Stochastic Delay Equations

Michael Scheutzow

March 2, 2018

Note: This is a preliminary and incomplete version. Updates will appear on my homepage several times before the school starts!

Abstract

We introduce stochastic delay equations, also known as stochastic delay differential equations (SDDEs) or stochastic functional differential equations (SFDEs) driven by Brownian motion. We start with some examples. Then we prove existence and uniqueness of (strong) solutions to a large class of such equations under a monotonicity assumption on the coefficients. We then show that the solutions generate a Markov process taking values in some function space. Large parts of these notes are devoted to the question of existence and uniqueness of an invariant probability measure for this Markov process as well as the question whether in this case all transition probabilities converge to the invariant probability measure. We will also say a few words about the convergence rate. It turns out that for stochastic delay equations the right concept of convergence of transition probabilities is weak convergence rather than total variation convergence. The results in the second part of the notes are quite recent and are based on joint work with Oleg Butkovsky and Alexey Kulik. Several of these results can also be applied to other infinite dimensional Markov processes like those generated by some stochastic partial differential equations (SPDEs).

The reader is supposed to know basic concepts and results of stochastic analysis such as the Itô integral and continuous-time martingales and to know the concept of a Markov process. Knowledge of stochastic differential equations (without delay) is useful but not required.

1 Introduction

This section is supposed to be a gentle introduction to stochastic delay equations with as little theory as possible.

1.1 Existence and uniqueness of solutions: a special case

Consider the equation

$$\mathrm{d}X(t) = g(X(t-1))\,\mathrm{d}W(t), \tag{1.1}$$

where $W(t)$, $t \geq 0$ is a one-dimensional Brownian motion defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $g : \mathbb{R} \to \mathbb{R}$ is continuous. We want to solve the equation. This means that we look for a stochastic process $X(t)$, $t \geq 0$ which satisfies (almost surely)

$$X(t) = X(0) + \int_0^t g(X(s-1))\,\mathrm{d}W(s) \tag{1.2}$$

for every $t \geq 0$. In order to have any chance of obtaining a *unique* solution, we need to specify an initial condition and it is clearly not enough to specify $X(0)$ only: we need to specify $X(s)$ for all $s \in [-1, 0]$ (or some other interval of length 1). Let η be an arbitrary (deterministic) function in $\mathcal{C} := C([-1,0], \mathbb{R})$. Then we can solve equation (1.1) with initial condition η explicitly and uniquely on the interval $[0, 1]$ by

$$X(t) = \eta(0) + \int_0^t g(\eta(s-1))\,\mathrm{d}W(s).$$

The solution is a continuous process on the interval $[0, 1]$ and therefore serves as an initial condition to solve the equation on $[1, 2]$ and so on. Observe that unlike the case of stochastic differential equations without delay we do not need any growth condition on g in order to guarantee existence of a global solution (for such an equation). Neither do we require g to be locally Lipschitz continuous to ensure uniqueness of solutions (even continuity of g is not really needed but we will still assume this property to hold in the following).

1.2 Continuous dependence on the initial condition

Next, we can ask whether the unique solution X depends continuously on the initial condition η. We always equip the space $\mathcal{C}$ with the supremum norm $\|\cdot\|_\infty$. In order to answer the question, we need to clarify which topology we are working with. In order to ease the exposition, let us assume that g satisfies a global Lipschitz condition, i.e. there exists some $L \geq 0$ such that $|g(x) - g(y)| \leq L|x - y|$ for all $x, y \in \mathbb{R}$. Let us temporarily denote the unique solution of (1.2) with initial condition η by X^η and let us just look at the solution on the interval $[0, 1]$.

Proposition 1.1. *The map $\eta \mapsto X^\eta(t)$ is continuous from $\mathcal{C}$ to $L^2(\Omega, \mathcal{F}, \mathbb{P})$ for every $t \in [0, 1]$.*

Proof. Let $\eta, \phi \in \mathcal{C}$. Then

$$\begin{aligned}
&\mathbb{E}|X^\eta(t) - X^\phi(t)|^2 \\
&= \mathbb{E}\Big|\eta(0) + \int_0^t g(\eta(s-1))\,\mathrm{d}W(s) - \phi(0) - \int_0^t g(\phi(s-1))\,\mathrm{d}W(s)\Big|^2 \\
&\le 2|\eta(0) - \phi(0)|^2 + 2\mathbb{E}\Big|\int_0^t g(\eta(s-1)) - g(\phi(s-1))\,\mathrm{d}W(s)\Big|^2 \\
&= 2|\eta(0) - \phi(0)|^2 + 2\int_0^t \big|g(\eta(s-1)) - g(\phi(s-1))\big|^2\,\mathrm{d}s \\
&\le 2(1 + L^2)\|\eta - \phi\|_\infty^2,
\end{aligned}$$

so the claim follows. □

Remark 1.2. Note that we have proved more than we claimed in Proposition 1.1.

The statement of the previous proposition is certainly no surprise. The following proposition (which goes back to S. Mohammed [27]) is more surprising because there is no analog for nondelay equations. For ease of exposition we just consider the linear case $g(x) = x$ here.

Proposition 1.3. *Consider the solution X^η of (1.1) with $g(x) = x$ and initial condition $\eta \in \mathcal{C}$. For any modification of the solutions X^η, the map $\eta \mapsto X^\eta(1)$ is almost surely discontinuous from $\mathcal{C}$ to $\mathbb{R}$.*

Proof. We show that the map $\eta \mapsto X^\eta(1)$ is almost surely discontinuous at the point $\eta \equiv 0$. To see this, let $\varepsilon > 0$ and define $\eta_n^\varepsilon \in \mathcal{C}$ as $\eta_n^\varepsilon(s) := \varepsilon \sin(2\pi s n)$, $s \in [-1, 0]$, $n \in \mathbb{N}$. Then,

$$Y_n := X^{\eta_n^\varepsilon}(1) = \varepsilon \int_0^1 \sin(2\pi s n)\,\mathrm{d}W(s).$$

The sequence $Y_1, \ldots$ is i.i.d. with $\mathcal{L}(Y_1) = \mathcal{N}(0, \frac{1}{2}\varepsilon^2)$. In particular, the sequence $Y_1(\omega), Y_2(\omega), \ldots$ is almost surely unbounded and therefore the image of an arbitrary neighborhood of $\eta \equiv 0$ under the map $\phi \mapsto X^\phi(1)$ is almost surely unbounded. This means that the map $\eta \mapsto X^\eta(1)$ is almost surely discontinuous. □

1.3 Reconstruction property

Next, we discuss the so-called *reconstruction property*. Imagine that you secretly write down an initial condition $\eta \in \mathcal{C}$. Then you simulate the solution X of equation (1.1) with initial condition η and you tell me the values of $X(t)$ for $t \in [99, 100]$ say. Can I recover η without knowing the Brownian motion trajectory $t \mapsto W(t)$? For a non-degenerate SDE like $\mathrm{d}Y(t) = -Y(t)\,\mathrm{d}t + \mathrm{d}W(t)$ this is certainly impossible. We claim however that for the solution of (1.1) this is possible almost surely in case $g : \mathbb{R} \to \mathbb{R}$ is one-to-one and strictly positive. I will try to convince you of this in the lecture. Let me give you the following hint: the reconstruction uses the (local) law of the iterated logarithm for Brownian motion.

1.4 Affine equations

Next, we consider the linear (or affine) equation with *additive noise*

$$\mathrm{d}X(t) = -aX(t-1)\,\mathrm{d}t + \mathrm{d}W(t), \tag{1.3}$$

where $a > 0$ and we continue to assume that W is one-dimensional Brownian motion. The equation is not of the type (1.1) but it follows as before that for every $\eta \in \mathcal{C}$, equation (1.3) has a unique solution $X(t)$, $t \geq 0$ with initial condition η. The equation resembles the SDE

$$\mathrm{d}Y(t) = -aY(t)\,\mathrm{d}t + \mathrm{d}W(t),$$

the solution of which is called *Ornstein-Uhlenbeck process* which is a real-valued Gaussian process and which has the property that for every initial condition $y \in \mathbb{R}$, the law of the solution $Y(t)$ with initial condition y converges in the total variation norm to $\mathcal{N}(0, \frac{1}{2a})$ (the definition of total variation convergence will be given later). Is the same true for solutions to (1.3)? Well, this depends on the value of a. If a is too large, then solutions start to oscillate more and more as time increases and the law of $X(t)$ will not converge to anything in any reasonable sense when $t \to \infty$. It is not hard to show that the critical value of a is $\pi/2$: for $a \in (0, \pi/2)$, the law of $X(t)$ converges to a Gaussian distribution in the total variation norm as $t \to \infty$ for any initial condition while for $a \geq \pi/2$ this is not the case.

Generally, affine SFDEs with additive noise (also multi-dimensional) can be solved explicitly in terms of the *fundamental solution* of the homogeneous equation just like inhomogeneous affine FDEs by a variation-of-constant formula. It is apparent from that formula that the solution of such an SFDE with deterministic initial condition is a Gaussian process. More general affine SFDEs have been treated in [14], [28], and [17].

1.5 The linear chain trick

Finally, we will briefly discuss the so-called *linear chain trick* which is well-known in the mathematical literature of deterministic FDEs and which allows to convert certain classes of SFDEs into an equivalent finite system of SDEs. We start with the following example:

$$\mathrm{d}X(t) = a\Big(\int_{-\infty}^{0} X(t+s)\,\mathrm{e}^{\lambda s}\,\mathrm{d}s\Big)\,\mathrm{d}t + b\Big(\int_{-\infty}^{0} X(t+s)\,\mathrm{e}^{\mu s}\,\mathrm{d}s\Big)\,\mathrm{d}W(t), \tag{1.4}$$

where $\lambda, \mu > 0$ and W is standard Brownian motion as before. This equation has unbounded delay and therefore we should specify an initial condition η which is a continuous function from $(-\infty, 0]$ to $\mathbb{R}$ such that $\big|\int_{-\infty}^{0} \mathrm{e}^{\nu s}\,\eta(s)\,\mathrm{d}s\big| < \infty$ for $\nu \in \{\lambda, \mu\}$. Defining

$$Y(t) := \int_{-\infty}^{0} X(t+s)\,\mathrm{e}^{\lambda s}\,\mathrm{d}s, \quad Z(t) := \int_{-\infty}^{0} X(t+s)\,\mathrm{e}^{\mu s}\,\mathrm{d}s$$

we can replace (1.4) by the following equivalent system of SDEs *without* delay

$$\begin{aligned}
\mathrm{d}X(t) &= a(Y(t))\,\mathrm{d}t + b(Z(t))\,\mathrm{d}W(t)\\
\mathrm{d}Y(t) &= \big(X(t) - \lambda Y(t)\big)\,\mathrm{d}t\\
\mathrm{d}Z(t) &= \big(X(t) - \mu Z(t)\big)\,\mathrm{d}t.
\end{aligned}$$

Analyzing this system is often much easier than analyzing (1.4) directly. This *linear chain trick* works for more general SFDEs but it does not work for every SFDE. In particular it never works for SFDEs with bounded delay (unless the SFDE is actually an SDE).

Further reading: The article [27] by S. Mohammed does not only contain the above proof that the solutions of the SDDE (1.1) with $g(x) = x$ almost surely fail to depend continuously upon the initial condition. He also shows that at least there is a modification which almost surely depends *measurably* upon the initial condition (with respect to the L^2 norm), but (quite surprisingly!) any modification which depends measurably upon the initial condition is necessarily almost surely a *nonlinear* function of the initial condition. [30] contains a larger class of examples of linear SFDEs which are almost surely discontinuous with respect to the initial condition.

2 Existence and uniqueness of solutions

In this section, we will provide sufficient conditions for existence and uniqueness of strong solutions of an SFDE with bounded memory. Let $r > 0$ be the maximal delay, $m, d \geq 1$, and let $\mathcal{C} := C([-r,0], \mathbb{R}^d)$ be the space of continuous functions from $[-r, 0]$ to $\mathbb{R}$ equipped with the supremum norm $\|.\|_\infty$ and the Borel-σ-field

$\mathcal{B}(\mathcal{C})$. Recall that $\mathcal{C}$ is a *Polish* space, i.e. separable and complete with respect to $\|\cdot\|_\infty$. Let W be m-dimensional Brownian motion. Consider the equation

$$\begin{aligned} \mathrm{d}X(t) &= f(X_t)\,\mathrm{d}t + g(X_t)\,\mathrm{d}W(t),\ t \geq 0 \\ X_0 &= x, \end{aligned} \tag{2.1}$$

where $x \in \mathcal{C}$, $f : \mathcal{C} \to \mathbb{R}^d$ and $g : \mathcal{C} \to \mathbb{R}^{d\times m}$ are continuous and bounded on bounded subsets and where we used the standard notation $X_t(s) := X(t+s)$, $s \in [-r, 0]$. For $A \in \mathbb{R}^{d\times m}$, we denote the *Frobenius norm* by

$$\|A\| := \big(\mathrm{Tr}(AA^*)\big)^{1/2} = \Big(\sum_{i,j} a_{ij}^2\Big)^{1/2}.$$

Theorem 2.1. *Assume that there exists a constant K such that f, g satisfy the following one-sided Lipschitz condition:*

$$2\langle f(x) - f(y), x(0) - y(0)\rangle + \|g(x) - g(y)\|^2 \leq K\|x - y\|_\infty^2$$

for all $x, y \in \mathcal{C}$. Then, equation (2.1) *has a unique strong (global) solution X. Further, for each $t > 0$, the map $x \mapsto X_t^x$ which maps the initial condition to the solution segment at time t is continuous in probability. In fact the following holds: for every $p \in (0, 1)$ there exist universal constants $c_1(p), c_2(p)$ such that for every $T \geq 0$ and $x, y \in \mathcal{C}$ we have*

$$\mathbb{E}\sup_{0\leq t\leq T} |X^x(t) - X^y(t)|^{2p} \leq (K+1)^p\|x - y\|_\infty^{2p} c_2(p)\exp\{c_1(p)KT\},$$

where X^z denotes the solution of (2.1) *with initial condition $z \in \mathcal{C}$.*

Remark 2.2. Let us check first that the one-sided Lipschitz condition in the theorem is weaker than a Lipschitz condition. Assume that both f and g satisfy a global Lipschitz condition, i.e. there exist constants L_f and L_g such that

$$|f(x) - f(y)| \leq L_f\|x - y\|, \quad \|g(x) - g(y)\| \leq L_g\|x - y\|_\infty, \text{ for all } x, y \in \mathcal{C}.$$

Then, by the Cauchy-Schwarz inequality,

$$\begin{aligned} 2\langle f(x) - f(y), x(0) - y(0)\rangle + \|g(x) - g(y)\|^2 &\leq 2L_f\|x - y\||x(0) - y(0)| + L_g^2\|x - y\|_\infty^2 \\ &\leq \big(2L_f + L_g^2\big)\|x - y\|_\infty^2, \end{aligned}$$

so the assumption in the theorem holds for $K = 2L_f + L_g^2$.

As an example which satisfies the condition of the theorem but does not satisfy a global Lipschitz condition take $d = m = 1$, $g \equiv 0$ and $f(x) = -x(0)^3$ or $f(x) = -\mathrm{sign}(x(0))\sqrt{|x(0)|}$.

Before proving Theorem 2.1, we quote the following *Stochastic Gronwall lemma* which is proved in [34].

Proposition 2.3. *Let Z be a non-negative process with continuous paths defined on $[0,\infty)$ which is adapted with respect to a given filtration $(\mathcal{F}_t)$, $t \geq 0$ and which satisfies the inequality*

$$Z(t) \leq K \int_0^t Z^*(u)\,\mathrm{d}u + M(t) + C,$$

where $C, K \geq 0$ and M is a continuous local martingale with respect to the same filtration satisfying $M(0) = 0$ and $Z^*(u) := \sup_{0 \leq s \leq u} Z(s)$ is the running supremum of Z. Then, for each $0 < p < 1$, there exist universal finite constants $c_1(p)$, $c_2(p)$ (not depending on M, C, K or T) such that

$$\mathbb{E}(Z^*(T))^p \leq C^p c_2(p) \exp\{c_1(p) KT\} \text{ for every } T \geq 0.$$

Proof of Theorem 2.1. We skip the proof of existence of a strong solution and just mention that it can be achieved by first defining an Euler approximation ϕ^n with step size $1/n$ and then showing that this sequence - stopped when the process becomes large - is a Cauchy sequence in an appropriate complete metric space. One then has to show that the limit ϕ really solves the equation. Details of the proof can be found in [34] (under even slightly weaker assumptions), see also [40] for the somewhat easier proof in the non-delay case.

Let us now show uniqueness (which is easier than showing existence). Let X and $\bar{X}$ be two solutions with the same initial condition $x \in \mathcal{C}$. Then, using Itô's formula, we get for $t \geq 0$

$$\mathrm{d}|X(t) - \bar{X}(t)|^2 = 2\langle X(t) - \bar{X}(t), f(X_t) - f(\bar{X}_t)\rangle\,\mathrm{d}t + \|g(X_t) - g(\bar{X}_t)\|^2\,\mathrm{d}t + \mathrm{d}M(t),$$

where M is a continuous local martingale satisfying $M(0) = 0$. Using the assumption in the theorem, we obtain

$$|X(t) - \bar{X}(t)|^2 \leq K \int_0^t \|X_s - \bar{X}_s\|_\infty^2\,\mathrm{d}s + M(t).$$

Define $Z(t) := |X(t) - \bar{X}(t)|^2$. Then Proposition 2.3 implies $Z \equiv 0$, so $X \equiv \bar{X}$ almost surely.

Finally, we show the continuity property in the statement of the theorem. Let $x, y \in \mathcal{C}$ and denote the solutions starting at x and y by X^x and X^y respectively. Then we have, for $t \geq 0$,

$$\mathrm{d}|X^x(t) - X^y(t)|^2 = 2\langle X^x(t) - X^y(t), f(X_t^x) - f(X_t^y)\rangle\,\mathrm{d}t + \|g(X_t^x) - g(X_t^y)\|^2\,\mathrm{d}t + \mathrm{d}M(t),$$

where M is a continuous local martingale satisfying $M(0) = 0$. Using the one-sided Lipschitz condition in the theorem, we get

$$|X^x(t) - X^y(t)|^2 \le |x(0) - y(0)|^2 + K \int_0^t \|X_s^x - X_s^y\|_\infty^2 \, \mathrm{d}s + M(t).$$

Define $Z(t) := |X^x(t) - X^y(t)|^2$, $t \ge 0$. Then, for $t \ge 0$,

$$Z(t) \le |x(0) - y(0)|^2 + K \int_0^t Z^*(s) \, \mathrm{d}s + K\|x - y\|_\infty^2 + M(t)$$

and Proposition 2.3 implies for $T \ge 0$ and $p \in (0,1)$

$$\mathbb{E}(Z^*(T))^p \le ((K+1)\|x - y\|^2)^p c_2(p) \exp\{c_1(p)KT\},$$

so the continuity claim follows. □

Further reading: The idea to prove existence and uniqueness of solutions to an SDE which satisfies a one-sided Lipschitz condition goes back to Krylov (see [16]). An extended version of these results for SDEs can be found in [24]. [25] contains an existence and uniqueness proof for SFDEs with jumps and random and time-dependent coefficients. Existence and uniqueness of weak solutions for SFDEs with additive noise and very general drift f can be found in [36].

3 Markov property, strong Markov property and Feller property

In this section we prove the Markov and Feller property of solutions to SFDEs.

Proposition 3.1. *Let the assumptions of Theorem 2.1 be satisfied. Then the solution process X_t, $t \ge 0$ is a $\mathcal{C}$-valued Markov process. In fact, X_t, $t \ge 0$ is even a* Feller process*, i.e. the map $x \mapsto \mathbb{E}\phi(X_t^x)$ is continuous from $\mathcal{C}$ to $\mathbb{R}$ for every $t > 0$ and every bounded and continuous function ϕ from $\mathcal{C}$ to $\mathbb{R}$.*

Proof. We follow the proof of Proposition 4.1. in [2]. Fix $0 \le s \le t$ and let $h : \mathcal{C} \to \mathbb{R}$ be bounded and measurable. Our goal is to show that

$$\mathbb{E}(h(X_t)|\mathcal{F}_s) = \mathbb{E}(h(X_t)|X_s), \tag{3.1}$$

where $\mathcal{F}_s$ denotes the complete σ-field generated by $W(u)$, $0 \le u \le s$ which implies the Markov property. To establish (3.1) consider the equation

$$X^{(s,x)}(v) = x(0) + \int_s^v f(X_u^{(s,x)}) \, \mathrm{d}u + \int_s^v g(X_u^{(s,x)}) \, \mathrm{d}W(u), \; v \ge s, \; x \in \mathcal{C}. \tag{3.2}$$

It follows from the previous section that this equation has a unique strong solution and $X_t^{(s,x)}$ is $(\mathcal{G}_{s,t},\mathcal{B}(\mathcal{C}))$-measurable, where $\mathcal{G}_{s,t}$ is the complete σ-field generated by $W(u)-W(s)$, $s \le u \le t$. Introduce the function $\Phi : \mathcal{C}\times\Omega \to \mathcal{C}$, $(x,\omega) \mapsto X_t^{(s,x)}(\omega)$. We saw that for fixed $x \in \mathcal{C}$ the function $\Phi(x,\cdot)$ is $(\mathcal{G}_{s,t},\mathcal{B}(\mathcal{C}))$-measurable. The last statement in Theorem 2.1 says that the map $x \mapsto \Phi(x,\cdot)$ is continuous in probability. Since $\mathcal{C}$ is Polish, the map Φ has a modification $\tilde{\Phi}$ which is jointly measurable (see [10]). Strong uniqueness of solutions yields that

$$X_t(\omega) = \tilde{\Phi}(X_s,\omega) \text{ a.s.},$$

where we used the fact that X_s is $\mathcal{F}_s$-measurable and the σ-fields $\mathcal{F}_s$ and $\mathcal{G}_{s,t}$ are independent.

The fact that $\tilde{\Phi}$ is jointly measurable shows that $h(\tilde{\Phi}(X_s,\cdot))$ is $\sigma(X_s,\mathcal{G}_{s,t})$-measurable. Using the independence of $\mathcal{F}_s$ and $\mathcal{G}_{s,t}$ once more, we obtain

$$\mathbb{E}(h(X_t)|\mathcal{F}_s) = \mathbb{E}(h(\tilde{\Phi}(X_s,\cdot))|\mathcal{F}_s) = \mathbb{E}f(\tilde{\Phi}(x,\cdot))|_{x=X_s}$$

and

$$\mathbb{E}(h(X_t)|X_s) = \mathbb{E}(h(\tilde{\Phi}(X_s,\cdot))|X_s) = \mathbb{E}f(\tilde{\Phi}(x,\cdot))|_{x=X_s}$$

and therefore identity (3.1) holds.

The Feller property is an immediate consequence of the last statement of Theorem 2.1. □

We will not give a rigorous definition or proof of the strong Markov property which roughly speaking states that the Markov property does not only hold for fixed times but also for stopping times. A well-known theorem states that a Feller process with (right-)continuous trajectories automatically enjoys the strong Markov property (see for example [40], Theorem 4.2.5 or [35], Theorem 3.3.1) showing that the strong Markov property holds in our set-up.

4 Flows

In this section, we will provide sufficient conditions which guarantee that an SFDE generates a stochastic semi-flow (we will not discuss the related question under which an SFDE generates a *random dynamical system*, cf. the comments at the end of this section). Since by definition a semi-flow depends continuously upon the initial condition and we already know that this property fails to hold for some equations for which the coefficient in front of the noise depends on the past, we only discuss equations for which only the drift depends on the past. Let us consider

$$\mathrm{d}X(t) = f(X_t)\,\mathrm{d}t + \sum_{i=1}^{m} \sigma_i(X(t))\,\mathrm{d}W_i(t). \tag{4.1}$$

It turns out that under slight regularity assumptions one can represent the semi-flow generated by the solution X via the flow generated by the same equation but without drift (or at least without the part of the drift which depends on the past). We assume the following:

- f can be decomposed in the form $f(\eta) = H(\eta) + b(\eta(0))$, $\eta \in \mathcal{C}$.
- H is globally Lipschitz continuous, i.e. there exists L such that $|H(\eta) - H(\bar{\eta})| \le L\|\eta - \bar{\eta}\|_\infty$ for all η, $\bar{\eta} \in \mathcal{C}$.
- $b, \sigma_1, ..., \sigma_m \in C_{\mathrm{b}}^{1,\delta}$ for some $\delta \in (0,1)$.

Consider the SDE (without delay)

$$\begin{aligned} \mathrm{d}Y(t) &= b(Y(t))\,\mathrm{d}t + \sum_{i=1}^{m} \sigma_i(Y(t))\,\mathrm{d}W_i(t),\ t \ge s \\ Y(s) &= x, \end{aligned} \tag{4.2}$$

where $s \ge 0$.

We start with the following lemma which is a special case of Theorem 4.6.5 in [21].

Lemma 4.1. *There exists a process $\Psi : [0,\infty)^2 \times \mathbb{R}^d \times \Omega \to \mathbb{R}^d$ which satisfies the following:*

(i) For each $s \ge 0$, $x \in \mathbb{R}^d$, the process $\Psi_{s,t}(x,\omega)$, $t \ge s$ solves (4.2).

(ii) For each $s \ge 0$, $x \in \mathbb{R}^d$ and $\omega \in \Omega$ we have $\Psi_{s,s}(x,\omega) = x$.

(iii) The maps $(s,t,x) \mapsto \Psi_{s,t}(x,\omega)$ and $(s,t,x) \mapsto \mathrm{D}_x\Psi_{s,t}(x,\omega)$ are continuous for each $\omega \in \Omega$. Further, $\Psi_{s,t}(.,\omega)$ is a C^1-diffeomorphism for each $s,t \ge 0$ and $\omega \in \Omega$.

(iv) For each $s,t,u \ge 0$, and $\omega \in \Omega$ the following semi-flow property *holds:*

$$\Psi_{s,u}(.,\omega) = \Psi_{t,u}(.,\omega) \circ \Psi_{s,t}(.,\omega).$$

Note that by (ii) and (iv) we have $\Psi_{s,t}(.,\omega) = \big(\Psi_{t,s}(.,\omega)\big)^{-1}$.

Let $\Psi(u,x,\omega) := \Psi_{0,u}(x,\omega)$. We define the processes $\xi : [0,\infty) \times \mathbb{R}^d \times \Omega \to \mathbb{R}^d$ and $F : [0,\infty) \times \mathbb{R}^d \times \mathcal{C} \times \Omega \to \mathbb{R}^d$ by

$$\begin{aligned} \xi(u,x,\omega) &:= \Psi(u,\cdot,\omega)^{-1}(x) = \Psi_{u,0}(x,\omega), \\ F(u,x,\eta,\omega) &:= \big(\mathrm{D}_x\Psi_{s,t}(u,x,\omega)\big)^{-1} H(\eta) \end{aligned}$$

and consider the (random) equation

$$x(t,\omega) = \Psi\Big(t, \big[\xi(s,\eta(0),\omega) + \int_s^t F\big(u, \xi(u, x(u,\omega),\omega), x_u(\omega)\big)\,\mathrm{d}u\big], \omega\Big) \tag{4.3}$$

for $t \geq s$ with initial data

$$x(t,\omega) = \eta(t-s) \text{ for } t \in [s-r, s],$$

where η is an $\mathcal{F}_s$-measurable $\mathcal{C}$-valued random variable. We suppress the dependence of x on s and η for notational simplicity.

It turns out that solving (4.3) for x with initial condition η is equivalent to solving (4.1) with initial condition η at time s. Moreover, (4.3) can be solved ω-wise and the solution has very good regularity properties. At first, it is not clear whether (4.3) admits *global* solutions (which exist for all $t \geq s$). One direction of the equivalence claim can be shown as follows (see Proposition 2.2 in [4] for details and for the converse): assume that x solves (4.3) for some $s \geq 0$ and η up to some stopping time. Then x is a continuous semimartingale to which we can apply Itô's formula and obtain that x solves (4.1).

Next, one can check that on a set of full measure Ω_0, (4.3) admits a unique local solution up to the explosion time for all s and η. This solution depends continuously on (t,η) for each s and the solution enjoys the following semi-flow property for every $0 \leq s \leq t \leq u$, $\eta \in \mathcal{C}$, and $\omega \in \Omega_0$:

$$x(s,\eta,u,\omega) = x(t, x_t(s,\eta,t,\omega), u, \omega),$$

up to explosion (see Proposition 2.3 in [4] for a proof).

It is natural to ask whether the solution of (4.3) can be shown to exist globally. It is reasonable to conjecture that this is true since we assumed H to satisfy a global Lipschitz condition. It seems however that this conjecture has not been proven so far and it might even be wrong. The following sufficient conditions for (4.1) to admit a global solution have been proven in [31] (see also Proposition 2.4 in [4]).

Proposition 4.2. *Under the assumptions above, each of the following conditions guarantee that the solutions to* (4.3) *exist globally:*

(i) There exist $c \geq 0$ and $\gamma \in [0,1)$ such that $|H(\eta)| \leq c\big(1 + \|\eta\|_\infty^\gamma\big)$ for all $\eta \in \mathcal{C}$.

(ii) There exists $\beta \in (0,r)$ such that $H(\eta) = H(\bar{\eta})$ whenever $\eta|_{[-r,-\beta]} = \bar{\eta}|_{[-r,-\beta]}$.

(iii) For all $\omega \in \Omega$ and $T \in (0,\infty)$, we have that

$$\sup_{0 \leq u \leq T, x \in \mathbb{R}^d} \big\| \big(\mathrm{D}_x \Psi_{0,u}(x,\omega)\big)^{-1} \big\| < \infty$$

Note that property (iii) does not always hold under our assumptions.

The reader might ask if we really need to exclude all kind of delays in the coefficient in front of the noise if we want to have a chance that the SFDE generates a semi-flow. This is not the case. As an example, consider the scalar SFDE

$$\mathrm{d}X(t) = a(X_t)\,\mathrm{d}t + \sigma\Big(\int_{t-1}^{t} X(s)\,\mathrm{d}s\Big)\,\mathrm{d}W(t),$$

where $d = m = r = 1$. Writing $Y(t) := \int_{t-1}^{t} X(s)\,\mathrm{d}s$, $t \geq 0$, we obtain the equivalent system of SFDEs

$$\begin{aligned}\mathrm{d}X(t) &= a(X_t)\,\mathrm{d}t + \sigma(Y(t))\,\mathrm{d}W(t)\\ \mathrm{d}Y(t) &= \big(X(t) - X(t-1)\big)\,\mathrm{d}t.\end{aligned}$$

If a satisfies (ii) in Proposition 4.2 and $\sigma \in C_{\mathrm{b}}^{1,\delta}$ for some $\delta \in (0,1)$, then the previous result shows that the equation generates a stochastic semi-flow if we use the decomposition $H(\eta) := (a(\eta_1), -\eta_2(-1))$ and $b(x,y) = (0,x)$. Note that – in general – neither of the conditions in the previous proposition holds if we use the decomposition $H(\eta) = \big(a(\eta), \eta_2(0) - \eta_2(-1)\big)$, $b \equiv 0$ instead.

Remark 4.3. The reader may wonder why we need extra conditions in Proposition 4.2 to guarantee non-explosion of the flow. After all, Theorem 2.1 tells us that a global solution exists when f and g (i.e. H, b, and σ) satisfy a global Lipschitz condition. Well, there is a subtle difference between global existence of solutions and global existence of semi-flows and the former does not imply the latter even if a local semi-flow exists. The former property is ofen referred to as *weak completeness* and the latter as *strict* or *strong completeness*. [22] contains an example of a two-dimensional SDE without drift and with σ bounded and C^∞ for which strong completeness does not hold (but weak completeness does).

Further reading: The presentation of the material in this section is based on [31] and [4] (the latter being more general by also allowing for random and time-dependent coefficients and more general driving martingales). [31] and [32] focus on random dynamical systems: we show that the class of SFDEs treated above do not only generate a semi-flow but even a *random dynamical system*. [32] contains a statement about stable and unstable manifolds of the linearization of the solution around a stationary trajectory.

5 Invariant measures: existence

Let $\mathbb{T}$ be either $[0,\infty)$ or $\mathbb{N}_0$. Let X_t, $t \in \mathbb{T}$ be a Markov process taking values in a measurable space $(E,\mathcal{E})$. Whenever (E,d) is a metric space, $\mathcal{E}$ is supposed to be the Borel-σ-algebra on E. In any case we assume that all singletons in E belong to $\mathcal{E}$. We denote the space of bounded measurable functions on $(E,\mathcal{E})$ by $\mathrm{b}\mathcal{E}$. If (E,d) is metric, then $C_\mathrm{b}(E)$ denotes the space of all bounded and continuous functions on E. $\mathcal{M}_1(E)$ denotes the set of probability measures on $(E,\mathcal{E})$.

Let us briefly recall some basic concepts:

- A family P_t, $t \in \mathbb{T}$ of Markov kernels on $(E,\mathcal{E})$ is called a (normal) Markov transition function if $P_0 = \mathrm{Id}$ and $P_{t+s} = P_t \circ P_s$ for all $s,t \in \mathbb{T}$.
- To each Markov kernel K on E we associate a map from $\mathrm{b}\mathcal{E}$ to $\mathrm{b}\mathcal{E}$ by $K\varphi(x) = \int \varphi(y)K(x,\mathrm{d}y)$ (we will use the same symbol for the kernel and this map).
- To each Markov kernel K on E we associate a map K^* from $\mathcal{M}_1(E)$ to $\mathcal{M}_1(E)$ by $K^*\mu(A) = \int K(x,A)\,\mathrm{d}\mu(x)$, where $A \in \mathcal{E}$.

Definition 5.1. A probability measure μ on $(E,\mathcal{E})$ is called *invariant (probability) measure* (abbreviated *ipm*) of the Markov semi-group (P_t), $t \in \mathbb{T}$ if

$$\mu(A) = \int_E P_t(x,A)\,\mathrm{d}\mu(x) \text{ for all } t \in \mathbb{T},\ A \in \mathcal{E}.$$

Note that when we speak of an *invariant measure* then we always mean *invariant probability measure*. An important question in the theory of Markov processes is whether a given Markov process (or the associated Markov semi-group) admits an invariant measure μ and if so whether or not μ is unique. If so, then it is natural to ask whether the transition probabilities $P_t(x,.)$ converge to μ as $t \to \infty$ for every $x \in E$ (for this question to make sense we need to specify a topology on $\mathcal{M}_1(E)$). Finally, if this is true, then one can ask about the speed of convergence. In this section we treat the existence question only.

The following paragraphs including Theorem 5.3 and its proof are adapted from [5]. Missing proofs can be found there. We start with a well-known sufficient criterion for the existence of an invariant measure which is due to Krylov and Bogoliubov. We formulate the result for continuous time (it should be clear how to treat the easier discrete time case). We will assume until further notice that the state space (E,d) is Polish, i.e. separable with complete metric d. We denote the open ball with radius δ around $x \in E$ by $B(x,\delta)$.

Definition 5.2. Let (P_t) be a Markov semi-group on (E,d).

- (P_t) is called *stochastically continuous* if

$$\lim_{t\downarrow 0} P_t(x,B(x,\delta)) = 1,\ \text{for all } x \in E, \delta > 0.$$

- A stochastically continuous Markov semi-group (P_t) is called *Feller semigroup* if $P_t(C_b(E)) \subseteq C_b(E)$ holds for every $t \geq 0$.
- A Markov process (X_t) is called *Feller process* if its associated semigroup is Feller.

For a Feller semigroup (P_t) we define

$$R_T(x,B) := \frac{1}{T}\int_0^T P_t(x,B)\,\mathrm{d}t,\ x \in E,\ T > 0,\ B \in \mathcal{E},$$

(measurability of the integrand with respect to (t,x) is shown in [5]). R_T is a Markov kernel, so we can define R_T^*.

Theorem 5.3. *(Krylov-Bogoliubov). Let (P_t) be a Feller semigroup on (E,d). If, for some $\nu \in \mathcal{M}_1(E)$ and some sequence $T_n \uparrow \infty$, the sequence $R_{T_n}^*\nu$, $n \in \mathbb{N}$ is tight, then (P_t) has at least one invariant probability measure.*

Proof. Since E is Polish, Prohorov's theorem tells us that tightness of a sequence of probability measures implies the existence of some $\mu \in \mathcal{M}_1(E)$ and a subsequence of $R_{T_n}^*\nu$ which converges to μ weakly (we denote the subsequence with the same symbol for notational simplicity). Fix $v > 0$ and $\varphi \in C_b(E)$. Then $P_v\varphi \in C_b(E)$ and – abbreviating $\langle\psi,\kappa\rangle := \int \psi\,\mathrm{d}\kappa$ for $\psi \in C_b(E)$ and a finite measure κ on E –

$$\begin{aligned}
\langle\varphi, P_v^*\mu\rangle &= \langle P_v\varphi,\mu\rangle = \langle P_v\varphi, \lim_{n\to\infty} R_{T_n}^*\nu\rangle \\
&= \lim_{n\to\infty}\frac{1}{T_n}\langle P_v\varphi, \int_0^{T_n} P_s^*\nu\,\mathrm{d}s\rangle \\
&= \lim_{n\to\infty}\frac{1}{T_n}\langle \varphi, \int_v^{T_n+v} P_s^*\nu\,\mathrm{d}s\rangle \\
&= \lim_{n\to\infty}\Big(\frac{1}{T_n}\langle\varphi,\int_0^{T_n} P_s^*\nu\,\mathrm{d}s\rangle + \frac{1}{T_n}\langle\varphi,\int_{T_n}^{T_n+v} P_s^*\nu\,\mathrm{d}s\rangle - \frac{1}{T_n}\langle\varphi,\int_0^{v} P_s^*\nu\,\mathrm{d}s\rangle\Big) \\
&= \langle\varphi,\mu\rangle.
\end{aligned}$$

Since this holds for all $\varphi \in C_b(E)$ it follows that $P_v^*\mu = \mu$ and the theorem is proved. □

Note that the previous theorem shows in particular, that a Feller transition function on a *compact* state space E admits at least on invariant measure.

Unfortunately the conditions of Theorem 5.3 can rarely be checked directly in case E is non-compact. A well-known method is to construct a *Lyapunov function*.

Let us look at the special case of SFDEs and find sufficient conditions on the coefficients for which the assumptions of Theorem 5.3 are satisfied and therefore an invariant measure exists. We will always assume that the coefficients satisfy the assumptions of Theorem 2.1. We denote by B_R the closed ball in $\mathcal{C}$ around 0 with radius R. (P_t) is the Feller semigroup associated to the solution process (X_t), $t \geq 0$ of the SFDE (2.1) and $\mathbb{P}_x(\cdot)$ denotes the probability of an event subject to the condition $X_0 = x$.

The following result is a straightforward consequence of the last statement in Theorem 2.1.

Proposition 5.4. *For every $R > 0$, $\varepsilon > 0$, and $T > 0$ there exists $S \geq R$ such that*

$$\inf_{x \in B_R} \mathbb{P}_x\big(|X(t)| \leq S \text{ for all } t \in [0,T]\big) \geq 1 - \varepsilon.$$

Proposition 5.5. *Assume that there exists an initial condition $x \in \mathcal{C}$ such that*

$$\lim_{R\to\infty} \liminf_{t\to\infty} \frac{1}{t} \int_0^t P_s\big(x, B_R\big)\,\mathrm{d}s = 1.$$

Then (P_t) admits an invariant measure.

Proof. Fix $\varepsilon > 0$ and $R > 0$ such that

$$\liminf_{t\to\infty} \frac{1}{t} \int_0^t P_s\big(x, B_R\big)\,\mathrm{d}s \geq 1 - \varepsilon/2.$$

Choose a compact set C in $\mathcal{C}$ such that

$$\inf_{y \in B_R} P_r(y, C) \geq 1 - \varepsilon/2.$$

It is easy to see that such a set C exists by using the previous proposition and the fact that g is bounded on bounded subsets of $\mathcal{C}$.

Then

$$\begin{aligned} P_s(x,C) &= \mathbb{P}_x\big(X_s \in C\big) = \mathbb{P}_x\big(X_s \in C | X_{s-r} \in B_R\big)\mathbb{P}_x\big(X_{s-r} \in B_R\big) \\ &\geq \big(1 - \varepsilon/2\big)\mathbb{P}_x\big(X_{s-r} \in B_R\big). \end{aligned}$$

Hence,

$$\begin{aligned} \liminf_{t\to\infty} \frac{1}{t} \int_0^t P_s(x,C)\,\mathrm{d}s &\geq \big(1 - \varepsilon/2\big) \liminf_{t\to\infty} \frac{1}{t} \int_0^t P_{s-r}(x, B_R)\,\mathrm{d}s \\ &\geq \big(1 - \varepsilon/2\big)^2 \geq 1 - \varepsilon, \end{aligned}$$

and it follows easily that the assumptions of Theorem 5.3 are satisfied for $\nu = \delta_x$ and any sequence $T_n \to \infty$. □

Remark 5.6. The reader may wonder if the lim inf in the statement of Proposition 5.5 can be replaced by lim sup. My answer is: I don't know (if you know the answer: please tell me).

Proposition 5.7. *Assume that there exists some $T > 0$ such that*

$$\lim_{R\to\infty} \liminf_{k\to\infty} \frac{1}{k} \sum_{m=1}^{k} P_{mT}\big(x, B_R\big) = 1.$$

Then (P_t) admits an invariant measure.

Proof. Using Proposition 5.5 we see that for $\varepsilon \in (0,1)$ and $R > 0$ there exists $S \ge R$ such that

$$\inf_{x\in B_R} \mathbb{P}_x\big(|X(t)| \le S \text{ for all } t \in [0,T]\big) \ge 1 - \varepsilon.$$

This implies that for each $m \in \mathbb{N}_0$ and $s \in [mT, (m+1)T]$, we have

$$\mathbb{P}_x\big(X_s \in B_S\big) \ge (1-\varepsilon)\mathbb{P}_x\big(X_{mT} \in B_R\big)$$

and therefore

$$\liminf_{t\to\infty} \frac{1}{t} \int_0^t P_s(x, B_S)\,\mathrm{d}s \ge (1-\varepsilon) \liminf_{k\to\infty} \frac{1}{k} \sum_{m=1}^{k} P_{mT}\big(x, B_R\big)$$

so the assertion follows from the previous proposition by letting $R \to \infty$ and using the assumption in the proposition. □

The following proposition provides a useful criterion for the existence of an invariant measure of an SFDE in terms of a Lyapunov function V.

Proposition 5.8. *Assume that $V : \mathcal{C} \to [0,\infty)$ is measurable and there exist $T > 0$ and an increasing function $\varphi : [0,\infty) \to \mathbb{R}$ such that $\lim_{R\to\infty} \varphi(R) = \infty$ and $\mathbb{E}_x V(X_T) \le V(x) - \varphi(\|x\|_\infty)$. Then P admits an invariant measure.*

Proof. Note that $\varphi(0) \le 0$. For $k \in \mathbb{N}$ and $R > 0$, we have

$$\begin{aligned}
\mathbb{E}_x V(X_{kT}) - V(x) &= \sum_{m=1}^{k} \mathbb{E}_x\big(V(X_{mT}) - V(X_{(m-1)T})\big)\\
&\le \sum_{m=1}^{k} \Big(-\varphi(R)\mathbb{P}_x\big(\|X_{(m-1)T}\| > R\big) - \varphi(0)\mathbb{P}_x\big(\|X_{(m-1)T}\| \le R\big)\Big)\\
&= \sum_{m=0}^{k-1} \Big(-\varphi(R) + (\varphi(R) - \varphi(0))\mathbb{P}_x\big(X_{mT} \in B_R\big)\Big).
\end{aligned}$$

Since $V \geq 0$, we get

$$0 \leq -\varphi(R) + (\varphi(R) - \varphi(0)) \liminf_{k\to\infty} \frac{1}{k} \sum_{m=0}^{k-1} \mathbb{P}_x\big(X_{mT} \in B_R\big).$$

If R is so large that $\varphi(R) = \varphi(0) > 0$, then

$$\liminf_{k\to\infty} \frac{1}{k} \sum_{m=0}^{k-1} \mathbb{P}_x\big(X_{mT} \in B_R\big) \geq \frac{\varphi(R)}{\varphi(R) - \varphi(0)}$$

and the claim follows from Proposition 5.7 since $\lim_{R\to\infty} \varphi(R) = \infty$. □

As an application, we present the following example.

Proposition 5.9. *Consider the equation*

$$\mathrm{d}X(t) = \big(f_1(X(t)) + f_2(X_t)\big)\,\mathrm{d}t + g(X_t)\,\mathrm{d}W(t),\ X_0 = x \in \mathcal{C},$$

where (in addition to the assumptions in Theorem 2.1) f_2 and g are bounded and $\limsup_{|v|\to\infty} \langle f_1(v), v\rangle / v^2 < 0$. Then the equation admits an invariant measure.

Idea of the proof. We try the Lyapunov function $V(x) := |x(0)|^2$. Using Itô's formula and Young's inequality, we get

$$\begin{aligned} \mathrm{d}|X(t)|^2 &= 2\langle X(t), f_1(X(t))\rangle\,\mathrm{d}t + 2\langle X(t), f_2(X_t)\rangle\,\mathrm{d}t + \|g(X_t)\|^2\,\mathrm{d}t + \mathrm{d}M(t) \\ &\leq -\kappa_1|X(t)|^2\,\mathrm{d}t + \kappa_2\,\mathrm{d}t + \mathrm{d}M(t). \end{aligned}$$

where κ_1, $\kappa_2 > 0$ and M is a continuous local martingale. It is rather clear that this should imply that the assumptions of Proposition 5.5 hold (we will not check this here), but unfortunately $V(x) := |x(0)|^2$ does not satisfy the assumptions of Proposition 5.8 since $V(x) = 0$ in case $x(0) = 0$ but still $\|x\|_\infty$ may be arbitrarily large and it is impossible that $\mathbb{E}_x V(X_T) \leq V(x) - c_1$ in that case.

Alternatively, one can try $V(x) = |x(0)|^2 + \delta D(x)^2$ for some $\delta > 0$, where $D(x) := \sup_{s,t\in[-r,0]} |x(t) - x(s)|$ is the diameter of the range of $x \in \mathcal{C}$. Roughly speaking, if $\|x\|_\infty$ is large, then either $|x(0)|^2$ is large in which case $\mathbb{E}|X(r)|^2$ will be much smaller due to the strong drift towards the origin or $|x(0)|^2$ is small and $D(x)$ is large in which case $\mathbb{E}|X(r)|^2$ will not be much larger than $|x(0)|^2$ but $\mathbb{E}D(X_r)^2$ will be much smaller than $D(x)$. □

Next, we look at a more complicated example (for proofs and further results, see [2]). For $x \in \mathcal{C}$, $D(x) := \sup_{s,t\in[-r,0]} |x(t) - x(s)|$ denotes the diameter of the range of $x \in \mathcal{C}$ as in the previous example.

Proposition 5.10. *Assume that the coefficients f, g in SFDE* (2.1) *satisfy, in addition to our general hypotheses in Theorem 2.1,*

- *there exists $K \geq 0$ such that for all $x, y \in \mathcal{C}$ we have*

$$\langle f(x) - f(y), x(0) - y(0)\rangle_+ + \|g(x) - g(y)\|^2 \leq K\|x - y\|_\infty^2$$

- $\sup_{x \in \mathcal{C}} \|g^{-1}(x)\| < \infty$

- *g is globally bounded and f is sublinear, i.e. there exist $\beta \in [0,1)$ and $\bar{K} \geq 0$ such that*

$$|f(x)| \leq \bar{K}\big(1 + \|x\|_\infty^\beta\big)$$

for all $x \in \mathcal{C}$.

- *There exist constants $\alpha_1, \alpha_2 > 0$ and a function $\kappa : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\lim_{z\to\infty}\big(\kappa(z)z^{-\beta}\big) = \infty$ and*

$$\langle f(x), x(0)\rangle \leq -\alpha_1|x(0)|, \text{ for any } x \in \mathcal{C} \text{ with } D(x) \leq \kappa(|x(0)|) \text{ and } |x(0)| \geq M.$$

Then, the Markov process generated by the solution process admits an invariant measure.

Remark 5.11. In fact the conditions even guarantee uniqueness of an invariant measure and exponential convergence (in a suitable sense) of the transition probabilities to the invariant measure, so one should certainly expect that the non-degeneracy assumption on g in the theorem is not needed if one just wants to show the existence of an invariant measure. The proof of the proposition is a bit lengthy. The Lyapunov function we use is

$$V(x) := \exp\{\lambda|x(0)| + (D(x) - \gamma|x(0)|^\beta)_+\},\ x \in \mathcal{C},$$

where $\lambda, \gamma > 0$ are suitably chosen constants. Note that the first condition in the proposition is slightly stronger than the one in Theorem 2.1. It means in particular that g satisfies a global Lipschitz condition. It is not clear to me if everything remains true if the + after the $\rangle$ is deleted.

Further reading: General results on the existence of invariant probability measures can be found in the monograph [26]. [36] contains results on the existence of invariant measures for solutions of SFDEs with non-degenerate *additive* noise which are not necessarily Feller (and hence Theorem 5.3 cannot be applied). [7] contains an existence result for an invariant measure for an SFDE which is similar to the one in Proposition 5.9: f_2 and g are not necessarily bounded but linearly bounded and f_1 grows superlinearly, i.e. $\lim_{|v|\to\infty}\langle f_1(v), v\rangle/|v|^2 = -\infty$.

6 Uniqueness and stability of invariant measures: Coupling

A classical way to prove uniqueness of an invariant probability measure of a Feller semigroup is to show that the semigroup is *irreducible* and *strong Feller*. Hasminskii's theorem then states that the semigroup is t_0-*regular* for some $t_0 > 0$ and Doob's theorem states that a t_0-regular Feller semigroup can have at most one invariant probability measure μ and if μ exists, then all transition probabilities converge to μ (in total variation). We will neither explain these terms nor provide proofs and refer the interested reader to Chapter 4 of the monograph [5] instead. We point out that a rather general version of Doob's theorem (more general than the statement in [5] with a rather short and transparent proof based on coupling) has been shown in [19]. This uniqueness criterion can be applied to a large class of SFDEs but even if the noise coefficient of the SFDE is non-degenerate (and therefore the Markov semigroup is irreducible) the strong Feller property, which says that the semigroup maps $b\mathcal{E}$ to $C_b(E)$, fails to hold in many interesting examples. It fails in particular whenever the reconstruction property addressed at the beginning of these notes holds. We will therefore not pursue this approach any further. Instead, we will introduce the *coupling* technique which can be employed to show uniqueness of an invariant measure as well as convergence of transition probabilities to the invariant measure (the latter property is often referred to as *stability* or *asymptotic stability* of the invariant measure). We point out that Doeblin seems to have been the first person to employ the coupling technique to show uniqueness of an invariant measure and convergence (including rate of convergence) of the transition probabilities (see [6]). The coupling technique has been used to show convergence of transition probabilities for irreducible and aperiodic discrete time Markov chains with countable state space but has also proven useful in many other cases (we just mention the Propp-Wilson algorithm as an example).

We start by defining the two most important convergence concepts for measures. Let (E, d) be a metric space with Borel-σ-algebra $\mathcal{E}$.

Definition 6.1. For $\mu, \nu \in \mathcal{M}_1(E)$ we call

$$d_{\mathrm{TV}}(\mu,\nu) := \sup_{A\in\mathcal{E}} |\mu(A) - \nu(A)|$$

total variation distance of μ and ν.

It is easy to check that d_{TV} is a metric on $\mathcal{M}_1(E)$ which takes values in $[0,1]$. Note that $\mu, \nu \in \mathcal{M}_1(E)$ are *singular* if and only if $d_{\mathrm{TV}}(\mu,\nu) = 1$ (recall that the probability measures μ and ν are called singular if there exists some $A \in \mathcal{E}$ such that $\mu(A) = 1$ and $\nu(A) = 0$). Another possibility to define convergence in $\mathcal{M}_1(E)$ is *weak convergence*.

Definition 6.2. We will say that μ_t, $t \in \mathbb{T}$ in $\mathcal{M}_1(E)$ converges *weakly* to $\mu \in \mathcal{M}_1(E)$, if

$$\lim_{t\to\infty} \int f \,\mathrm{d}\mu_t = \int f \,\mathrm{d}\mu$$

holds for every $f \in C_\mathrm{b}(E)$. In this case we write $\mu_t \Rightarrow \mu$.

Note that the total variation distance does not depend on the topology on E. It makes sense for any measurable space $(E, \mathcal{E})$.

If the space (E, d) is separable, then one can find a metric ρ on $\mathcal{M}_1(E)$ such that for any sequence $\mu_n \in \mathcal{M}_1(E)$, $n \in \mathbb{N}_0$ and $\mu \in \mathcal{M}_1(E)$ we have $\mu_n \Rightarrow \mu$ if and only if $\lim_{n\to\infty} \rho(\mu_n, \mu) = 0$. One such metric is the L^1-Wasserstein metric.

Definition 6.3. Let (E, d) be a separable metric space and $\mu, \nu \in \mathcal{M}_1(E)$. Then the L^1-*Wasserstein distance* of μ and ν is given by

$$\rho(\mu, \nu) := \inf \int_{E\times E} d(x, y) \wedge 1 \,\mathrm{d}\xi(x, y), \tag{6.1}$$

where the infimum is taken over all $\xi \in \mathcal{M}_1(E \times E)$ such that $\xi\pi_1^{-1} = \mu$ and $\xi\pi_2^{-1} = \nu$, where $\pi_i : E \times E \to E$ is defined as $\pi_i(e_1, e_2) = e_i$, $i \in \{1, 2\}$.

A probability measure $\xi \in E \times E$ as in the previous definition is called a *coupling* of μ and ν. Note that we have $\rho(\mu, \nu) \le d_{\mathrm{TV}}(\mu, \nu)$ (this follows from (6.1) by using the estimate $d(x, y) \wedge 1 \le 1 - \delta_{x,y}$ and applying the so-called *coupling lemma* which can be found in [23], p.19), so on a separable space (E, d) total variation convergence implies weak convergence. It is easy to see that the converse is not true.

There are many examples of SFDEs for which the transition probabilities converge to the unique invariant measure in the total variation metric (we will see this below), but there are also cases, in which the best we can hope for is weak convergence. One example is the following:

$$\mathrm{d}X(t) = -X(t)\,\mathrm{d}t + g(X(t-1))\,\mathrm{d}W(t),$$

where $g : \mathbb{R} \to \mathbb{R}$ is strictly positive, bounded and bounded away from 0, strictly increasing, and C^1 with bounded derivative. We will see below that the corresponding Markov process has a unique invariant measure μ and that all transition probabilities converge to μ weakly. It is however clear that the transition probabilities cannot possibly converge to μ in total variation: just as in the example discussed in the introduction, the equation has the reconstruction property and this means in particular that for $P_t(x,.)$ and $P_t(y,.)$ are singular for any $t \geq 0$ whenever $x \neq y$. Therefore, we have $d_{\mathrm{TV}}(P_t(x,.),P_t(y,.)) = 1$ and therefore it is impossible that both $P_t(x,.)$ and $P_t(y,.)$ converge to μ in total variation.

Let us now state two propositions which provide sufficient conditions in terms of the existence of couplings for uniqueness and asymptotic stability of an invariant measure, the first for convergence with respect to d_{TV} and the second for weak convergence.

Proposition 6.4. *Let $(E,\mathcal{E})$ be a measurable space for which the diagonal $\Delta := \{(x,x) : x \in E\} \in \mathcal{E} \otimes \mathcal{E}$ and let $(X_t)_{t\in\mathbb{T}}$ be an E-valued Markov process with semigroup (P_t). Assume that for each pair $x,y \in E$, there exists a probability space $(\Omega,\mathcal{F},\mathbb{P})$ and processes (X_t^x) and (X_t^y) whose laws coincide with that of the Markov process started at x respectively y and which are coupled in such a way that $\lim_{t\to\infty} \mathbb{P}(X_t^x = X_t^y) = 1$. If there exists an invariant probability measure π, then it is unique and for each $x \in E$ we have*

$$\lim_{t\to\infty} d_{\mathrm{TV}}\big(P_t^*\delta_x, \pi\big) = 0. \tag{6.2}$$

Proof. Note that uniqueness follows from (6.2). If π is an invariant probability measure and $x \in E$, then

$$\begin{aligned}
d_{\mathrm{TV}}\big(P_t^*\delta_x, \pi\big) &= \sup_{A\in\mathcal{E}} |\pi(A) - (P_t^*\delta_x)(A)| \\
&= \sup_{A\in\mathcal{E}} \big|(P_t^*\pi)(A) - (P_t^*\delta_x)(A)\big| \\
&= \sup_{A\in\mathcal{E}} \Big| \int_E (P_t^*\delta_y)(A)\,\pi(\mathrm{d}y) - (P_t^*\delta_x)(A)\Big| \\
&= \sup_{A\in\mathcal{E}} \Big| \int_E (P_t^*\delta_y)(A) - (P_t^*\delta_x)(A)\,\pi(\mathrm{d}y)\Big| \\
&\leq \int_E \sup_{A\in\mathcal{E}} \big|(P_t^*\delta_y)(A) - (P_t^*\delta_x)(A)\big|\,\pi(\mathrm{d}y) \\
&\leq \int_E \sup_{A\in\mathcal{E}} \big|\mathbb{P}\big(X_t^y \in A\big) - \mathbb{P}\big(X_t^x \in A\big)\big|\,\pi(\mathrm{d}y) \\
&\leq \int_E \big|\mathbb{P}\big(X_t^y \neq X_t^x\big)\big|\,\pi(\mathrm{d}y)
\end{aligned}$$

so the claim follows using the assumption thanks to the dominated convergence theorem. □

We will now formulate a similar proposition which instead of total variation convergence yields only weak convergence. On the other hand, we do not require that the trajectories starting at different initial conditions couple in the sense that they become equal but we require only that they are close with high probability.

Proposition 6.5. *Let (E,d) be a Polish space and let $(X_t)_{t\in\mathbb{T}}$ be an E-valued Markov process. Assume that for each pair $x,y\in E$, there exists a probability space $(\Omega,\mathcal{F},\mathbb{P})$ and processes (X_t^x) and (X_t^y) whose laws coincide with that of the Markov process started at x respectively y and which are coupled in such a way that $d(X_t^x,X_t^y)$ converges to 0 in probability as $t\to\infty$. If there exists an invariant probability measure π, then it is unique and for each $x\in E$ we have*

$$\lim_{t\to\infty} P_t^*\delta_x = \pi,$$

where the limit is to be understood in the sense of weak convergence.

Proof. If π is an invariant probability measure, $x\in E$ and $f\in \mathrm{Lip}_b(E)$ with Lipschitz constant L, then

$$\begin{aligned}\Big|\int_E f(y)\,\pi(\mathrm{d}y)-\int_E f(y)\,\big(P_t^*\delta_x\big)(\mathrm{d}y)\Big| &= \Big|\int_E \mathbb{E}f(X_t^z)\,\pi(\mathrm{d}z)-\int_E \mathbb{E}f(X_t^x)\,\pi(\mathrm{d}z)\Big|\\ &\le \int_E \mathbb{E}\big|f(X_t^z)-f(X_t^x)\big|\,\pi(\mathrm{d}z)\\ &\le \int_E \mathbb{E}\Big(\big(Ld(X_t^z,X_t^x)\big)\wedge\big(2\|f\|_\infty\big)\Big)\,\pi(\mathrm{d}z)\end{aligned}$$

so the claim follows using the assumption thanks to the dominated convergence theorem. □

Example 6.6. This is an example of an SDDE to which the previous proposition can be applied (see [33] for more general equations for which the method works). For ease of exposition take $d=m=1$, $r=1$, $a<0$ and let $g:\mathbb{R}\to\mathbb{R}$ be bounded and Lipschitz with constant L. Consider the SFDE

$$\mathrm{d}X(t) = -a\,X(t)\,\mathrm{d}t + g(X(t-1))\,\mathrm{d}W(t).$$

We consider the solution X^x of this equation with initial condition $x\in\mathcal{C}$. Note that this defines a coupling for any pair of initial conditions $x,y\in\mathcal{C}$ (we drive X^x and X^y with the same Brownian motion). Let $Z(t):=X^x(t)-X^y(t)$. Itô's formula now shows

$$\mathrm{d}|Z(t)|^2 = -2a|Z(t)|^2\mathrm{d}t + \big(g(X_t^x)-g(X_t^y)\big)^2\mathrm{d}t + 2Z(t)\big(g(X_t^x)-g(X_t^y)\big)\,\mathrm{d}W(t).$$

Integrating the equation, we see that (since g is bounded) $\sup_{0\le t\le T}\mathbb{E}|Z(t)|^2<\infty$ for each $T>0$ and

$$\mathbb{E}|Z(t)|^2 \le |Z(0)|^2 - 2a\int_0^t \mathbb{E}|Z(s)|^2\,\mathrm{d}s + L\int_0^t \mathbb{E}|Z(s-1)|^2\,\mathrm{d}s.$$

This is an inequality for the (deterministic) function $t \mapsto \mathbb{E}|Z(t)|^2$ and it is not hard to show that $\lim_{t\to\infty} \mathbb{E}|Z(t)|^2 = 0$ in case L is sufficiently small compared to a. At this point the proof is not yet finished: we have to show that also $\lim_{t\to\infty} \mathbb{E}|Z_t|^2 = 0$ (we will not do this here). Once this is done, then it is clear that the assumptions of Proposition 6.5 hold and we obtain uniqueness of an invariant measure. Note that existence of an invariant measure follows from Proposition 5.9.

One certainly expects to have uniqueness of an invariant measure in this example in case g is non-degenerate even when the Lipschitz constant L of g is large but it is impossible to prove this fact using the coupling above: there are examples of functions g for which the difference of solutions starting at different initial conditions (and which are driven by the same Brownian motion W) does *not* converge to 0. In this case one can try a different coupling by driving equations with different initial conditions with different Brownian motions. A more systematic way to prove uniqueness of an invariant measure will be presented in the next section.

7 Uniqueness and stability of invariant measures: Generalized couplings

We saw in the previous section that even if we know that a unique invariant measure exists, it may be difficult to show uniqueness by constructing an appropriate coupling. In this section, we relax the conditions by defining the concept of a *generalized coupling*. It turns out that generalized couplings are often much easier to construct than couplings and that they still allow to conclude uniqueness and (under additional assumptions) weak convergence of transition probabilities and sometimes even rates of convergence.

Let us start with an example which is supposed to illustrate the usefulness of a generalized coupling.

Example 7.1. Consider SFDE (2.1). In addition to the assumptions in Theorem 2.1 we assume that g is non-degenerate in the sense that for every $x \in \mathcal{C}$, $g(x)$ has a right inverse $g^{-1}(x)$ such that $\sup_{x\in\mathcal{C}} \|g^{-1}(x)\| < \infty$. Consider the pair of processes

$$\begin{aligned} \mathrm{d}X(t) &= f(X_t)\,\mathrm{d}t + g(X_t)\,\mathrm{d}W(t),\ X_0 = x \\ \mathrm{d}Y(t) &= f(Y_t)\,\mathrm{d}t + \lambda(X(t) - Y(t))\,\mathrm{d}t + g(Y_t)\,\mathrm{d}W(t),\ Y_0 = y \end{aligned}$$

where $\lambda > 0$ is a constant. Note that Y does not solve the original equation. On the other hand, the additional drift $\lambda(X(t) - Y(t))$ helps to push Y to X. It is not unreasonable to hope that for $\lambda > 0$ sufficiently large, the difference $Z(t) := X(t) - Y(t)$ will almost surely converge to 0 as $t \to \infty$ (which is generally not true when $\lambda = 0$). Even though we have changed the law of Y by adding a drift we can hope that the change is not too severe. Indeed, if $\lim_{t\to\infty} Z(t) = 0$ holds, then (due to the non-degeneracy assumption on g) we may hope to be able to apply Girsanov's theorem to see that the law of $Y(t)$, $t \geq 0$ is equivalent (or at least absolutely continuous) with respect to the law of the same process with $\lambda = 0$ (i.e. the true solution). If there exist two different ergodic invariant measures μ_1 and μ_2 say, then in some sense they cannot be very close to each other and this contradicts the existence of a generalized coupling.

Let us first formalize the concept of a generalized coupling (and repeat that of a coupling).

Definition 7.2. Let $(\tilde{E}, \tilde{\mathcal{E}})$ be a measurable space and let $\mu, \nu \in \mathcal{M}_1(\tilde{E})$. Then the set

$$C(\mu,\nu) := \{\xi \in \mathcal{M}_1(\tilde{E}\times\tilde{E}) : \xi\pi_1^{-1} = \mu,\, \xi\pi_2^{-1} = \nu\}$$

is called the set of *couplings* of μ and ν and either of the sets

$$\widetilde{C}(\mu,\nu) := \{\xi \in \mathcal{M}_1(\tilde{E}\times\tilde{E}) : \xi\pi_1^{-1} \sim \mu,\, \xi\pi_2^{-1} \sim \nu\}$$
$$\widehat{C}(\mu,\nu) := \{\xi \in \mathcal{M}_1(\tilde{E}\times\tilde{E}) : \xi\pi_1^{-1} \ll \mu,\, \xi\pi_2^{-1} \ll \nu\}$$

is called set of *generalized couplings* of μ and ν.

Here, "$\ll$" means that the measure on the left hand side is absolutely continuous with respect to that on the right hand side and "$\sim$" means that both measures are equivalent, i.e. mutually absolutely continuous.

The following theorem makes the consideration in the previous example precise in a rather general set-up. It can be found in [20] and generalizes a similar result in [12]. The statement is formulated in discrete time. Note that uniqueness of the invariant measure of a continuous-time Markov process follows from the uniqueness of the invariant measure for the time-discretized Markov process, so there is no loss of generality in considering the discrete time case only.

Theorem 7.3. *Let μ_1, μ_2 be ergodic invariant measures of the Markov kernel P on the Polish space (E,d). If there exists a set $M \in \mathcal{E}\otimes\mathcal{E}$ for which $\mu_1\otimes\mu_2(M) > 0$ such that for all $(x,y) \in M$ and all $\varepsilon > 0$, there exists $\xi \in \widehat{C}(\mathbb{P}_x,\mathbb{P}_y)$ s.t.*

$$\alpha_{x,y} := \lim_{\varepsilon\downarrow 0}\limsup_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\xi\Big((\bar{x},\bar{y}) \in E^{\mathbb{N}_0}\times E^{\mathbb{N}_0} : d(\bar{x}_i,\bar{y}_i) \leq \varepsilon\Big) > 0,$$

then $\mu_1 = \mu_2$.

Proof. If $\mu_1 \neq \mu_2$, then $\mu_1 \perp \mu_2$. Choose $m \in \mathbb{N}$. Since E is Polish and every probability measure on a Polish space is known to be inner regular, there exist disjoint compact sets K_1^m, K_2^m such that $\mu_i(K_i^m) \geq 1 - 1/m$, $i = 1, 2$. Then there exists some $\varepsilon_m > 0$ such that $\varepsilon_m < \operatorname{dist}(K_1^m, K_2^m)$.

Let

$$A_i^m := \Big\{\bar{x} \in E^{\mathbb{N}_0} : \frac{1}{n}\sum_{j=0}^{n-1} 1_{K_i^m}(\bar{x}_j) \to \mu_i(K_i^m)\Big\},\ i = 1, 2.$$

Birkhoff's ergodic theorem shows that

$$\mathbb{P}_x\big(A_i^m\big) = 1 \text{ for } \mu_i\text{-almost every } x \in E, i = 1, 2,\ m \in \mathbb{N}.$$

Therefore, the set

$$M \cap \{(x, y) \in E \times E : \mathbb{P}_x(A_1^m) = 1, \mathbb{P}_y(A_2^m) = 1 \text{ for all } m \in \mathbb{N}\}$$

has positive $\mu_1 \otimes \mu_2$-measure. Fix some pair (x, y) in that intersection and let ξ^m be as in the theorem (with $\varepsilon = \varepsilon_m$). Note that the pair (x, y) does not depend on m! Since $\mathbb{P}_x(A_1^m) = 1$ and $\xi_1^m \ll \mathbb{P}_x$ (and similarly for y), we get $\xi_i^m(A_i^m) = 1$, $i = 1, 2$ and, by dominated convergence,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \xi_i^m\big(\bar{x} : \bar{x}_j \in K_i^m\big) = \mu_i\big(K_i^m\big) \geq 1 - 1/m,\ i = 1, 2. \tag{7.1}$$

By assumption, for $m \in \mathbb{N}$ we find a sequence $n_j \to \infty$ in $\mathbb{N}$ such that

$$\frac{\alpha_{x,y}}{2} < \frac{1}{n_j}\sum_{j=0}^{n_j-1} \xi^m\big((\bar{x}, \bar{y}) \in E^{\mathbb{N}_0} \times E^{\mathbb{N}_0} : d(\bar{x}_j, \bar{y}_j) \leq \varepsilon_m\big).$$

Hence, using (7.1)

$$\frac{\alpha_{x,y}}{2} < \frac{1}{n_j}\sum_{j=0}^{n_j-1} \Big(\xi_1^m\big(\bar{x}_j \notin K_1^m\big) + \xi_2^m\big(\bar{x}_j \notin K_2^m\big)\Big) \leq 2/m$$

which is a contradiction in case $m \geq \alpha_{x,y}/4$, so we have proved the statement in the theorem. □

Remark 7.4. Theorem 7.3 remains true if we relax the conditions on d as follows. We just assume that d is a positive definite and lower semi-continuous function from $E \times E$ to $[0, \infty)$ (such a function is sometimes called a *premetric*). This relaxation sometimes allows to construct generalized couplings more easily. It is easy to check that the proof still works: all we have to check is that the d-distance of two disjoint compact sets in E is strictly positive and this follows immediately from lower semi-continuity.

Remark 7.5. The result is wrong without taking averages. As an example take $E = [0,1)$, $\mu_1 = \delta_0$, $\mu_2 =$ Lebesgue measure and consider the deterministic *Baker's transform* $x \mapsto 2x \bmod 1$.

Remark 7.6. The assumptions in Theorem 7.3 are insufficient to show weak convergence of transition probabilities. As an example (with discrete time), take $E = \mathbb{N}_0$ with transition probabilities $p_{0,0} = 1$, $p_{j,j-1} = 1/3$, $p_{j,j+1} = 2/3$, $j \geq 1$. The assumptions of Theorem 7.3 are satisfied with $M = E \times E$ (and hence the invariant measure δ_0 is unique) but clearly transition probabilities starting from $j \geq 1$ do not converge to δ_0 (neither in total variation nor weakly).

Next, we address the question of weak convergence of transition probabilities. [12] contains general results in this direction which are then applied to SFDEs (without using generalized couplings). Here, we present the approach in [20] which uses generalized couplings. We just present the statements and basic ideas and refer the reader to [20] for proofs. We will use $d(\mu,\nu)$ (rather than $\rho(x,y)$ to denote the Wasserstein distance associated to d which was introduced in Definition 6.3.

Theorem 7.7. *Let P be Feller. Let μ be an ergodic invariant measure and $\mu \otimes \mu(M) = 1$. If for each $(x,y) \in M$*

$$\sup_{\xi \in \widehat{C}(\mathbb{P}_x, \mathbb{P}_y)} \lim_{\varepsilon \downarrow 0} \liminf_{n \to \infty} \xi\Big(d(X_n, Y_n) \leq \varepsilon\Big) > 0,$$

then

$$\mu\Big(x : d\big(P_n(x,.), \mu\big) > \varepsilon\Big) \to 0,\ n \to \infty,\ \varepsilon > 0.$$

We can give a stronger conclusion if we assume more regularity on the semigroup.

Definition 7.8. X is called an *e-chain* w.r.t. the metric d if for any $x \in E$, $\varepsilon > 0$ there exists $\delta > 0$ such that

$$d(P_n(x,.), P_n(y,.)) \leq \varepsilon,\ n \in \mathbb{N}_0,\ d(x,y) < \delta.$$

Proposition 7.9. *If, in addition, X is an* e-chain*, then*

$$P_n(y,.) \Rightarrow \mu \text{ for } \mu\text{-a.a. } y \in E.$$

Example 7.10. $E = \{0\} \cup \{2^{-k}, k \in \mathbb{N}_0\}$. Consider the deterministic map

$$\ldots \mapsto \frac{1}{8} \mapsto \frac{1}{4} \mapsto \frac{1}{2} \mapsto 1 \mapsto 0 \mapsto 0.$$

The chain is Feller (with respect to the Euclidean distance). It is *not* an e-chain, but satisfies all other previous assumptions and conclusions. The chain does not satisfy the *asymptotic strong Feller property* introduced by Hairer and Mattingly (for those who know this concept).

The next theorem provides sufficient conditions for the convergence of the transition probabilities for a *given* initial condition x.

Proposition 7.11. *Under the conditions of Theorem 7.7, the chain X is* mixing *w.r.t. $\mathbb{P}_\mu$, i.e. $\mathbb{E}_\mu\big(f(X_0)g(X_n)\big) \to \mathbb{E}_\mu f(X_0)\mathbb{E}_\mu g(X_0)$ for all $f,g \in \mathrm{b}\mathcal{E}$.*

Theorem 7.12. *Assume that μ is an ergodic invariant measure such that X is mixing. Fix $x \in E$ and assume that there exists $M \in \mathcal{E}$ such that $\mu(M) > 0$ and for every $y \in M$ there exists $\xi \in \widehat{C}(\mathbb{P}_x, \mathbb{P}_y)$ such that $\pi_1(\xi) \sim \mathbb{P}_x$ and*

$$\lim_{n\to\infty} \xi\big(d(X_n, Y_n) \le \varepsilon\big) = 1$$

for every $\varepsilon > 0$. Then $P_n(x,.) \Rightarrow \mu$.

Remark 7.13. In previous theorem, the condition $\pi_1(\xi) \sim \mathbb{P}_x$ cannot be dropped! We take the same example as in Remark 7.6: $E = \{0,1,2,...\}$, $p_{0,0} = 1$, $p_{i,i-1} = 1/3$, $p_{i,i+1} = 2/3$, $i = 1,2,....$

For each $i \neq 0$ there exists $\xi \in \widehat{C}(\mathbb{P}_i, \mathbb{P}_0)$ such that $\pi_2(\xi) = \mathbb{P}_0$ and the trajectories a.s. meet after finite time. Nevertheless, the transition probabilities from $i \neq 0$ do not converge to $\mu = \delta_0$.

Remark 7.14. All of the previous results (uniqueness and convergence of all transition probabilities) can be applied to SFDEs which satisfy the non-degeneracy assumption in Example 7.1.

Idea of proof of Theorem 7.7. **Step 1:** Show that the assumption in terms of $\xi \in \widehat{C}(\mathbb{P}_x, \mathbb{P}_y)$ implies

$$\lim_{\varepsilon\downarrow 0} \liminf_{n\to\infty} \sup_{\xi\in C(\mathbb{P}_x,\mathbb{P}_y)} \xi\big(d(X_n, Y_n) \le \varepsilon\big) > 0.$$

(observe the different order of lim, lim inf, and sup compared to the statement of the theorem!)

Step 2: Define

$$\gamma_{x,y}^{n,\varepsilon} := \sup_{\xi\in C(\mathbb{P}_x,\mathbb{P}_y)} \xi\big(d(X_n, Y_n) \le \varepsilon\big),\ \Gamma^{n,\varepsilon} := \int \gamma_{x,y}^{n,\varepsilon}\, \mu(\mathrm{d}x)\mu(\mathrm{d}y).$$

Show that $\lim_{n\to\infty} \Gamma^{n,\varepsilon} = 1$ for each $\varepsilon > 0$.

Then it follows that

$$\mu\big(x : d\big(P_n(x,.), \mu\big) > \varepsilon\big) \to 0,\ n \to \infty,\ \varepsilon > 0.$$

□

8 Exponential growth rate of a simple linear SDDE

This section is based on the article [38] to which we refer the reader for full proofs. We go back to the linear equation in the first section, namely

$$\mathrm{d}X(t) = X(t-1)\,\mathrm{d}W(t), \tag{8.1}$$

where W is one-dimensional Brownian motion. For an initial condition $\eta \in \mathcal{C}$ we denote the solution by X^η. Let

$$\Lambda(\eta,\omega) := \limsup_{t\to\infty} \frac{1}{t} \log \|X_t^\eta\|_\infty$$

We ask:

- Is it true that for every $\eta \not\equiv 0$, $\Lambda(\eta,.)$ is almost surely a limit rather than just a limsup?
- If so, does there exist a deterministic number Λ such that for every $\eta \not\equiv 0$ we have $\Lambda(\eta,.) = \Lambda$ almost surely?

We will give a positive answer to both questions by applying the previous results about uniqueness of an invariant measure.

We start by looking at a similar but easier problem which has been studied several decades ago. Consider a linear SDE of the form

$$\mathrm{d}X(t) = A\,X(t)\,\mathrm{d}t + \sum_{i=1}^{m} B_i X(t)\,\mathrm{d}W_i(t), \tag{8.2}$$

where $W_1,...,W_m$ are independent standard Brownian motions and A, $B_1,...,B_m \in \mathbb{R}^{d\times d}$ are given matrices. If the coefficients are sufficiently non-degenerate then we expect that there exists some $\Lambda \in \mathbb{R}$ such that $\lim_{t\to\infty} \frac{1}{t}\log|X_t^x| = \Lambda$ almost surely for each $x \in \mathbb{R}\backslash\{0\}$. The following classical way to prove such a result goes back to Furstenberg and Hasminskii: project the solution to equation (8.2) onto the (Euclidean) unit sphere in $\mathbb{R}^d$. Due to the linearity of the equation the projected process ξ is also a Markov process, in fact even a Feller process. Since the state space S^{d-1} of ξ is compact, it follows from Theorem 5.3 that ξ has at least one invariant measure μ (this is true even if all B_i are 0). Using Itô's formula, one can see that the radial part $|X^x(t)|$ of the solution starting at $x \in \mathbb{R}^d\backslash\{0\}$ has a representation of the form

$$|X^x(t)| = |x| \exp\Big\{ \int_0^t q_0(\xi(s,x(0)/|x(0)|)\,\mathrm{d}s + \sum_{i=1}^{m} \int_0^t q_i(\xi(s,x(0)/|x(0)|)\,\mathrm{d}W_i(s) \Big\}$$

for some explicit continuous functions q_i, $i = 0, \cdots, m$. If the invariant measure μ is ergodic, then Birkhoff's ergodic theorem shows that

$$\lim_{t \to \infty} \frac{1}{t} \log |X^x(t)| = \int_{S^{d-1}} q_0(y) \, \mathrm{d}\mu(y)$$

for μ-almost all $x \in S^{d-1}$ which is often called *Furstenberg-Hasminskii formula*. If μ is unique then it follows that the formula holds for every $x \in S^{d-1}$ ([13] provides conditions on the matrices for this to hold).

It is reasonable to hope that a corresponding result should hold for equation (8.1). Since the state space of this equation is the infinite dimensional space $\mathcal{C}$, we should project the solution process onto the unit sphere of $\mathcal{C}$ or some other infinite dimensional normed space. This unit sphere is not compact in the norm topology, so even existence of an invariant measure is not automatic.

It turns out, that ...

References

[1] O. Butkovsky (2014). Subgeometric rates of convergence of Markov processes in the Wasserstein metric. *Ann. Appl. Probab.*, **24** 526–552.

[2] O. Butkovsky, M. Scheutzow (2017). Invariant measures for stochastic functional differential equations. *Electronic J. Probab.*, **22** 1–23.

[3] O. Butkovsky, A. Kulik, M. Scheutzow (2018). Generalized couplings and rates of convergence. Preprint.

[4] I. Chueshov, M. Scheutzow (2013). Invariance and monotonicity for stochastic delay differential equations. *Discrete Contin. Dyn. Syst. Ser. B*, **18** 1533–1554.

[5] G. Da Prato, J. Zabczyk (1996). *Ergodicity for Infinite Dimensional Systems*, Cambridge University Press, Cambridge.

[6] W. Doeblin (1940). Elément d'une théorie générale des chaînes simples constantes de Markoff. *Ann. Sci. ENS*, **57** 61–111.

[7] A. Es-Sarhir, O. v. Gaans, M. Scheutzow (2010). Invariant measures for stochastic functional differential equations with superlinear drift. *Differential and Integral Equations*, **23** 189–200.

[8] A. Es-Sarhir, M. v. Renesse, M. Scheutzow (2009). Harnack inequality for functional SDEs with bounded memory. *Electronic Comm. Probab.*, **14** 560–565.

[9] A. Es-Sarhir, M. Scheutzow, J. Toelle, O. v. Gaans (2013). Invariant measures for monotone SPDE's with multiplicative noise term. *Appl. Math. Optim.*, **68** 275–287.

[10] D. Gusak, A. Kukush, A. Kulik, Y. Mishura, A. Pilipenko (2010). *Theory of Stochastic Processes: With Application to Financial Mathematics and Risk Theory*. Springer, New York.

[11] M. Hairer, J. Mattingly (2011). Yet another look at Harris' ergodic theorem for Markov chains. In: Seminar on Stochastic Analysis, Random Fields and Applications VI. Progress in Probability 63 109–117. Birkhäuser, Basel.

[12] M. Hairer, J. Mattingly, M. Scheutzow (2011). Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations. *Probab. Theory Related Fields*, **149** 223–259.

[13] R.Z. Hasminskii (1967). Necessary and sufficient conditions for asymptotic stability of linear stochastic systems. *Theory Probability Appl.*, **12** 144–147.

[14] K. Itô, M. Nisio (1964). On stationary solutions of a stochastic differential equation. *J. Math.Kyoto Univ.*, **4** 1–75.

[15] T. Komorowski, S. Peszat, D. Szarek (2010). On ergodicity of some Markov processes. *Ann. Probab.*, **38** 1401–1443.

[16] N. Krylov (1999). On Kolmogorov's equations for finite dimensional diffusions. In: Stochastic PDE's and Kolmogorov Equations in Infinite Dimensions (Cetaro, 1998). LNM 1715, pp. 1–63 . Springer, Berlin.

[17] U. Küchler, B. Mensch (1992). Langevin stochastic differential equation extended by a time-delayed term. *Stochastics*, **40** 23–42.

[18] A. Kulik (2018). *Ergodic Behavior of Markov Processes*. De Gruyter, Berlin.

[19] A. Kulik, M. Scheutzow (2015). A coupling approach to Doob's theorem. *Atti Accad. Naz. Lincei Rend. Lincei Mat. Appl.*, **26** 83–92.

[20] A. Kulik, M. Scheutzow (2018). Generalized couplings and convergence of transition probabilities. *Probab. Theory Related Fields*, to appear.

[21] H. Kunita (1990). *Stochastic Flows and Stochastic Differential Equations.* Cambridge University Press, Cambridge.

[22] X-M. Li, M. Scheutzow (2011). Lack of strong completeness for stochastic flows. *Ann. Probab.*, **39** 1407–1421.

[23] T. Lindvall (1992). *Lectures on the Coupling Method.* Wiley, New York.

[24] W. Liu, M. Röckner (2015). *Stochastic Partial Differential Equations: An Introduction.* Springer, Heidelberg.

[25] S. Mehri, M. Scheutzow, W. Stannat, B. Zangeneh (2018). Propagation of chaos for stochastic spatially structured neuronal networks with fully path dependent delays and monotone coefficients driven by jump diffusion noise. Preprint.

[26] S. Meyn, R. Tweedie (1993). *Markov Chains and Stochastic Stability.* Springer, New York.

[27] S. Mohammed (1986). Nonlinear flows of stochastic linear delay equations. *Stochastics*, **17** 207–213.

[28] S. Mohammed, M. Scheutzow (1990). Lyapunov exponents and stationary solutions for affine stochastic delay equations. *Stochastics*, **29** 259–283.

[29] S. Mohammed, M. Scheutzow (1996). Lyapunov exponents of linear stochastic functional differential equations driven by semimartingales. Part I: The multiplicative ergodic theory. *Ann. Inst. Henri Poincaré*, **32** 69–105.

[30] S. Mohammed, M. Scheutzow (1997). Lyapunov exponents of linear stochastic functional differential equations. Part II: Examples and case studies. *Ann. Probab.*, **25** 1210–1240.

[31] S. Mohammed, M. Scheutzow (2003). The stable manifold theorem for nonlinear stochastic systems with memory I: Existence of the semiflow. *Journal of Functional Analysis*, **205** 271–305.

[32] S. Mohammed, M. Scheutzow (2004). The stable manifold theorem for nonlinear stochastic systems with memory II: The local stable manifold theorem. *Journal of Functional Analysis*, **206** 253–306.

[33] M. Reiss, M. Riedle, O. van Gaans (2006). Delay differential equations driven by Lévy processes: stationarity and Feller properties. *Stoch. Proccess. Appl.*. **116** 1409–1432.

[34] M. von Renesse, M. Scheutzow (2010). Existence and uniqueness of solutions of stochastic functional differential equations. *Random Operators and Stochastic Equations*, **18** 267–284.

[35] D. Revuz, M. Yor (2013). *Continuous Martingales and Brownian Motion.* Springer, Heidelberg.

[36] M. Scheutzow (1984). Qualitative behaviour of stochastic delay equations with a bounded memory. *Stochastics*, **12** 41–80.

[37] M. Scheutzow (2005). Exponential growth rates for stochastic delay differential equations, *Stochastics and Dynamics*, **5** 163–174.

[38] M. Scheutzow (2013). Exponential growth rate for a singular linear stochastic delay differential equation. *Discrete Contin. Dyn. Syst. Ser. B*, **18** 1533–1554.

[39] M. Scheutzow (2013). A stochastic Gronwall lemma. *Infin. Dimens. Anal. Quantum Probab. Relat. Top.*, **16** 4p.

[40] M. Scheutzow (2018). *Stochastic Processes II, WT3.* Lecture Notes. URL: http://page.math.tu-berlin.de/~scheutzow/