0^0) = (\mathbf{E}(\mathbf{1}_A \,|\, \mathcal{F}_0^0))^2\;.$$

Since on the other hand $\mathbf{1}_A^2 = \mathbf{1}_A$ and $\mathbf{E}(\mathbf{1}_A \,|\, \mathcal{F}_0^0) \in [0,1]$, one has $\mathbf{E}(\mathbf{1}_A \,|\, \mathcal{F}_0^0) \in \{0,1\}$ almost surely. Let $\hat{A}$ denote the points such that $\mathbf{E}(\mathbf{1}_A \,|\, \mathcal{F}_0^0) = 1$, so that $\hat{A} \in \mathcal{F}_0^0$ by the definition of conditional expectations. Furthermore, the same definition yields $\mathbf{P}(\hat{A} \cap B) = \mathbf{P}(A \cap B)$ for every set $B \in \mathcal{F}_0^0$ and (using the same reasoning as above for $1 - \mathbf{1}_A$) $\mathbf{P}(\hat{A}^c \cap B) = \mathbf{P}(A^c \cap B)$ as well. Using this for $B = \hat{A}$ and $B = \hat{A}^c$ respectively shows that $A \sim \hat{A}$, as required. □

Corollary 5.3 *Let again* $\mathbf{P}$ *be the law of a stationary Markov process. Then, for every set* $A \in \mathcal{I}$ *there exists a measurable set* $\bar{A} \subset \mathcal{X}$ *such that* $A \sim \bar{A}^{\mathbf{Z}}$.

Proof. We know by Proposition 5.2 that $A \in \bar{\mathcal{F}}_0^0$, so that the event A is equivalent to an event of the form $\{x_0 \in \bar{A}\}$ for some $\bar{A} \subset \mathcal{X}$. Since $\mathbf{P}$ is stationary and $A \in \mathcal{I}$, the time 0 is not distinguishable from any other time, so that this implies that A is equivalent to the event $\{x_n \in \bar{A}\}$ for every $n \in \mathbf{Z}$. In particular, it is equivalent to the event $\{x_n \in \bar{A}$ for every $n\}$. □

Note that this result is crucial in the proof of the structure theorem, since it allows us to relate invariant sets $A \in \mathcal{I}$ to invariant sets $\bar{A} \subset \mathcal{X}$, in the following sense:

Definition 5.4 Let $\mathcal{P}$ be a transition operator on a space $\mathcal{X}$ and let μ be an invariant probability measure for $\mathcal{P}$. We say that a measurable set $\bar{A} \subset \mathcal{X}$ is *μ-invariant* if $P(x, \bar{A}) = 1$ for μ-almost every $x \in \bar{A}$.

With this definition, we have

Corollary 5.5 *Let* $\mathcal{P}$ *be a transition operator on a space* $\mathcal{X}$ *and let* μ *be an invariant probability measure for* $\mathcal{P}$. *Then* μ *is ergodic if and only if every* μ*-invariant set* $\bar{A}$ *is of* μ*-measure* 0 *or* 1.

Proof. It follows immediately from the definition of an invariant set that one has $\mu(\bar{A}) = \mathbf{P}_\mu(\bar{A}^{\mathbf{Z}})$ for every μ-invariant set $\bar{A}$.

Now if μ is ergodic, then $\mathbf{P}_\mu(\bar{A}^{\mathbf{Z}}) \in \{0,1\}$ for every set $\bar{A}$, so that in particular $\mu(\bar{A}) \in \{0,1\}$ for every μ-invariant set. If μ is not ergodic, then there exists a set $A \in \mathcal{I}$ such that $\mathbf{P}_\mu(A) \notin \{0,1\}$. By Corollary 5.3, there exists a set $\bar{A} \subset \mathcal{X}$ such that $A \sim \{x_0 \in \bar{A}\} \sim \bar{A}^{\mathbf{Z}}$. The set $\bar{A}$ must be μ-invariant, since otherwise the relation $\{x_0 \in \bar{A}\} \sim \bar{A}^{\mathbf{Z}}$ would fail. □

Proof of Theorem 5.1. Assume first that $\mu \in \mathcal{J}(\mathcal{P})$ is not extremal, *i.e.* it is of the form $\mu = t\mu_1 + (1-t)\mu_2$ with $t \in (0,1)$ and $\mu_i \in \mathcal{J}(\mathcal{P})$. (Note that therefore $\mathbf{P}_\mu = t\mathbf{P}_{\mu_1} + (1-t)\mathbf{P}_{\mu_2}$.) Assume by contradiction that μ is ergodic, so that $\mathbf{P}_\mu(A) \in \{0,1\}$

for every $A \in \mathcal{I}$. If $\mathbf{P}_\mu(A) = 0$, then one must have $\mathbf{P}_{\mu_1}(A) = \mathbf{P}_{\mu_2}(A) = 0$ and smilarly if $\mathbf{P}_\mu(A) = 1$. Therefore, $\mathbf{P}_{\mu_1}$ and $\mathbf{P}_{\mu_2}$ agree on $\mathcal{I}$, so that both $\mathbf{P}_{\mu_1}$ and $\mathbf{P}_{\mu_2}$ are ergodic. Let now $f\colon \mathcal{X}^{\mathbf{Z}} \to \mathbf{R}$ be an arbitrary bounded measurable function and consider the function $f^*\colon \mathcal{X}^{\mathbf{Z}} \to \mathbf{R}$ which is defined by

$$f^*(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(\theta^k(x)) ,$$

on the set E on which this limit exists and by $f^*(x) = 0$ otherwise. Denote by E_i the set of points x such that $f^*(x) = \int f(x)\,\mathbf{P}_{\mu_i}(dx)$. By Corollary 3.5, one has $\mathbf{P}_{\mu_i}(E_i) = 1$, so that in particular $\mathbf{P}_\mu(E_1) = \mathbf{P}_\mu(E_2) = 1$. Since f was arbitrary, one can choose it so that $\int f(x)\,\mathbf{P}_{\mu_1}(dx) \neq \int f(x)\,\mathbf{P}_{\mu_2}(dx)$, which would imply $E_1 \cap E_2 = \phi$, thus contradicting the fact that $\mathbf{P}_\mu(E_1) = \mathbf{P}_\mu(E_2) = 1$.

Let now $\mu \in \mathcal{J}(\mathcal{P})$ be an invariant measure that is not ergodic, we want to show that it can be written as $\mu = t\mu_1 + (1-t)\mu_2$ for some $\mu_i \in \mathcal{J}(\mathcal{P})$ and $t \in (0,1)$. By Corollary 5.5, there exists a set $\bar{A} \subset \mathcal{X}$ such that $\mu(\bar{A}) = t$ and such that $P(x, \bar{A}) = 1$ for μ-almost every $x \in \bar{A}$. Furthermore, one has $\mu(\bar{A}^c) = 1 - t$ and the stationarity of μ implies that one must have $P(x, \bar{A}^c) = 1$ for μ-almost every $x \in \bar{A}^c$. This invariance property immediately implies that the measures μ_i defined by

$$\mu_1(B) = \frac{1}{t}\mu(\bar{A} \cap B) , \qquad \mu_2(B) = \frac{1}{1-t}\mu(\bar{A}^c \cap B) ,$$

belong to $\mathcal{J}(\mathcal{P})$ and therefore have the required property.

The statement about the mutual singularity of any two elements of $\mathcal{E}(\mathcal{P})$ follows immediately from Corollary 3.5. Let indeed μ_1 and μ_2 be two distinct ergodic invariant probability measures. Since they are distinct, there exists a measurable bounded function $f\colon \mathcal{X} \to \mathbf{R}$ such that $\int f(x)\,\mu_1(dx) \neq \int f(x)\,\mu_2(dx)$. Let us denote by $\{x_n\}$ the Markov process with transition operator $\mathcal{P}$ starting at x_0. Then, using the shift map θ in Corollary 3.5, we find that the equality

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(x_n) = \int f(x)\,\mu_i(dx)$$

holds almost surely for μ_i-almost every initial condition x_0 (which is the same as to say that it holds for $\mathbf{P}_{\mu_i}$-almost every sequence x). Since $\int f(x)\,\mu_1(dx) \neq \int f(x)\,\mu_2(dx)$ by assumption, this implies that μ_1 and μ_2 are mutually singular.

The proof of the fact that every invariant measure can be obtained as a convex combination of ergodic invariant measures is a consequence of the ergodic decomposition and will not be given here. □

This structure theorem allows to draw several important conclusions concerning the set of all invariant probability measures of a given Markov process. For example, we have that

Corollary 5.6 *If a Markov process with transition operator $\mathcal{P}$ has a unique invariant measure μ, then μ is ergodic.*

Proof. In this case $\mathcal{J}(\mathcal{P}) = \{\mu\}$, so that μ is an extremal point of $\mathcal{J}(\mathcal{P})$. □

6 Existence of an invariant measure

This section is devoted to the study of criteria for the existence of an invariant measure for a Markov semigroup $(\mathcal{P}_t)_{t\geq 0}$ over a Polish space $\mathcal{X}$. In most of this section, we will assume that the semigroup has the *Feller property*, that is it maps the space $\mathcal{C}_b(\mathcal{X})$ of continuous bounded functions into itself. This is a sufficient regularity property to be able to apply the following general criterion:

Theorem 6.1 (Krylov-Bogolioubov) *Let $(\mathcal{P}_t)_{t\geq 0}$ be a Feller Markov semigroup over a Polish space $\mathcal{X}$. Assume that there exists $\mu_0 \in \mathscr{M}_1(\mathcal{X})$ such that the sequence $\{\mathcal{P}_t\mu_0\}$ is tight. Then, there exists at least one invariant probability measure for $(\mathcal{P}_t)_{t\geq 0}$.*

Proof. Let $\mathcal{R}_t$ be the sequence of probability measures defined by

$$\mu_t(A) = \frac{1}{t}\int_0^t (\mathcal{P}_s\mu_0)(A)\,ds\;.$$

Since we assumed that $\{\mathcal{P}_t\mu_0\}$ is tight, it is straightforward to check that $\{\mu_t\}$ is also tight (just take the same compact set). Therefore, there exists at least one accumulation point μ_* and a sequence t_n with $t_n \to \infty$ such that $\mu_{t_n} \to \mu_*$ weakly. Take now an arbitrary test function $\varphi \in \mathcal{C}_b(\mathcal{X})$. One has

$$\begin{aligned}
|(\mathcal{P}_t\mu_*)(\varphi) - \mu_*(\varphi)| &= |\mu_*(\mathcal{P}_t\varphi) - \mu_*(\varphi)| = \lim_{n\to\infty}|\mu_{t_n}(\mathcal{P}_t\varphi) - \mu_{t_n}(\varphi)| \\
&= \lim_{n\to\infty}|\mu_{t_n}(\mathcal{P}_t\varphi) - \mu_{t_n}(\varphi)| \\
&= \lim_{n\to\infty}\frac{1}{t_n}\Big|\int_{t_n}^{t+t_n}\mu_0(\mathcal{P}_s\varphi)\,ds - \int_0^t \mu_0(\mathcal{P}_s\varphi)\,ds\Big| \\
&\leq \lim_{n\to\infty}\frac{2t}{t_n}\sup_{x\in\mathcal{X}}|\varphi(x)| = 0\;.
\end{aligned}$$

Here, the second equality relies on the fact that $\mathcal{P}_t\varphi$ is continuous since $\mathcal{P}_t$ was assumed to be Feller. Since both φ and t were arbitrary, this shows that $\mathcal{P}_t\mu_* = \mu_*$ for every t as requested. □

Remark 6.2 It is a straightforward exercise to show that the same proof also works for discrete times by taking $\mu_N(A) = \frac{1}{N}\sum_{n=1}^N (\mathcal{P}^n\mu_0)(A)$.

Example 6.3 Take $\mathcal{X} = [0,1]$ and consider the transition probabilities defined by

$$\mathcal{P}(x,\cdot) = \begin{cases} \delta_{x/2} & \text{if } x > 0 \\ \delta_1 & \text{if } x = 0. \end{cases}$$

It is clear that this Markov operator cannot have any invariant probability measure. Indeed, assume that μ is invariant. Clearly, one must have $\mu(\{0\}) = 0$ since $\mathcal{P}(x,\{0\}) = 0$ for every x. Since, for $x \neq 0$, one has $\mathcal{P}(x,\{(1/2,1]\}) = 0$, one must also have $\mu((1/2,1]) = 0$. Proceeding by induction, we have that $\mu((1/2^n,1]) = 0$ for every n and therefore $\mu((0,1]) = 0$. Therefore, $\mu(\mathcal{X}) = 0$ which is a contradiction.

Endowing $\mathcal{X}$ with the usual topology, it is clear that the 'Feller' assumption of the Krylov-Bogolioubov criteria is not satisfied around 0. The tightness criterion however is satisfied since $\mathcal{X}$ is a compact space. On the other hand, we could add the set $\{0\}$ to the topology of $\mathcal{X}$, therefore really interpreting it as $\mathcal{X} = \{0\} \sqcup (0,1]$. Since $\{0\}$

already belongs to the Borel σ-algebra of $\mathcal{X}$, this change of topology does not affect the Borel sets. Furthermore, the space $\mathcal{X}$ is still a Polish space and it is easy to check that the Markov operator $\mathcal{P}$ now has the Feller property! However, the space $\mathcal{X}$ is no longer compact and a sequence $\{x_n\}$ accumulating at 0 is no longer a precompact set, so that it is now the tightness assumption that is no longer satisfied.

7 A simple yet powerful uniqueness criterion

Many uniqueness criteria for the invariant measure of a Markov process rely at a fundamental level on the following simple lemma:

Lemma 7.1 *If the set $\mathcal{J}(\mathcal{P})$ of invariant measure for a Markov operator $\mathcal{P}$ over a Polish space $\mathcal{X}$ contains more than one element, then there exist at least two elements $\mu_1, \mu_2 \in \mathcal{J}(\mathcal{P})$ such that μ_1 and μ_2 are mutually singular.*

Proof. Assume that $\mathcal{J}(\mathcal{P})$ has at least two elements. Since, by Theorem 5.1, every invariant measure can be obtained as a convex combination of ergodic ones, $\mathcal{J}(\mathcal{P})$ must contain at least two distinct ergodic invariant measures, say μ_1 and μ_2, which are mutually singular by Theorem 5.1. □

As a consequence of this lemma, if $\mathcal{J}(\mathcal{P})$ contains more than one invariant measure, the state space $\mathcal{X}$ can be partitioned into (at least) two disjoint parts $\mathcal{X} = \mathcal{X}_1 \sqcup \mathcal{X}_2$ with the property that if the process starts in $\mathcal{X}_1$, then it will stay in $\mathcal{X}_1$ for all times almost surely and the same applies to $\mathcal{X}_2$. The intuition that derives from this consideration is that uniqueness of the invariant measure is a consequence of the process visiting a "sufficiently large" portion of the phase space, independently of its initial position. The remainder of this section is devoted to several ways of formalising this intuition.

The following definition captures what we mean by the fact that a given point of the state space can be 'visited' by the dynamic:

Definition 7.2 Let $\{\mathcal{P}_t\}$ be a Markov semigroup over a Polish space $\mathcal{X}$ and let $x \in \mathcal{X}$. Define the resolvent operator $\mathcal{R}_\lambda$ for $\mathcal{P}_t$ by

$$\mathcal{R}_\lambda(y, U) = \lambda \int_0^\infty e^{-\lambda t} \mathcal{P}_t(y, U)\, dt ,$$

which is again a Markov operator over $\mathcal{X}$. We say that x is *accessible* for $\{\mathcal{P}_t\}$ if, for every $y \in \mathcal{X}$ and every open neighborhood U of x, one has $\mathcal{R}_\lambda(y, U) > 0$. (Note that this definition does not depend on the choice of λ.)

It is straightforward to show that if a given point is reachable, then it must belong to the topological support of every invariant measure of the semigroup:

Lemma 7.3 *Let $\{\mathcal{P}_t\}$ be a Markov semigroup over a Polish space $\mathcal{X}$ and let $x \in \mathcal{X}$ be accessible. Then, $x \in \operatorname{supp} \mu$ for every $\mu \in \mathcal{J}(\mathcal{P})$.*

Proof. Let μ be invariant for the Markov semigroup $\{\mathcal{P}_t\}$, let $\lambda > 0$, and let $U \subset \mathcal{X}$ be an arbitrary neighborhood of x. The invariance of μ implies that

$$\mu(U) = \int_{\mathcal{X}} \mathcal{R}_\lambda(y, U)\, \mu(dy) > 0 ,$$

as required. □

It is important to realise that this definition depends on the topology of $\mathcal{X}$ and not just on the Borel σ-algebra. Considering again Example 6.3, we see that the point 0 is reachable when $[0, 1]$ is endowed with its usual topology, whereas it is *not* reachable if we interpret the state space as $\{0\} \sqcup (0, 1]$. Therefore, as in the previous section, this definition can be useful only in conjunction with an appropriate regularity property of the Markov semigroup. The following example shows that the Feller property is too weak to serve our purpose.

Example 7.4 (Ising model) The Ising model is one of the most popular toy models of statistical mechanics. It is one of the simplest models describing the evolution of a ferromagnet. The physical space is modelled by a lattice $\mathbf{Z}^d$ and the magnetisation at each lattice site is modelled by a 'spin', an element of $\{\pm 1\}$. The state space of the system is therefore given by $\mathcal{X} = \{\pm 1\}^{\mathbf{Z}^2}$, which we endow with the product topology. This topology can be metrized for example by the distance function

$$d(x, y) = \sum_{k \in \mathbf{Z}^2} \frac{|x_k - y_k|}{2^{|k|}} ,$$

and the space $\mathcal{X}$ endowed with this distance function is easily seen to be separable.

The (Glauber) dynamic for the Ising model depends on a parameter β and can be described in the following way. At each lattice site, we consider independent clocks that ring at Poisson distributed times. Whenever the clock at a given site (say the site k) rings, we consider the quantity $\delta E_k(x) = \sum_{j \sim k} x_j x_k$, where the sum runs over all sites j that are nearest neighbors of k. We then flip the spin at site k with probability $\min\{1, \exp(-\beta\, \delta E_k(x))\}$.

Let us first show that every point is accessible for this dynamic. Fix an arbitrary configuration $x \in \mathcal{X}$ and a neighbourhood U containing x. By the definition of the product topology, U contains an 'elementary' neighbourhood $U_N(x)$ of the type $U_N(x) = \{y \in \mathcal{X} \,|\, y_k = x_k \ \forall |k| \leq N\}$. Given now an arbitrary initial condition $y \in \mathcal{X}$, we can find a sequence of m spin flips at distinct locations $k_1, \ldots, k_m$, all of them located inside the ball $\{|k| \leq N\}$, that allows to go from y into $U_N(x)$. Fix now $t > 0$. There is a very small but nevertheless strictly positive probability that within that time interval, the Poisson clocks located at $k_1, \ldots, k_m$ ring exactly once and exactly in that order, whereas all the other clocks located in the ball $\{|k| \leq N + 2\}$ do not ring. Furthermore, there is a strictly positive probability that all the corresponding spin flips do actually happen. As a consequence, the Ising model is *topologically irreducible* in the sense that for any state $x \in \mathcal{X}$, any open set $U \subset \mathcal{X}$ and any $t > 0$, one has $\mathcal{P}_t(x, U) > 0$.

It is also relatively straightforward to show that the dynamic has the Feller property, but this is outside the scope of these notes. However, despite the fact that the dynamic is both Feller and topologically irreducible, one has the following:

Theorem 7.5 *For $d \geq 2$ there exists $\beta_c > 0$ such that the Ising model has at least two distinct invariant measures for $\beta > \beta_c$.*

The proof of this theorem is not simple and we will not give it here. It was a celebrated tour de force by Onsager to be able to compute the critical value $\beta_c = \ln(1 + \sqrt{2})/2$ explicitly in [Ons44] for the case $d = 2$. We refer to the monograph [Geo88] for a more detailed discussion of this and related models.

This example shows that if we wish to base a uniqueness argument on the accessibility of a point or on the topological irreducibility of a system, we need to combine this with a stronger regularity property than the Feller property. One possible regularity property that yields the required properties is the *strong Feller* property:

Definition 7.6 A Markov operator $\mathcal{P}$ over a Polish space $\mathcal{X}$ has the *strong Feller* property if, for every function $\varphi \in \mathcal{B}_b(\mathcal{X})$, one has $\mathcal{P}\varphi \in \mathcal{C}_b(\mathcal{X})$.

With this definition, one has:

Proposition 7.7 *If a Markov operator $\mathcal{P}$ over a Polish space $\mathcal{X}$ has the strong Feller property, then the topological supports of any two mutually singular invariant measures are disjoint.*

Proof. Let μ and ν be two mutually singular invariant measures for $\mathcal{P}$. Since they must be mutually singular by Theorem 5.1, there exists a set $A \subset \mathcal{X}$ such that $\mu(A) = 1$ and $\nu(A) = 0$. The invariance of μ and ν then implies that $\mathcal{P}(x, A) = 1$ for μ-almost every x and $\mathcal{P}(x, A) = 0$ for ν-almost every x.

Set $\varphi = \mathcal{P}\mathbf{1}_A$, where $\mathbf{1}_A$ is the characteristic function of A. It follows from the previous remarks that $\varphi(x) = 1$ μ-almost everywhere and $\varphi(x) = 0$ ν-almost everywhere. Since φ is continuous by the strong Feller property, the claim now follows from the fact that if a continuous function is constant μ-almost everywhere, it must be constant on the topological support of μ. □

Actually, by looking at the proof of Proposition 7.7, one realises that one could have introduced a notion of being strong Feller at a point $x \in \mathcal{X}$ in a natural way by imposing that $\mathcal{P}\varphi$ is continuous at x for every bounded measurable function φ. With this notation, the same proof as above allows to conclude that if $\mathcal{P}$ is strong Feller at x, then x can belong to the support of at most one invariant probability measure. This leads to one of the most general uniqueness criteria commonly used in the literature:

Corollary 7.8 *If $\mathcal{P}$ is strong Feller at an accessible point $x \in \mathcal{X}$ for $\mathcal{P}$, then it can have at most one invariant measure.*

Proof. Combine Proposition 7.7 with Lemma 7.3 and the fact that if $\mathcal{J}(\mathcal{P})$ contains more than one element, then by Theorem 5.1 there must be at least two distinct ergodic invariant measures for $\mathcal{P}$. □

Exercise 7.9 Let ξ_n be an i.i.d. sequence of real-valued random variables with law μ. Define a real-valued Markov process x_n by $x_{n+1} = \frac{1}{2}x_n + \xi_n$. Show that if μ has a continuous density with respect to Lebesgue measure, then the corresponding Markov operator has the strong Feller property.

8 Hörmander's condition

The Markov processes considered in this section are diffusions with smooth coefficients:

$$dx(t) = f_0(x(t))\,dt + \sum_{i=1}^{m} f_i(x(t)) \circ dw_i(t)\;. \tag{8.1}$$

Here, the process x is $\mathbf{R}^n$-valued, the functions $f_j \colon \mathbf{R}^n \to \mathbf{R}^n$ are assumed to be $\mathcal{C}^\infty$ with bounded derivatives of all orders, and the w_i's are i.i.d. standard Wiener processes.

It is a standard result from stochastic analysis that (8.1) has a unique solution for every initial condition $x_0 \in \mathbf{R}^n$ and that these solutions have the Markov property [Øks03b, Kry95].

Denote by $\mathcal{P}_t$ the Markov semigroup associated to solutions of (8.1), that is

$$(\mathcal{P}_t \varphi)(x) = \mathbf{E}(\varphi(x(t)) \,:\, x(0) = x) \;.$$

The aim of this section is to provide a criteria that is not difficult to verify in practice and that guarantees that $\mathcal{P}_t \varphi$ is smooth for every bounded measurable function φ.

Given two smooth vector fields $f, g\colon \mathbf{R}^n \to \mathbf{R}^n$, we define their Lie bracket by

$$[f, g](x) = Dg(x)f(x) - Df(x)g(x) \;.$$

Here, Df and Dg denote the Fréchet derivatives of f and g. With this notation at hand, we define an increasing sequence of families of vector fields recursively by

$$\begin{aligned} A_0 &= \{f_j \,:\, j = 1, \ldots, m\} \;, \\ A_{k+1} &= A_k \cup \{[g, f_j] \,:\, g \in A_k,\; j = 0, \ldots, m\} \;. \end{aligned}$$

Denoting $A_\infty = \bigcup_{k\geq 0} A_k$, we can associate to each point $x \in \mathbf{R}^n$ a family of subspaces $\bar{A}_k(x)$ of $\mathbf{R}^n$ by

$$\bar{A}_k(x) = \text{span}\{g(x) \,:\, g \in A_k\} \;.$$

We will say that *Hörmander's condition holds* at a point $x \in \mathbf{R}^n$ if $\bar{A}_\infty(x) = \mathbf{R}^n$. With this notation, we have the following result:

Theorem 8.1 (Hörmander) *If Hörmander's condition holds at some $x \in \mathbf{R}^n$ then, for every bounded measurable function φ, the function $\mathcal{P}_t \varphi$ is smooth in a neighbourhood of x.*

Remark 8.2 The easiest way for Hörmander's condition to hold is if $\bar{A}_0(x) = \mathbf{R}^n$ for every x. In this case, the generator

$$\mathcal{L} = f_0 \nabla + \sum_{i=1}^m (f_i \nabla)^2 \tag{8.2}$$

of the diffusion (8.1) is an elliptic operator. If Hörmander's condition is satisfied at every x, then we will say that (8.2) is a *hypoelliptic* operator.

This result was first proven by Hörmander in a slightly different form in [Hör67] (see also the monograph [Hör85]) by using purely functional-analytic techniques. Since this result has such strong implications in probability theory, there was a rush in the seventies to obtain a probabilistic proof. This finally emerged from the works of Malliavin, Bismut, Stroock and Kusuoka [Mal78, Bis81, KS84, KS85, KS87], using Malliavin calculus techniques. We are now going to give a short overview along the lines of [Nor86] on how this probabilistic proof works, but without entering into all the technical details. For a complete proof, we refer for example to the monograph [Nua95].

The main technical tool used for the proof of Hörmander's theorem is Malliavin's integration by part formula, so let us start by giving a short survey of Malliavin calculus. The idea is to generalise to Wiener space the fact that, if μ is a product of d independent standard Gaussian measures with variance σ^2 and $f\colon \mathbf{R}^d \to \mathbf{R}$ is a smooth function, one has the identity

$$\int_{\mathbf{R}^d} D_v f(x)\,\mu(dx) = \frac{1}{\sigma^2}\int_{\mathbf{R}^d} f(x)\sum_{k=1}^d v_k x_k\,\mu(dx)\;, \tag{8.3}$$

where D_v denotes the directional derivative in the direction $v \in \mathbf{R}^d$.

Denote by $(\Omega, \mathbf{P})$ standard Wiener space (say on the time interval $[0,1]$) endowed with Wiener measure. It is a basic fact in the theory of Brownian motion that $\mathbf{P}$ is quasi-invariant under translations by elements in the Cameron-Martin space $\mathcal{H} = H^1 = \{\int_0^\cdot v(t)\,dt \,:\, v \in L^2([0,1])\}$. Furthermore, although $\mathcal{H}$ itself is of $\mathbf{P}$-measure 0, it can be characterised as the intersection of all linear subspaces of Ω that have full measure [Bog98]. This suggests that the 'right' set of directions in which to consider directional derivatives in Wiener space is given by the Cameron-Martin space $\mathcal{H}$. Furthermore, if we 'slice' the time interval $[0,1]$ into small slices of size δt, then all these increments are independent Gaussians with variance δt. A perturbation of Brownian motion into the direction $h = \int_0^\cdot v(s)\,ds$ then corresponds precisely to a perturbation of each increment into the direction $v(t)\,\delta t$. Equation (8.3) then yields

$$\begin{aligned}\int_\Omega D_h f(w)\,\mathbf{P}(dw) &\approx \frac{1}{\delta t}\int_\Omega f(w)\sum v(t)\delta t\delta w(t)\,\mathbf{P}(dw)\\ &\approx \int_\Omega f(w)\int_0^1 h(t)\,dw(t)\,\mathbf{P}(dw)\;.\end{aligned}$$

These approximations can be justified rigorously, and lead to the integration by parts formula:

$$\mathbf{E}\mathscr{D}_v f(w) = \mathbf{E}\Big(f(w)\int_0^1 v(t)\,dw(t)\Big)\;. \tag{8.4}$$

Here, $f\colon \Omega \to \mathbf{R}$ is a random variable with sufficient smoothness, $v \in L^2([0,1])$, and the Malliavin derivative $\mathscr{D}_v f$ of f in the direction v is defined as

$$\mathscr{D}_v f = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}(\tau_{\varepsilon v} f - f)\;, \tag{8.5}$$

where the translation operatur τ is defined by $(\tau_v f)(w) = f(w + \int_0^\cdot v(s)\,ds)$ and the limit in (8.5) is a limit in probability. It turns out that (8.4) is still true if v is not a deterministic element of $\mathcal{H}$, but an $\mathcal{H}$-valued random variable (which is a stochastic process with sample paths that are almost surely L^2-valued). If it turns out that v is adapted, then the stochastic integral appearing in the right hand side of (8.4) is the usual Itô integral. If v is not adapted, then it is a Skorokhod integral, see [Nua95].

We are not going to enter into the details of the definition of the Skorokhod integral in these notes. Suffices to say that one has an extension of the usual Itô isometry which is given by

$$\mathbf{E}\Big(\int_0^1 v(s)dw(s)\Big)^2 = \mathbf{E}\int_0^1 v^2(s)\,ds + \mathbf{E}\int_0^1\int_0^1 \mathscr{D}_s v(t)\mathscr{D}_t v(s)\,ds\,dt\;,$$

where, at least at a formal level, $\mathscr{D}_t f$ is given by the limit of $\mathscr{D}_v f$ as v approaches a Dirac delta-function centred in t. Of course, for this to be well-defined, additional regularity is required on v and simply being an L^2-valued process is not sufficient. An exception is the situation where v is an adapted process. In this case, $v(t)$ does not depend on the increments of the Wiener process before time t, so that $\mathscr{D}_s v(t) = 0$ for $s < t$. Therefore, one of the two factors appearing in the double integral always vanishes and one recovers the usual Itô isometry.

How does all this help for the proof of Hörmander's theorem? Using the chain rule and Fubini's theorem, we see that if φ is a sufficiently smooth function and ξ is an arbitrary element of $\mathbf{R}^n$, one has the identity

$$D_\xi \mathcal{P}_t \varphi(x) = \mathbf{E}(D\varphi(x_t) D_\xi x_t) = \mathbf{E}(D\varphi(x_t) J_{0,t}\xi) ,$$

were we denoted by $J_{s,t}$ the Jacobian of the solution map of (8.1) between two times s and t. Suppose now that, given $\xi \in \mathbf{R}^n$, we can find a process $v_\xi \in L^2([0,t], \mathbf{R}^m)$ such that the derivative $J_{0,t}\xi$ of the solution to (8.1) in the direction ξ with respect to its initial condition is *equal* to its Malliavin derivative $\mathscr{D}_{v_\xi} x_t$ in the direction of the process v_ξ. We could then use (8.4) to write

$$D_\xi \mathcal{P}_t \varphi(x) = \mathbf{E}(D\varphi(x_t) \mathscr{D}_{v_\xi} x_t) = \mathbf{E}(\mathscr{D}_{v_\xi}(\varphi(x_t))) = \mathbf{E}\Big(\varphi(x_t) \int_0^t v_\xi(s)\, dw(s)\Big) ,$$

thus obtaining a bound on the derivative of $\mathcal{P}_t\varphi$ which is uniform over all functions φ with a given supremum bound. Iterating such a procedure would then lead to the proof of Hörmander's theorem. The main moral that one should take home from this story is that Malliavin calculus allows to transform a regularity problem (showing that $\mathcal{P}_t$ has a smoothing property) into a linear control problem (find a control v such that perturbing the noise by v has the same effect as a given perturbation in the initial condition).

The aim of the remainder of this section is to give an idea on how to construct such a 'control' v_ξ in the framework given by the assumptions of Hörmander's theorem. The main insight required to perform such a construction is to realise that the Malliavin derivative of x_t is intimately related to the Jacobian. Formally taking the derivative of (8.1) in the direction of the w_i indeed yields

$$d\mathscr{D}_v x_t = Df_0(x_t)\mathscr{D}_v x_t\, dt + \sum_{i=1}^m Df_i(x_t)\mathscr{D}_v x_t \circ dw_i(t) + \sum_{i=1}^m f_i(x_t) v_i(t)\, dt ,$$

endowed with the initial condition $\mathscr{D}_v x_0 = 0$, whereas taking derivatives with respect to the initial condition yields the very similar expression

$$dJ_{s,t}\xi = Df_0(x_t) J_{s,t}\xi\, dt + \sum_{i=1}^m Df_i(x_t) J_{s,t}\xi \circ dw_i(t) ,$$

endowed with the initial condition $J_{s,s}\xi = \xi$. This allows to solve the equation for $\mathscr{D}_v x_t$ using the variation of constants formula, thus obtaining the expression

$$\mathscr{D}_v x_t = \int_0^t J_{s,t} f_i(x_s) v_i(s)\, ds = J_{0,t} \int_0^t J_{0,s}^{-1} f_i(x_s) v_i(s)\, ds \equiv J_{0,t}\mathscr{A}_{0,t} v ,$$

where the (random) linear operator $\mathscr{A}_{0,t}$ maps $L^2([0,t], \mathbf{R}^n)$ into $\mathbf{R}^n$. With these notations in place, our control problem is now to find a control v_ξ such that one has the identity $J_{0,t}\mathscr{A}_{0,t} v = J_{0,t}\xi$ which, since the Jacobian is invertible for the class of problems that we consider, is equivalent to the identity

$$\mathscr{A}_{0,t} v = \xi . \tag{8.6}$$

At this stage, if we knew that the operator $\mathscr{M}_{0,t} \equiv \mathscr{A}_{0,t}\mathscr{A}_{0,t}^*$ was invertible, we could solve (8.6) by setting

$$v = \mathscr{A}_{0,t}^* \mathscr{M}_{0,t}^{-1} \xi \;.$$

It therefore remains to argue that $\mathscr{M}_{0,t}$ is indeed invertible, provided that Hörmander's condition holds at the initial point x.

Note first that, given $\xi \in \mathbf{R}^n$, the expression $\langle \xi, \mathscr{M}_{0,t}\xi\rangle$ is given by

$$\langle \xi, \mathscr{M}_{0,t}\xi\rangle = \sum_{i=1}^m \int_0^t \langle \xi, J_{0,s}^{-1} f_i(x_s)\rangle^2 \, ds \;. \tag{8.7}$$

We will argue that if ξ is any deterministic element in $\mathbf{R}^n$, then the probability of $\langle \xi, \mathscr{M}_{0,t}\xi\rangle$ being small is very small. We hope that it is plausible to the reader in view of Exercise 8.3 below that such a statement can indeed be turned into a more quantifiable statement on the invertibility of $\mathscr{M}_{0,t}$. The main technical tool at this stage is Norris' lemma which is a quantitative version of the Meyer-Doob decomposition.

Exercise 8.3 Let M be a random positive semidefinite $d \times d$ matrix such that $\|M\| \le 1$ almost surely. Assume that for every $p > 0$ one can find a constant $C_p > 0$ such that the bound

$$\sup_{\|\xi\|=1} \mathbf{P}(\langle \xi, M\xi\rangle \le \varepsilon) \le C_p \varepsilon^p \;,$$

holds for ε sufficiently small. Show that this implies the existence of a possibly different family of constants C_p' such that the bound

$$\mathbf{P}\Big(\inf_{\|\xi\|=1} \langle \xi, M\xi\rangle \le \varepsilon\Big) \le C_p' \varepsilon^p \;,$$

holds for ε small enough. Deduce that the matrix M is then invertible almost surely and that its inverse has moments of all orders.

Hint: Decompose the sphere $\|\xi\| = 1$ into small patches of radius ε^2 and argue separately on each patch.

Let a and b be two adapted real-valued and $\mathbf{R}^m$-valued process respectively satisfying sufficient regularity assumptions and consider the process z defined by

$$z(t) = \int_0^t a(s)\, ds + \int_0^t b(s) \circ dw(s) \;.$$

Then Norris' lemma states that if z is small then, with high probability, both a and b are small. Using the fact that the inverse of the Jacobian satisfies the equation

$$dJ_{s,t}^{-1} = -J_{s,t}^{-1} Df_0(x_t)\, dt - \sum_{i=1}^m J_{s,t}^{-1} Df_i(x_t) \circ dw_i(t) \;,$$

It is now straightforward to check that if g is any smooth vector field, then the process $z(t) = \langle \xi, J_{0,s}^{-1} g(x_s)\rangle$ satisfies the SDE

$$dz(t) = \langle \xi, J_{0,s}^{-1} [f_0, g](x_s)\rangle \, dt + \sum_{i=1}^m \langle \xi, J_{0,s}^{-1} [f_i, g](x_s)\rangle \circ dw_i(t) \tag{8.8}$$

In order to show that $\langle \xi, \mathscr{M}_{0,t}\xi\rangle$ cannot be too small, we now argue by contradiction. Assume that $\langle \xi, \mathscr{M}_{0,t}\xi\rangle$ is very small then, by (8.7), the processes $\langle \xi, J_{0,s}^{-1} f_i(x_s)\rangle$ must all be very small as well. On the other hand, Norris' lemma combined with (8.8) shows that if any process of the type $\langle \xi, J_{0,s}^{-1} g(x_s)\rangle$ is small, then the processes $\langle \xi, J_{0,s}^{-1}[f_i, g](x_s)\rangle$ must also be small for $j = 0, \ldots, m$. In particular, $\langle \xi, [f_i, g](x_0)\rangle$ must be small.

Iterating this argument shows that if $\langle \xi, \mathscr{M}_{0,t}\xi\rangle$ is very small, then $\langle \xi, g(x_0)\rangle$ must be small for all $g \in A_\infty$, which is in direct contradiction with Hörmander's condition.

Hörmander's theorem is a very neat way of showing that a diffusion has the strong Feller property. Combined with Stroock-Varadhan's support theorem, it allows very often to verify that the assumptions of Corollary 7.8 hold:

Theorem 8.4 (Stroock-Varadhan) *Given a diffusion (8.1) written in Stratonowich form, we associate to it the control problem*

$$\frac{dx(t)}{dt} = f_0(x(t)) + \sum_{i=1}^m f_i(x(t))\, u_i(t)\;, \quad x(0) = x\;. \tag{8.9}$$

Then, the support of the transition probabilities $\mathcal{P}_t(x, \cdot\,)$ is precisely given by the closure of all points in $\mathbf{R}^d$ that can be reached in time t by solutions to (8.9) with the u_i given by arbitrary smooth functions.

Exercise 8.5 Consider the folloving Langevin equation:

$$dq = p\, dt\;, \qquad dp = -\nabla V(q)\, dt - p\, dt + dw(t)\;, \tag{8.10}$$

Here, both p and q take values in $\mathbf{R}^n$, $V\colon \mathbf{R}^n \to \mathbf{R}$ is a smooth function that we assume to grow to infinity at least at algebraic speed and w is a standard n-dimensional Wiener process. Show that (8.10) can have at most one invariant probability measure.

Exercise 8.6 A slight elaboration on the previous example is given by a finite chain of nonlinear oscillators coupled to heat baths at the ends:

$$\begin{aligned}
dq_i &= p_i\, dt\;, \qquad i = 0, \ldots, N\;,\\
dp_0 &= -\nabla V_1(q_0)\, dt - \nabla V_2(q_0 - q_1)\, dt - p_0\, dt + \sqrt{2T_L}\, dw_L(t)\;,\\
dp_j &= -\nabla V_1(q_j)\, dt - \nabla V_2(q_j - q_{j-1})\, dt - \nabla V_2(q_j - q_{j+1})\, dt\;,\\
dp_N &= -\nabla V_1(q_N)\, dt - \nabla V_2(q_N - q_{N-1})\, dt - p_N\, dt + \sqrt{2T_R}\, dw_R(t)\;.
\end{aligned}$$

Show that if the coupling potential V_2 is strictly convex (so that its Hessian is strictly positive definite in every point), then this equation does satisfy Hörmander's condition, so that it satisfies the strong Feller property. *Harder:* Show that every point is reachable, using Stroock-Varadhan's support theorem.

9 What about the infinite-dimensional case?

There are various problems that arise if one tries to transpose the kind of arguments presented previously to the setting of stochastic PDEs, where the Markov process x takes values in some infinite-dimensional space of functions (either a Banach space or a Hilbert space).

First, the Jacobian for a parabolic PDE (or SPDE) is not usually an invertible operator, so we can not reduce ourselves to the situation (8.6) where the process appearing in the definition of $\mathscr{A}$ is adapted. Furthermore, the question of the invertibility of the operator $\mathscr{M}_{0,t}$ is much more subtle in infinite dimensions. As a consequence, one cannot expect in general that a smoothing theorem along the lines of the statement of Hörmander's theorem given previously holds in infinite dimensions.

Let us for example consider the following infinite-dimensional system of SDEs:

$$dx_k = -x_k\,dt + e^{-k^2}\,dw_k(t)\;, \quad k \in \mathbf{Z}\;, \tag{9.1}$$

which we consider as an evolution in the Hilbert space ℓ^2 of square-summable sequences. It follows from the variation of constants formula that the solution to (9.1) is given by

$$x(t) = e^{-t}x_0 + \int_0^t e^{-(t-s)} Q\,dW(t)\;. \tag{9.2}$$

Here, we denote by $Q\colon \ell^2 \to \ell^2$ the operator acting on the canonical unit vectors as $Qe_k = e^{-k^2}e_k$, and W is a cylindrical Wiener process on ℓ, that is

$$W(t) = \sum_{k\in\mathbf{Z}} e_k\,w_k(t)\;, \tag{9.3}$$

with the w_k's a sequence of i.i.d. standard Wiener processes. Note that the sum in (9.3) does *not* converge in ℓ^2, but one can convince oneself that the expression (9.2) is well-defined and does take values in ℓ^2.

Exercise 9.1 Show that the solution to (9.1) does indeed live in ℓ^2 almost surely and that the Markov semigroup given by (9.2) has the Feller property.

Consider now the subset $A \subset \ell^2$ of sequences with fast decay:

$$A = \Bigl\{x \in \ell^2 \;:\; \sup_k |x_k||k|^N < \infty \quad \forall N > 0\Bigr\}\;.$$

It a straightforward calculation to show that, if $x_0 = 0$, then the right hand side of (9.2) belongs to A almost surely. Therefore, since A is a vector space, one has $x(t) \in A$ if and only if $x_0 \in A$. In other words, the characteristic function of the set A is left invariant by the Markov semigroup associated to (9.2), thus showing that it does *not* have the strong Feller property, despite the diffusion (9.1) looking perfectly 'elliptic'.

This is however not the generic situation. Before we show an example showing that the strong Feller property can sometimes also be satisfied in infinite-dimensional spaces, let us recall some basics of the theory of Gaussian measures on Hilbert spaces. A measure μ on a (separable) Banach space $\mathcal{B}$ is called *Gaussian* if the law of $\ell^*\mu$ is Gaussian for every continuous linear functional $\ell\colon \mathcal{B} \to \mathbf{R}$. The *covariance operator* C_μ of μ is the bounded linear operator from $\mathcal{B}^*$ to $\mathcal{B}$ such that the identity

$$u(C_\mu v) = \int_{\mathcal{H}} u(x)v(x)\,\mu(dx)\;,$$

holds for any $u, v \in \mathcal{B}^*$. The two main theorems from Gaussian measure theory are then given by:

Theorem 9.2 (Fernique) *Let μ be a Gaussian measure on a Banach space $\mathcal{B}$. Then, the norm of $x \in \mathcal{B}$ has Gaussian tails under μ.*

Theorem 9.3 (Cameron-Martin) *Let μ be a Gaussian measure on a separable Banach space $\mathcal{B}$. Then, there exists a Hilbert space $\mathcal{H}_\mu \subset \mathcal{B}$ which can be equivalently characterised as*

- *The closure of the space*

$$\hat{\mathcal{H}}_\mu = \{h \in \mathcal{B} \,:\, \exists\, h^* \in \mathcal{B}^* \text{ with } C_\mu(h^*, \ell) = \ell(h)\ \forall \ell \in \mathcal{B}^*\} ,$$

  under the norm $\|h\|_\mu^2 = C_\mu(h^*, h^*)$.
- *The intersection of all measurable subspaces of $\mathcal{B}$ that have measure 1 under μ.*
- *The set of all elements $h \in \mathcal{B}$ such that $\tau_h^*\mu$ is absolutely continuous with respect to μ.*

Furthermore, the correspondence $h \leftrightarrow h^*$ extends to an isometry between $\mathcal{H}_\mu$ and the closure $\mathcal{R}_\mu$ of $\mathcal{B}^*$ viewed as a subset of $L^2(\mathcal{B}, \mu)$. For any $h \in \mathcal{H}_\mu$, the Radon-Nikodym density of $\tau_h^*\mu$ with respect to μ is given by $\exp(h^*(x) - \frac{1}{2}\|h\|_\mu^2)$.

Exercise 9.4 If μ is a Gaussian measure on a Hilbert space $\mathcal{H}$, the Riesz's representation theorem allows to identify C_μ with an operator from $\mathcal{H}$ to $\mathcal{H}$. As a consequence of Fernique's theorem, show that in this case, the operator C_μ is selfadjoint, positive semidefinite, and trace class, that is $\sum_k \langle e_k, C_\mu e_k\rangle < \infty$ for any orthonormal basis $\{e_k\}$ of $\mathcal{H}$.

Remark 9.5 If y is an $\mathcal{H}$-valued random variable with covariance C and A is a bounded linear operator on $\mathcal{H}$, then Ay is Gaussian with covariance ACA^*. This suggests that $C^{-1/2}y$ would be Gaussian with the identity as its covariance operator. If $\mathcal{H}$ is infinite-dimensional, then the identity is not trace class, so such a random variable obviously doesn't exist. However, for $h \in \mathcal{H}$, it is always possible to extend $C^{-1/2}h$ to a measurable linear functional on $\mathcal{H}$ such that the variance of $\langle C^{-1/2}h, Y\rangle$ is $\|h\|^2$, see [Bog98]. This shows that one can 'pretend' that $C^{-1/2}y$ is an $\mathcal{H}$-valued random variable with the identity as its covariance operator, as long as one only considers

Exercise 9.6 Again in the case of a Gaussian measure on a Hilbert space, show that $\mathcal{H}_\mu$ is given by the range of $C_\mu^{1/2}$ and that $\|h\|_\mu = \|C_\mu^{-1/2}h\|$.

Consider now a general linear SPDE with additive noise on a Hilbert space $\mathcal{H}$ driven by additive noise written as an evolution equation:

$$dx = -Ax\,dt + Q\,dW(t)\;. \tag{9.4}$$

Here A is assumed to be the generator of a strongly continuous semigroup e^{-At} on $\mathcal{H}$, W is a cylindrical process on $\mathcal{H}$, and $Q\colon \mathcal{H} \to \mathcal{H}$ is a bounded operator. By the variation of constants formula, the solution to (9.4) is given by

$$x(t) = e^{-At}x_0 + \int_0^t e^{-A(t-s)}Q\,dW(s)\;.$$

In other words, the law of $x(t)$ is a Gaussian measure centred at $e^{-At}x_0$ with covariance

$$Q_t = \int_0^t e^{-A(t-s)}QQ^* e^{-A^*(t-s)}\,ds\;. \tag{9.5}$$

In view of Exercise 9.4, we obtain the following condition for a linear evolution equation on a Hilbert space to possess the strong Feller property:

Proposition 9.7 *The Markov operator $\mathcal{P}_t$ associated to (9.4) has the strong Feller property if and only if the range of e^{-At} is contained in the range of $Q_t^{1/2}$.*

Proof. Let us first show that the condition is sufficient. Denote by μ the law of the centred Gaussian measure with covariance Q_t. It then follows from Theorem 9.3 that the transition probabilities $\mathcal{P}_t(x, dy)$ have a density $p_t(x, y)$ with respect to μ given by

$$p_t(x,y) = \exp\Big(\langle Q_t^{-1/2}e^{-At}x, Q_t^{-1/2}y\rangle - \frac{1}{2}\|Q_t^{-1/2}e^{-At}x\|^2\Big) .$$

In particular, it follows that the directional derivative $D_\xi\mathcal{P}_t\varphi$ in the direction $\xi \in \mathcal{H}$ is given by

$$\begin{aligned} D_\xi\mathcal{P}_t\varphi(x) &= \int (\langle Q_t^{-1/2}e^{-At}\xi, Q_t^{-1/2}(y - e^{-At}x_0)\rangle)\varphi(y)p_t(x,y)\mu(dy) \\ &\le \sqrt{\mathcal{P}_t\varphi^2(x)}\sqrt{\int |\langle Q_t^{-1/2}e^{-At}\xi, Q_t^{-1/2}y\rangle|^2\,\mu(dy)} \\ &= \|Q_t^{-1/2}e^{-At}\xi\|^2\sqrt{\mathcal{P}_t\varphi^2(x)} . \end{aligned}$$

Here, $Q_t^{-1/2}Y$ should be interpreted as in Remark 9.5. Note that all of these calculations were formal, but can easily be made rigorous by approximation. Since $Q_t^{-1/2}e^{-At}$ is a bounded operator by assumption, the right hand side is bounded uniformly for all bounded measurable functions φ, showing that $\mathcal{P}_t\varphi$ is uniformly Lipschitz, so that $\mathcal{P}_t$ is strong Feller.

Suppose now that the range of e^{-At} is *not* contained in the range of $Q_t^{1/2}$. In this case, one can find $x_0 \in \mathcal{H}$ such that $h \equiv e^{-At}x_0$ does *not* belong to the Cameron-Martin space of μ. By Theorem 9.3, the measures $\tau_{\varepsilon h}^*\mu$ are all mutually singular, so that we can find a measurable subset $B \subset \mathcal{H}$ such that $\mu(B) = 1$ and such that $(\tau_{\varepsilon h}^*\mu)(B) = 0$ for every rational ε. This shows that $\mathcal{P}_t\mathbf{1}_B$ is equal to 1 at the origin, but equal to 0 at εx_0 for every rational ε. It is therefore discontinuous at 0, showing that $\mathcal{P}_t$ is not strong Feller. □

10 The Bismut-Elworthy-Li formula

We now finally turn to the study of a class of nonlinear stochastic PDEs. The class that we are going to consider are PDEs written in 'Da Prato - Zabczyk form' [DPZ92, DPZ96] as

$$du = -Au\,dt + F(u)\,dt + Q\,dW(t) , \qquad u \in \mathcal{H} . \tag{10.1}$$

This really stands as a shorthand for the solution to the fixed point equation

$$u(t) = e^{-At}u_0 + \int_0^t e^{-A(t-s)}F(u(s))\,ds + \int_0^t e^{-A(t-s)}Q\,dW(s) .$$

Here, we make the standing simplifying assumptions that

1. The operator A is selfadjoint and positive definite in the separable Hilbert space $\mathcal{H}$.
2. There exists $\alpha \in [0, 1)$ such that $u \mapsto A^{-\alpha}F(u)$ is a function from $\mathcal{H}$ to $\mathcal{H}$ that has bounded Fréchet derivative on bounded sets.

3. W is a cylindrical Wiener process on $\mathcal{H}$ and Q and A are such that the solution to the linearised equation (that is (10.1) where we set $F \equiv 0$) has almost surely continuous sample paths in $\mathcal{H}$.

With these assumptions, one can show that (10.1) can be solved pathwise by the usual Picard iteration procedure, see for example [Hai08]. This solution can be continued in the usual way up to a (random) explosion time τ such that $\lim_{t \nearrow \tau} \|u(t)\| = \infty$. Since we do not wish to deal with exploding solution, we assume that one can obtain an *a priori* estimate on the size of the solution such that

4. The explosion time τ is infinite almost surely.

Exercise 10.1 Given $T > 0$ and $\alpha > 0$ and a positive definite selfadjoint operator A, show that there exists a constant C such that $\|e^{-At}A^\alpha\| \leq Ct^{-\alpha}$ holds for every $t \in [0, T]$.

Exercise 10.2 Show that all the simplifying assumptions are satisfied for *reaction diffusion equations*, that is $\mathcal{H} = L^2(D, \mathbf{R}^m)$ for some smooth domain $D \subset \mathbf{R}^d$, $A = -\Delta$ endowed with Dirichlet boundary conditions, $F(u)(x) = f(u(x))$ for some globally Lipschitz continuous function $f\colon \mathbf{R}^m \to \mathbf{R}^m$, and Q is a Hilbert-Schmidt operator (that is QQ^* is trace class) on $\mathcal{H}$.

Under these assumptions, one can show that, as a straightforward consequence of the implicit functions theorem, the solution to (10.1) is Fréchet differentiable with respect to its initial condition and its derivative $J_{s,t}\xi$ in the direction $\xi \in \mathcal{H}$ satisfies the equation

$$dJ_{s,t}\xi = -AJ_{s,t}\xi\, dt + DF(u(t))J_{s,t}\xi\, dt\;. \tag{10.2}$$

Furthermore, it is possible to show in the same way that the solution is Malliavin differentiable and that its Malliavin derivative $\mathscr{D}_v u_t$ in the direction $v \in L^2([0,t], \mathcal{H})$ is given by the formula

$$\mathscr{D}_v u_t = \int_0^t J_{s,t} Q v(s)\, ds\;. \tag{10.3}$$

this allows to prove the following formula for the derivative of $\mathcal{P}_t\varphi$ [EL94]:

Theorem 10.3 (Bismut-Elworthy-Li) *Assume that the Jacobian J is such that the range of $J_{0,t}$ is contained in the range of Q for $t > 0$ and that $\mathbf{E}\|Q^{-1}J_{0,t}\|^2 < \infty$ uniformly in t over any bounded time interval bounded away from* 0*. Then, for all Fréchet differentiable test functions $\varphi\colon \mathcal{H} \to \mathbf{R}$, one has the identity*

$$D_\xi \mathcal{P}_t\varphi(u) = \frac{2}{t}\mathbf{E}\Big(\varphi(u_t)\int_{t/4}^{3t/4} \langle Q^{-1}J_{0,s}\xi, dW(s)\rangle\Big)$$

Proof. The proof works just like the proof of Hörmander's theorem, except that since Q is invertible on the range of the Jacobian, one can find explicitly a solution to (8.6). For fixed t and ξ, we write

$$v_\xi(s) = \begin{cases} \frac{2}{t}Q^{-1}J_{0,s}\xi & \text{for } s \in [\frac{t}{4}, \frac{3t}{4}] \\ 0 & \text{otherwise} \end{cases}$$

It follows immediately from (10.3) that this particular choice of v_ξ satisfies the identity $\mathscr{D}_{v_\xi} u_t = J_{0,t}\xi$, so that

$$D_\xi \mathcal{P}_t\varphi(u) = \mathbf{E}\Big(D\varphi(u_t)J_{0,t}\xi\Big) = \mathbf{E}\Big(D\varphi(u_t)\mathscr{D}_{v_\xi} u_t\Big)$$

$$= \mathbf{E}\Big(\varphi(u_t)\int \langle v_\xi(s), dW(s)\rangle\Big) ,$$

which is the desired result. □

Verifying the conditions of Theorem 10.3 is not a trivial task by far in general. However, there is a heuristic argument that allows to 'guess' the right answer in many cases of interest. It is based on the following two facts:

- The solution to (10.1) has very often the same regularity as the solution to the linear equation with $F \equiv 0$. This is because most parabolic PDEs have some smoothing property that would cause the solutions to the deterministic equation ($Q \equiv 0$) to become $\mathcal{C}^\infty$ immediately. Therefore, the driving noise is the only factor that limits the regularity of the solutions.
- The Jacobian $J_{0,t}$ typically has $1 - \alpha$ powers of A more smoothness than the solutions to (10.1). (Here, the exponent α is the one appearing in assumption 2 above.) This is about the maximal amount of regularity that one can expect from (10.2). Indeed, the variation of constants formula yields

$$J_{0,t}\xi = \int_0^t e^{-A(t-s)} DF(u(s)) J_{0,s}\xi \, ds \;.$$

Even if one assumes that $J_{0,s}\xi$ is extremely smooth, due to assumption 2 one would in general expect $DF(u)$ to have α powers of A *less* regularity than u. The convolution with the semigroup generated by A however allows to gain one power of A in terms of regularity, since the operator $A^\beta e^{-At}$ behaves like $t^{-\beta}$ for small t, so that this singularity is integrable provided $\beta < 1$.

Of course, this heuristic can be expected to hold only if the range of Q can be described as a space of functions with a given degree of regularity. This is the case for example if Q is given by a negative power of the Laplacian or some other elliptic differential operator.

Combining these facts with the expression (9.5) for the covariance of the linear equation, we deduce from these heuristic considerations that the Bismut-Elworthy-Li formula is applicable provided that the operator $Q^{-1}A^{\alpha-1}e^{-At}Q$ is Hilbert-Schmidt and that its Hilbert-Schmidt norm is square-integrable at $t \approx 0$. If Q and A commute, the borderline case for this condition occurs when $A^{\alpha-\frac{3}{2}}$ is a Hilbert-Schmidt operator.

11 The asymptotic strong Feller property

As we have seen in the previous section, the strong Feller property often fails to hold in infinite dimensions, simply because it is far 'too easy' for two measures in such spaces to be mutually singular. It would therefore be extremely convenient to have a weaker property that still allows to get an equivalent statement to Proposition 7.7. This is the idea of the *asymptotic strong Feller* property which, instead of prescribing a smoothing property at a fixed time $t > 0$, prescribes some kind of smoothing property 'at time ∞'.

Since we are interested in invariant measures that are time-invariant objects, it is reasonable to expect that such an asymptotic smoothing property is sufficient to conclude that the topological supports of distinct ergodic invariant measures are disjoint. In order to give a precise definition of the asymptotic strong Feller property, we introduce the following notation:

Definition 11.1 Given a Polish space $\mathcal{X}$ and a metric d on $\mathcal{X}$, we lift d to the corresponding Wasserstein-1 metric on the space of probability measures on $\mathcal{X}$ via the formula

$$\|\mu - \nu\|_d = \sup_{\mathrm{Lip}\,\varphi = 1} \Big(\int \varphi(x)\mu(dx) - \int \varphi(x)\nu(dx) \Big) .$$

Here, $\mathrm{Lip}\,\varphi$ denotes the best Lipschitz constant for φ, taken with respect to the metric d.

The important fact about Wasserstein distances is:

Theorem 11.2 *If d is a bounded metric that generates the topology of $\mathcal{X}$, then the corresponding Wasserstein-1 metric generates the topology of weak convergence on the space of probability measures on $\mathcal{X}$.*

It is also possible to show that with this definition, the total variation distance between two probability measures (actually half of the usual total variation distance, so that the distance between mutually singular probability measures is normalised to 1) is given by the Wasserstein-1 distance corresponding to the metric

$$d_{\mathrm{TV}}(x,y) = \begin{cases} 1 & \text{if } x \neq y, \\ 0 & \text{if } x = y. \end{cases}$$

This is a metric that totally separates all the points of our space and therefore loses completely all information about the topology of $\mathcal{X}$. It suggests the following definition, which provides one way of approximating the total variation distance between two probability measures by a sequence of Wasserstein-1 distances.

Definition 11.3 A sequence d_n of bounded continuous metrics is said to be *totally separating* if $d_n(x,y) \nearrow 1$ as $n \to \infty$ for every $x \neq y$.

With these notations at hand, we define the following notion that was introduced in [HM04]:

Definition 11.4 A Markov transition semigroup $\mathcal{P}_t$ on a Polish space $\mathcal{X}$ is asymptotically strong Feller at x if there exists a totally separating system of metrics $\{d_n\}$ for $\mathcal{X}$ and a sequence $t_n > 0$ such that

$$\inf_{U \in \mathcal{U}_x} \limsup_{n\to\infty} \sup_{y \in U} \|\mathcal{P}_{t_n}(x,\cdot\,) - \mathcal{P}_{t_n}(y,\cdot\,)\|_{d_n} = 0 , \tag{11.1}$$

It is said to have the asymptotic strong Feller property if the above property holds at every $x \in \mathcal{X}$.

Remark 11.5 If $\mathcal{B}(x,\gamma)$ denotes the open ball of radius γ centered at x in some metric defining the topology of $\mathcal{X}$, then it is immediate that (11.1) is equivalent to

$$\lim_{\gamma\to 0} \limsup_{n\to\infty} \sup_{y \in \mathcal{B}(x,\gamma)} \|\mathcal{P}_{t_n}(x,\cdot\,) - \mathcal{P}_{t_n}(y,\cdot\,)\|_{d_n} = 0 .$$

Remark 11.6 If there exists $t > 0$ such that $\mathcal{P}_t$ is continuous in the total variation topology, then this is also the case for $\mathcal{P}_s$ with all $s > t$. In this case, it is a straightforward exercise to check that the semigroup $\{\mathcal{P}_t\}$ satisfies the asymptotic strong Feller property. On the other hand, it is a known fact [DM83, Sei01] if $\mathcal{P}_t$ is strong Feller, then $\mathcal{P}_{2t}$ is continuous in the total variation topology. This shows that the asymptotic strong Feller property is a genuine generalisation of the strong Feller property.

The interest consequence of a Markov semigroup having the asymptotic strong Feller property is that one does have the following analogue of Proposition 7.7:

Theorem 11.7 *If a Markov semigroup $\{\mathcal{P}_t\}$ over a Polish space $\mathcal{X}$ is asymptotically strong Feller at $x \in \mathcal{X}$, then x cannot belong simultaneously to the topological supports of two distinct ergodic invariant measures for $\mathcal{P}_t$.*

Proof. For every measurable set A, every $t > 0$, and every metric d on $\mathcal{X}$ with $d \le 1$, the triangle inequality for $\|\cdot\|_d$ implies

$$\|\mu - \nu\|_d \le 1 - \min\{\mu(A), \nu(A)\}\Big(1 - \max_{y,z\in A} \|\mathcal{P}_t(z,\cdot) - \mathcal{P}_t(y,\cdot)\|_d\Big) . \tag{11.2}$$

To see this, set $\alpha = \min\{\mu(A), \nu(A)\}$. If $\alpha = 0$ there is nothing to prove so assume $\alpha > 0$. Clearly there exist probability measures $\bar\nu$, $\bar\mu$, ν_A, and μ_A such that $\nu_A(A) = \mu_A(A) = 1$ and such that $\mu = (1-\alpha)\bar\mu + \alpha\mu_A$ and $\nu = (1-\alpha)\bar\nu + \alpha\nu_A$. Using the invariance of the measures μ and ν and the triangle inequality implies

$$\begin{aligned}\|\mu - \nu\|_d &= \|\mathcal{P}_t\mu - \mathcal{P}_t\nu\|_d \le (1-\alpha)\|\mathcal{P}_t\bar\mu - \mathcal{P}_t\bar\nu\|_d + \alpha\|\mathcal{P}_t\mu_A - \mathcal{P}_t\nu_A\|_d \\ &\le (1-\alpha) + \alpha \int_A\int_A \|\mathcal{P}_t(z,\cdot) - \mathcal{P}_t(y,\cdot)\|_d \mu_A(dz)\nu_A(dy) \\ &\le 1 - \alpha\Big(1 - \max_{y,z\in A}\|\mathcal{P}_t(z,\cdot) - \mathcal{P}_t(y,\cdot)\|_d\Big) .\end{aligned}$$

Continuing with the proof of the corollary, by the definition of the asymptotic strong Feller property there exist constants $N > 0$, a sequence of totally separating metrics $\{d_n\}$, and an open set U containing x such that $\|\mathcal{P}_{t_n}(z,\cdot) - \mathcal{P}_{t_n}(y,\cdot)\|_{d_n} \le 1/2$ for every $n > N$ and every $y, z \in U$. (Note that by the definition of totally separating pseudo-metrics $d_n \le 1$.)

Let now μ and ν be two distinct ergodic invariant measures and assume by contradiction that $x \in \operatorname{supp}\mu \cap \operatorname{supp}\nu$ and therefore that one has $\alpha = \min(\mu(U), \nu(U)) > 0$. Taking $A = U$, $d = d_n$, and $t = t_n$ in (11.2), we then get $\|\mu - \nu\|_{d_n} \le 1 - \frac{\alpha}{2}$ for every $n > N$. On the other hand, it is possible to show that if d_n is a totally separating sequence of metrics, then $\|\mu - \nu\|_{d_n} \to \|\mu - \nu\|_{\mathrm{TV}}$, so that $\|\mu - \nu\|_{\mathrm{TV}} \le 1 - \frac{\alpha}{2}$, thus leading to a contradiction with the fact that μ and ν are mutually singular. □

A useful criterion for checking that the strong Feller property holds for a given Markov semigroup is the following:

Proposition 11.8 *Let t_n and δ_n be two positive sequences with $\{t_n\}$ increasing to infinity and $\{\delta_n\}$ converging to zero. A semigroup $\mathcal{P}_t$ on a Hilbert space $\mathcal{H}$ is asymptotically strong Feller if, for all $\varphi : \mathcal{H} \to \mathbf{R}$ with $\|\varphi\|_\infty$ and $\|D\varphi\|_\infty$ finite one has the bound*

$$\|D\mathcal{P}_{t_n}\varphi(h)\| \le C(\|h\|)(\|\varphi\|_{L^\infty} + \delta_n\|D\varphi\|_{L^\infty}) \tag{11.3}$$

for all $n > 0$ and $h \in \mathcal{H}$, where $C : \mathbf{R}_+ \to \mathbf{R}$ is a fixed non-decreasing function.

Proof. For $\varepsilon > 0$, we define on $\mathcal{H}$ the distance $d_\varepsilon(h_1, h_2) = 1 \wedge \varepsilon^{-1}\|h_1 - h_2\|_{\mathcal{H}}$, and we denote by $\|\cdot\|_\varepsilon$ the corresponding Wasserstein-1 distance. It is clear that if δ_n is any decreasing sequence converging to 0, $\{d_{\delta_n}\}$ is a totally separating system of metrics for $\mathcal{H}$.

Note now that if φ is a Fréchet differentiable function with Lipschitz constant 1 with respect to d_ε, then $\|D\varphi(h)\| \le \varepsilon^{-1}$ for every $h \in \mathcal{H}$. It therefore follows from (11.3) that for every Fréchet differentiable function φ from $\mathcal{H}$ to **R** with $\|\varphi\|_\varepsilon \le 1$ one has

$$\int_{\mathcal{H}} \varphi(h)\,(\mathcal{P}_{t_n}(h_1,dh)-\mathcal{P}_{t_n}(h_2,dh)) \le \|h_1-h_2\| C(\|h_1\|\vee\|h_2\|)\Big(1+\frac{\delta_n}{\varepsilon}\Big)\;. \tag{11.4}$$

Choosing $\varepsilon = \delta_n$, we thus obtain the bound

$$\|\mathcal{P}_{t_n}(h_1,\cdot\,) - \mathcal{P}_{t_n}(h_2,\cdot\,)\|_{\delta_n} \le 2C(\|h_1\|\vee\|h_2\|)\|h_1-h_2\|\;,$$

which in turn implies that $\mathcal{P}_t$ is asymptotically strong Feller. □

How does on go about showing such a bound in practice for an SPDE of the type (10.1)? If we were able, given any $\xi \in \mathcal{H}$, to find a control v such that

$$J_{0,t}\xi = \mathscr{A}_{0,t} v = \int_0^t J_{s,t} Q v(s)\,ds\;,$$

then we could use the integration by parts formula in order to show that $\mathcal{P}_t$ is strong Feller. Suppose now that this relation is not satisfied exactly, but that we have some 'error term' $\varrho_t = J_{0,t}\xi - \mathscr{A}_{0,t} v$ left. Then, we obtain

$$\begin{aligned} D_\xi \mathcal{P}_t\varphi &= \mathbf{E}\Big(\varphi(x_t)\int_0^t \langle v_s, dW(s)\rangle\Big) + \mathbf{E}(D\varphi(x_t)\varrho_t) \\ &\le \|\varphi\|_{L^\infty}\mathbf{E}\Big|\int_0^t \langle v_s, dW(s)\rangle\Big| + \|D\varphi\|_{L^\infty}\mathbf{E}\|\varrho_t\|\;, \end{aligned}$$

which is precisely of the required form, provided that we can choose v in such a way that $\mathbf{E}(\int_0^\infty \langle v(s), dW(s)\rangle)^2 < \infty$ and $\lim_{t\to\infty}\mathbf{E}\|\varrho_t\| = 0$.

It turns out that the asymptotic strong Feller property is satisfied by a much larger class of stochastic PDEs than the strong Feller property. The reason is that such equations typically have infinitely many 'stable directions', where the dynamic itself takes care of driving ϱ_t to 0. The control problem of finding a suitable v can then typically be reduced to a situation that is much more similar to the finite-dimensional case. For example, we have the following result:

Theorem 11.9 *In the setting of before, assume furthermore that $F\colon \mathcal{H}\to\mathcal{H}$ is globally Lipschitz with Lipschitz constant K and that the range of Q contains the linear span $\mathcal{H}_\ell$ of the eigenvalues of A with values less than $K+1$. Then, the corresponding semigroup is asymptotically strong Feller.*

Proof. Note that ϱ_t satisfies the random evolution equation

$$\frac{d\varrho_t}{dt} = -A\varrho_t + DF(u_t)\,\varrho_t + Q\,v_t\;.$$

Denote now be ϱ_t^ℓ the orthogonal projection of ϱ_t onto $\mathcal{H}_\ell$ and by ϱ_t^h the projection onto its orthogonal complement. Since the range of Q contains $\mathcal{H}_\ell$, it is straightforward to find a control v such that Qv_t only takes values in $\mathcal{H}_\ell$ and such that $\varrho_\ell(t)\equiv 0$ for all $t\ge 1$, say. It then suffices to note that one has

$$\frac{1}{2}\frac{d}{dt}\|\varrho_t^h\|^2 = -\langle\varrho_t^h, A\varrho_t^h\rangle + \langle\varrho_t^h, DF(u_t)\varrho_t^h\rangle \le -(K+1)\|\varrho_t^h\|^2 + K\|\varrho_t^h\|^2 \le -\|\varrho_t^h\|^2\;,$$

so that one has the bound $\|\varrho_t\|^2 \le \|\varrho_1\|^2 e^{-(t-1)}$. We leave it as an exercise to verify that the control v achieving this is still square integrable. □

References

[Bis81] J.-M. BISMUT. Martingales, the Malliavin calculus and hypoellipticity under general Hörmander's conditions. *Z. Wahrsch. Verw. Gebiete* **56**, no. 4, (1981), 469–505.

[Bog98] V. I. BOGACHEV. *Gaussian measures*, vol. 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1998.

[Bro28] R. BROWN. A brief account of microscopical observations made in the months of june, july and august, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Phil. Mag.* **4**, (1828), 161–173.

[DM83] C. DELLACHERIE and P.-A. MEYER. *Probabilités et potentiel. Chapitres IX à XI*. Publications de l'Institut de Mathématiques de l'Université de Strasbourg [Publications of the Mathematical Institute of the University of Strasbourg], XVIII. Hermann, Paris, revised ed., 1983. Théorie discrète du potential. [Discrete potential theory], Actualités Scientifiques et Industrielles [Current Scientific and Industrial Topics], 1410.

[DPZ92] G. DA PRATO and J. ZABCZYK. *Stochastic Equations in Infinite Dimensions*. University Press, Cambridge, 1992.

[DPZ96] G. DA PRATO and J. ZABCZYK. *Ergodicity for Infinite Dimensional Systems*, vol. 229 of *London Mathematical Society Lecture Note Series*. University Press, Cambridge, 1996.

[Ein05] A. EINSTEIN. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik* **17**, (1905), 549–560.

[EL94] K. D. ELWORTHY and X.-M. LI. Formulae for the derivatives of heat semigroups. *J. Funct. Anal.* **125**, no. 1, (1994), 252–286.

[Fed69] H. FEDERER. *Geometric measure theory*. Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York, 1969.

[Geo88] H.-O. GEORGII. *Gibbs measures and phase transitions*, vol. 9 of *de Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 1988.

[Hai08] M. HAIRER. An introduction to stochastic PDEs, 2008. URL `http://www.hairer.org/Teaching.html`. Unpublished lecture notes.

[HM04] M. HAIRER and J. C. MATTINGLY. Ergodic properties of highly degenerate 2D stochastic Navier-Stokes equations. *C. R. Math. Acad. Sci. Paris* **339**, no. 12, (2004), 879–882.

[Hör67] L. HÖRMANDER. Hypoelliptic second order differential equations. *Acta Math.* **119**, (1967), 147–171.

[Hör85] L. HÖRMANDER. *The Analysis of Linear Partial Differential Operators I–IV*. Springer, New York, 1985.

[Kry95] N. V. KRYLOV. *Introduction to the theory of diffusion processes*, vol. 142 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1995. Translated from the Russian manuscript by Valim Khidekel and Gennady Pasechnik.

[KS84] S. KUSUOKA and D. STROOCK. Applications of the Malliavin calculus. I. In *Stochastic analysis (Katata/Kyoto, 1982)*, vol. 32 of *North-Holland Math. Library*, 271–306. North-Holland, Amsterdam, 1984.

[KS85] S. KUSUOKA and D. STROOCK. Applications of the Malliavin calculus. II. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **32**, no. 1, (1985), 1–76.

[KS87] S. KUSUOKA and D. STROOCK. Applications of the Malliavin calculus. III. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.* **34**, no. 2, (1987), 391–442.

[Mal78] P. MALLIAVIN. Stochastic calculus of variations and hypoelliptic operators. *Proc. Intern. Symp. SDE* 195–263.

[Nor86] J. NORRIS. Simplified Malliavin calculus. In *Séminaire de Probabilités, XX, 1984/85*, vol. 1204 of *Lecture Notes in Math.*, 101–130. Springer, Berlin, 1986.

[Nua95] D. NUALART. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, New York, 1995.

[Øks03a] B. ØKSENDAL. *Stochastic differential equations*. Universitext. Springer-Verlag, Berlin, sixth ed., 2003. An introduction with applications.

[Øks03b] B. ØKSENDAL. *Stochastic differential equations*. Universitext. Springer-Verlag, Berlin, sixth ed., 2003. An introduction with applications.

[Ons44] L. ONSAGER. Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Phys. Rev. (2)* **65**, (1944), 117–149.

[RY99] D. REVUZ and M. YOR. *Continuous martingales and Brownian motion*, vol. 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third ed., 1999.

[Sei01] J. SEIDLER. A note on the strong Feller property, 2001. Unpublished lecture notes.

[Sin94] Y. G. SINAĬ. *Topics in ergodic theory*, vol. 44 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 1994.

[Spi76] F. SPITZER. *Principles of random walks*. Springer-Verlag, New York, second ed., 1976. Graduate Texts in Mathematics, Vol. 34.

[Wal82] P. WALTERS. *An introduction to ergodic theory*, vol. 79 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1982.