

# Asymptotic coupling and a general form of Harris' theorem with applications to stochastic delay equations

October 21, 2009

**M. Hairer<sup>1,2</sup>, J. C. Mattingly<sup>3</sup>, M. Scheutzow<sup>4</sup>**

<sup>1</sup> Mathematics Institute, The University of Warwick, Coventry CV4 7AL, UK  
Email: M.Hairer@Warwick.ac.uk

<sup>2</sup> Courant Institute, New York University, New York NY 10012, USA  
Email: Martin.Hairer@courant.nyu.edu

<sup>3</sup> Department of Mathematics, Duke University, Durham NC 27708, USA  
Email: jonm@math.duke.edu

<sup>4</sup> Institut für Mathematik, Sekr. MA 7-5, Fakultät II – Mathematik und Naturwissenschaften, TU Berlin, Str. des 17. Juni 136, 10623 Berlin, Germany  
Email: ms@math.tu-berlin.de

## Abstract

There are many Markov chains on infinite dimensional spaces whose one-step transition kernels are mutually singular when starting from different initial conditions. We give results which prove unique ergodicity under minimal assumptions on one hand and the existence of a spectral gap under conditions reminiscent of Harris' theorem.

The first uses the existence of couplings which draw the solutions together as time goes to infinity. Such "asymptotic couplings" were central to [EMS01, Mat02b, Hai02, BM05] on which this work builds. As in [BM05] the emphasis here is on stochastic differential delay equations.

Harris' celebrated theorem states that if a Markov chain admits a Lyapunov function whose level sets are "small" (in the sense that transition probabilities are uniformly bounded from below), then it admits a unique invariant measure and transition probabilities converge towards it at exponential speed. This convergence takes place in a total variation norm, weighted by the Lyapunov function.

A second aim of this article is to replace the notion of a "small set" by the much weaker notion of a " $d$ -small set," which takes the topology of the underlying space into account via a distance-like function  $d$ . With this notion at hand, we prove an analogue to Harris' theorem, where the convergence takes place in a Wasserstein-like distance weighted again by the Lyapunov function.

This abstract result is then applied to the framework of stochastic delay equations. In this framework, the usual theory of Harris chains does not apply, since there are natural examples for which there exist *no* small sets (except for sets consisting of only one point). This gives a solution to the long-standing open problem of finding natural conditions under which a stochastic delay equation admits at most one invariant measure and transition probabilities converge to it.

*Keywords:* Stochastic delay equation, invariant measure, Harris' theorem, weak convergence, spectral gap, asymptotic coupling.

## 1 Introduction

There are many Markov chains on infinite dimensional spaces whose one-step transition kernels are mutually singular when starting from different initial conditions. Many standard techniques used in the study of Markov chains as exposed for example in [MT93] can not be applied to such a singular setting. In this article, we provide two sets of results which can be applied to general Markov processes even in such a singular setting. The first set of results gives minimal, verifiable conditions which are equivalent to the existence of at most one invariant measure. The second set of results gives a general form of Harris' theorem which proves the existence of a spectral gap under the existence of a Lyapunov function and a modified "small set" condition.

The study of the ergodic theory for stochastic partial differential equations (SPDEs) has been one of the principal motivations to develop this theory. While even simple, formally elliptic, linear SPDEs can have transition probabilities which are mutually singular, the bulk of recent work has been motivated by equations driven by noise which is "degenerate" to varying degrees [EH01, BKL01, EMS01, KS02, Mat02b, HM06, HM08a]. The current article focuses on stochastic delay differential equations (SDDEs) and makes use of the techniques developed in the SPDE context. That the SPDE techniques are applicable to the SDDE setting is not surprising since [EMS01] reduced the original SPDE, the stochastic Navier-Stokes equations, to an SDDE to prove unique ergodicity. In [BM05], the same ideas were applied directly to SDDEs. There the emphasis was on additive noise, here we generalize the results to the setting of state dependent noise. The works [EMS01, Mat02b, Hai02, BM05] all share the central idea of using a shift in the driving Wiener process to force solutions starting at different initial conditions together asymptotically as time goes to infinity. In [EMS01, Mat02b, BM05], the asymptotic coupling was achieved by driving a subset of the degrees of freedom together in finite time. Typically these were the dynamically unstable directions, which ensured the remaining degrees of freedom would converge to each other asymptotically. In [Hai02, HM06] the unstable directions were only stabilized sufficiently by shifting the driving Wiener processes to ensure that all of the degrees of freedom converged together asymptotically. This broadens the domain of applicability and is the tact taken in Section 2 to prove a very general theorem which gives verifiable conditions which are equivalent to unique ergodicity. In particular, this result applies to the setting when the transition probabilities are mutually singular for many initial conditions.

A simple, instructive example which motivates our discussion is the following SDDE:

$$dX(t) = -cX(t)dt + g(X(t-r))dW(t), \quad (1.1)$$

where  $r > 0$ ,  $W$  is a standard Wiener process,  $c > 0$ , and  $g : \mathbf{R} \rightarrow \mathbf{R}$  is a strictly positive, bounded and strictly increasing function. This can be viewed as a Markov process  $\{X_t\}_{t \geq 0}$  on the space  $\mathbf{X} = \mathcal{C}([-r, 0], \mathbf{R})$  which possesses an invariant measure for sufficiently large  $c$ . However, in this particular case, given the solution  $X_t$  for any  $t > 0$ , the initial condition  $X_0 \in \mathbf{X}$  can be recovered *with probability one*, exploiting the law of the iterated logarithm for Brownian motion (see [Sch05], Section 2). Thus, if the initial conditions in  $\mathcal{C}([-r, 0], \mathbf{R})$  do not agree, then the transition probabilities for any step of this chain are always mutually singular. In particular, the corresponding Markov semigroup does not have the strong Feller property and, even worse, the only "small sets" for this system are those consisting of one single point. The results in Section 2 nevertheless apply and allow us to show that (1.1) can have at most one invariant measure and that convergence toward it happens at exponential rate.

While the main application considered in this article is that of stochastic delay equations, the principal theorems are also applicable to a large class of stochastic PDEs driven by degenerate noise. In particular, Theorem 5.4 in [HM08c] yields a very large class of degenerate SPDEs (essentially semilin-

ear SPDEs with polynomial nonlinearities driven by additive noise, satisfying a Hörmander condition) for which it is possible to find a contracting distance  $d$ , see Section 5.3 below.

### 1.1 Overview of main results

We now summarise the two principal results of this article. The first is an abstract ergodic theorem which is useful in a number of different settings and gives conditions equivalent to unique ergodicity. The second result gives a general form of Harris' theorem which ensures the existence of a spectral gap if there exists an appropriate Lyapunov function.

#### 1.1.1 Asymptotic coupling and unique ergodicity

Let  $\mathbf{X}$  be a Polish space with metric  $d$  and let  $\mathbf{X}^\infty = \mathbf{X}^{\mathbb{N}_0}$  be the associated space of one-sided infinite sequences. Given a Markov transition kernel  $\mathcal{P}$  on  $\mathbf{X}$ , we will write  $\mathcal{P}_{[\infty]} : \mathbf{X} \rightarrow \mathcal{M}(\mathbf{X}^\infty)$  as the probability kernel defined by stepping with the Markov kernel  $\mathcal{P}$ . Here  $\mathcal{M}(\mathbf{X}^\infty)$  is the space of probability measures on  $\mathbf{X}^\infty$ . If  $\mu$  is a probability measure on  $\mathbf{X}$ , then we write  $\mathcal{P}_{[\infty]}\mu$  for the measure in  $\mathcal{M}(\mathbf{X}^\infty)$  defined by  $\int_{\mathbf{X}} \mathcal{P}_{[\infty]}(x, \cdot) \mu(dx)$ .

In general, we will denote by  $\mathcal{M}(\mathbf{Y})$  the set of probability measures over a Polish space  $\mathbf{Y}$ . Given  $\mu_1, \mu_2 \in \mathcal{M}(\mathbf{Y})$ ,  $\mathcal{C}(\mu_1, \mu_2)$  will denote the set of all couplings of the two measures. Namely,

$$\mathcal{C}(\mu_1, \mu_2) = \left\{ \Gamma \in \mathcal{M}(\mathbf{Y} \times \mathbf{Y}) : \Pi_{\#}^{(i)} \Gamma = \mu_i \text{ for } i = 1, 2 \right\},$$

where  $\Pi^{(i)}$  is the projection defined by  $\Pi^{(i)}(y_1, y_2) = y_i$  and  $f_{\#}\mu$  is the push-forward of the measure  $\mu$  defined by  $(f_{\#}\mu)(A) = \mu(f^{-1}(A))$ . We define the *diagonal at infinity*

$$\mathcal{D} = \left\{ (x^{(1)}, x^{(2)}) \in \mathbf{X}^\infty \times \mathbf{X}^\infty : \lim_{n \rightarrow \infty} d(x_n^{(1)}, x_n^{(2)}) = 0 \right\}$$

as the set of paths which converge to each other asymptotically. Given two measures  $m_1$  and  $m_2$  on  $\mathbf{X}^\infty$ , we say that  $\Gamma \in \mathcal{C}(m_1, m_2)$  is an *asymptotic coupling* of  $m_1$  and  $m_2$  if  $\Gamma(\mathcal{D}) = 1$ .

It is reasonable to expect that if two invariant measure  $\mu_1$  and  $\mu_2$  are such that there exists an asymptotic coupling of  $\mathcal{P}_{[\infty]}\mu_1$  and  $\mathcal{P}_{[\infty]}\mu_2$  then in fact  $\mu_1 = \mu_2$ . We will see that on the infinite product structure a seemingly weaker notion is sufficient to prove  $\mu_1 = \mu_2$ .

To this end we define

$$\tilde{\mathcal{C}}(\mu_1, \mu_2) = \left\{ \Gamma \in \mathcal{M}(\mathbf{Y} \times \mathbf{Y}) : \Pi_{\#}^{(i)} \Gamma \ll \mu_i \text{ for } i = 1, 2 \right\}, \quad (1.2)$$

where  $\Pi_{\#}^{(i)} \Gamma \ll \mu_i$  means that  $\Pi_{\#}^{(i)} \Gamma$  is absolutely continuous with respect to  $\mu_i$ .

To state the main results of this section, we recall that an invariant measure  $\mu$  for  $\mathcal{P}$  is said to be ergodic if for any invariant  $\varphi : \mathbf{X} \rightarrow \mathbf{R}$  ( $\mathcal{P}\varphi = \varphi$ ),  $\varphi$  is  $\mu$ -almost surely constant.

**Theorem 1.1** *Let  $\mathcal{P}$  be a Markov operator on a Polish space  $\mathbf{X}$  admitting two ergodic invariant measures  $\mu_1$  and  $\mu_2$ . The following statements are equivalent:*

1.  $\mu_1 = \mu_2$ .
2. There exists an asymptotic coupling of  $\mathcal{P}_{[\infty]}\mu_1$  and  $\mathcal{P}_{[\infty]}\mu_2$ .
3. There exists  $\Gamma \in \tilde{\mathcal{C}}(\mathcal{P}_{[\infty]}\mu_1, \mathcal{P}_{[\infty]}\mu_2)$  such that  $\Gamma(\mathcal{D}) > 0$ .

**Remark 1.2** In (1.2), we could have replaced absolute continuity by equivalence. If we have a ‘coupling’  $\Gamma$  satisfying the current condition, the measure  $\frac{1}{2}(\Gamma + \mathcal{P}_{[\infty]\mu_1} \otimes \mathcal{P}_{[\infty]\mu_2})$  satisfies that stronger condition.

**Remark 1.3** At first, it might seem surprising that it is sufficient to have a measure which couples asymptotically which is only *equivalent* and not *equal* to the law of the Markov process. However, it is important to recall that equivalence on an infinite time horizon is a much stronger statement than on a finite one. The key observation is that the time average of any function along a typical infinite trajectory gives almost surely the value of the integral of the function against some invariant measure. This was the key fact used in related results in [EMS01] and will be central to the proof below.

**Remark 1.4** This theorem was formulated in discrete time for simplicity. Since it only concerns uniqueness of the invariant measure, this is not a restriction since one can apply it to a continuous time system simply by subsampling it at integer times.

### 1.1.2 A general form of Harris’ Theorem

We now turn to an extension of the usual Harris theorem on the exponential convergence of Harris chains under a Lyapunov condition. Recall that, given a Markov semigroup  $\{\mathcal{P}_t\}_{t \geq 0}$  over a measurable space  $\mathbf{X}$ , a measurable function  $V: \mathbf{X} \rightarrow \mathbf{R}_+$  is called a *Lyapunov function* for  $\mathcal{P}_t$  if there exist strictly positive constants  $C_V, \gamma, K_V$  such that

$$\mathcal{P}_t V(x) \leq C_V e^{-\gamma t} V(x) + K_V, \quad (1.3)$$

holds for every  $x \in \mathbf{X}$  and every  $t \geq 0$ . Another omnipresent notion in the theory of Harris chains is that of a *small set*. Recall that  $A \subset \mathcal{X}$  is small for a Markov operator  $\mathcal{P}_t$  if there exists  $\delta > 0$  such that

$$\|\mathcal{P}_t(x, \cdot) - \mathcal{P}_t(y, \cdot)\|_{\text{TV}} \leq 1 - \delta, \quad (1.4)$$

holds for every  $x, y \in A$ .<sup>1</sup> (This is actually a slightly weaker notion of a small set than that found in [MT93], but it turns out to be sufficient for the results stated in this article.) With these definitions at hand, Harris’ theorem states that:

**Theorem 1.5** *Let  $\{\mathcal{P}_t\}_{t \geq 0}$  be a Markov semigroup over a measurable space  $\mathbf{X}$  admitting a Lyapunov function  $V$  and a time  $t_* > \gamma^{-1} \log C_V$  such that the level sets  $\{x \in \mathbf{X} : V(x) \leq C\}$  are small for  $\mathcal{P}_{t_*}$  for every  $C > 0$ . Then, there exists a unique probability measure  $\mu_*$  on  $\mathbf{X}$  that is invariant for  $\mathcal{P}_t$ . Furthermore, there exist constants  $\tilde{C} > 0$  and  $\tilde{\gamma} > 0$  such that the transition probabilities  $\mathcal{P}_t(x, \cdot)$  satisfy*

$$\|\mathcal{P}_t(x, \cdot) - \mu_*\|_{\text{TV}} \leq \tilde{C}(1 + V(x))e^{-\tilde{\gamma}t},$$

for every  $t \geq 0$  and every  $x \in \mathbf{X}$ .

**Remark 1.6** The definition of a Lyapunov function given in (1.3) is slightly different from the one usually given in the monograph [MT93], where  $K_V$  is replaced by  $K_V \mathbf{1}_C$  for some small set  $C$ . Since our aim is to precisely study situations where one does not wish to rely on small sets, (1.3) seems natural. Furthermore, in many situations,  $C$  can be taken as a level set of  $V$ , in which case (1.3) coincides with the definition in [MT93].

<sup>1</sup>In this article, we normalise the total variation distance in such a way that mutually singular probability measures are at distance 1 of each other. This differs by a factor 2 from the definition sometimes found in the literature.

**Remark 1.7** The convergence actually takes place in a stronger total variation norm weighted by  $V$ , in which the Markov semigroup then admits a spectral gap, see [MT93, HM08b].

While this result has been widely applied in the study of the long-time behaviour of Markov processes [MT93], it does not seem to be very suitable for the study of infinite-dimensional evolution equations because the notion of small sets requires that the transition measure not be mutually singular for nearby points.

This suggests that one should seek for a version of Harris' theorem that makes use of a relaxed notion of a small set, allowing for transition probabilities to be mutually singular. To this effect, we will introduce the notion of a  $d$ -small set for a given function  $d: \mathbf{X} \times \mathbf{X} \rightarrow [0, 1]$  used to measure distances between transition probabilities. This will be the content of Definition 4.4 below. If we lift  $d$  to the space of probability measures in the same way that one defines Wasserstein-1 distances, then this notion is just (1.4) with the total variation distance replaced by  $d$ .

However, we can of course not use any distance function  $d$  and expect to obtain a convergence result by combining it simply with Lyapunov stability. We therefore introduce the concept of a distance  $d$  that is *contracting* for  $\mathcal{P}_t$  if there exists  $\alpha < 1$  such that

$$d(\mathcal{P}_t(x, \cdot), \mathcal{P}_t(y, \cdot)) \leq \alpha d(x, y), \quad (1.5)$$

holds for any two points  $x, y \in \mathbf{X}$  such that  $d(x, y) < 1$ . This seems to be a very stringent condition at first sight (one has the impression that (1.5) alone is already sufficient to give exponential convergence of  $\mathcal{P}_t \mu$  to  $\mu_*$  when measured in the distance  $d$ ), but it is very important to note that (1.5) is *not* assumed to hold when  $d(x, y) = 1$ . Therefore, the interesting class of distance functions  $d$  will consist of functions that are equal to 1 for “most” pairs of points  $x$  and  $y$ . Compare this with the fact that the total variation distance can be viewed as the Wasserstein-1 distance corresponding to the trivial metric that is equal to 1 for any two points that are not identical.

With these definitions at hand, a slightly simplified version of our main abstract theorem states that:

**Theorem 1.8** *Let  $\{\mathcal{P}_t\}_{t \geq 0}$  be a Markov semigroup over a Polish space  $\mathbf{X}$  that admits a Lyapunov function  $V$ . Assume furthermore that there exists  $t_* > \gamma^{-1} \log C_V$  and a lower semi-continuous metric  $d: \mathbf{X} \times \mathbf{X} \rightarrow [0, 1]$  such that*

- $d^2$  is contracting for  $\mathcal{P}_{t_*}$ ,
- level sets of  $V$  are  $d$ -small for  $\mathcal{P}_{t_*}$ .

*Then there exists a unique invariant measure  $\mu_*$  for  $\mathcal{P}_t$  and the convergence  $d(\mathcal{P}_t(x, \cdot), \mu_*) \rightarrow 0$  is exponential for every  $x \in \mathbf{X}$ .*

## 1.2 Structure of paper

In Section 2, we give the proof of Theorem 1.1 as well as a result which under related hypotheses ensures that the transition probabilities starting from any point converge to the invariant measure. In Section 3, we apply the results of the preceding theorems to prove the unique ergodicity and convergence of transition probabilities for a wide class of SDDE, including those with state dependent coefficients. In Section 4, we prove a general form of Harris' ergodic theorem which implies exponential convergence in a type of weighted Wasserstein-1 distance on measures and an associated spectral gap. In Section 5, we apply these results to the SDDE setting under the additional assumption of a Lyapunov function in order to obtain a spectral gap result. Lastly, in Section 5.3, we show how to apply the results to the SPDE setting, thus providing an alternative proof to the results in [HM08a].

**Acknowledgments:** Work on this paper began while JCM and MS were visiting the Centro di Ricerca Matematica Ennio De Giorgi in 2006. MH and MS continued the work while both visited the Mittag-Leffler institute and later on during a workshop at the Oberwolfach mathematical research institute. Lastly, the end of the tunnel was seen while JCM visited the Warwick Maths Institute. We thank all of these fine institutions for their hospitality and role in this collaboration. We would also like to thank P. Bubak, M.S. Kinnally and R.J. Williams for pointing out several errors in a previous version of the manuscript.

MH gratefully acknowledges support from an EPSRC Advanced Research Fellowship (grant number EP/D071593/1). JCM gratefully acknowledges support from an NSF PECASE award (DMS-0449910) and a Sloan foundation fellowship. MS gratefully acknowledges support from the DFG Forschergruppe 718 ‘‘Analysis and Stochastics in Complex Physical Systems’’.

## 2 Asymptotic coupling

This section contains the proof of Theorem 1.1 and a number of results which use related ideas. The goal throughout this section is to use minimal assumptions. We also provide a criterion that yields convergence rates toward the invariant measure using a coupling argument. However, in the case of exponential convergence, a somewhat cleaner statement will be provided by the general form of Harris’ theorem in Section 4.

### 2.1 Proof of Theorem 1.1: Unique ergodicity through asymptotic coupling

The proof given below abstracts the essence of the arguments given in [EMS01, BM05]. A version of this theorem is presented in [Mat07]. The basic idea is to leverage the fact that if the initial condition is distributed as an ergodic invariant measure, then Birkhoff’s ergodic theorem implies that the average along a trajectory of a given function is an almost sure property of the trajectories.

*Proof of Theorem 1.1.* Throughout this proof we will use the shorthand notation  $m_i = \mathcal{P}_{[\infty]\mu_i}$  for  $i = 1, 2$ . Clearly 1. implies 2. because if  $\mu_1 = \mu_2$ , then we can take the asymptotic coupling to be the measure  $m_1$  (which then equals  $m_2$ ) pushed onto the diagonal of  $\mathbf{X}^\infty \times \mathbf{X}^\infty$ . Clearly 2. implies 3. since  $\mathcal{C}(\mathcal{P}_{[\infty]\mu_1}, \mathcal{P}_{[\infty]\mu_2}) \subset \tilde{\mathcal{C}}(\mathcal{P}_{[\infty]\mu_1}, \mathcal{P}_{[\infty]\mu_2})$  and the definition of asymptotic coupling implies that there exists a  $\Gamma \in \mathcal{C}(\mathcal{P}_{[\infty]\mu_1}, \mathcal{P}_{[\infty]\mu_2})$  with  $\Gamma(\mathcal{D}) = 1$ . We now turn to the meat of the result, proving that 3. implies 1.

Defining the shift  $\theta: \mathbf{X}^\infty \rightarrow \mathbf{X}^\infty$  by  $(\theta x)_k = x_{k+1}$  for  $k \in \mathbf{N}_0$ , we observe that  $\theta_\# m_i = m_i$  for  $i = 1, 2$ . Or in other words  $m_i$  is an invariant measure for the map  $\theta$ . In addition, one sees that  $m_i$  is an ergodic invariant measure for the shift  $\theta$  since  $\mu_i$  was ergodic for  $\mathcal{P}$ .

Fixing any bounded, globally Lipschitz  $\varphi: \mathbf{X} \rightarrow \mathbf{R}$  we extend it to a bounded function  $\tilde{\varphi}: \mathbf{X}^\infty \rightarrow \mathbf{R}$  by setting  $\tilde{\varphi}(x) = \varphi(x_0)$  for  $x \in \mathbf{X}^\infty$ . Now Birkhoff’s ergodic theorem and the fact that the  $m_i$  are ergodic for  $\theta$  ensure the existence of sets  $A_i^\varphi \subset \mathbf{X}^\infty$  with  $m_i(A_i^\varphi) = 1$  so that if  $x \in A_i^\varphi$  then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \varphi(x_k) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \tilde{\varphi}(\theta^k x) = \int_{\mathbf{X}^\infty} \tilde{\varphi}(x) m_i(dx) = \int_{\mathbf{X}} \varphi(z) \mu_i(dz). \quad (2.1)$$

Here, the first and last implications follow from the definition of  $\tilde{\varphi}$ . Now let  $\Gamma \in \tilde{\mathcal{C}}(m_1, m_2)$  with  $\Gamma(\mathcal{D}) > 0$ . Since  $\Pi_\#^{(i)} \Gamma \ll m_i$  and both are probability measures we know that  $(\Pi_\#^{(i)} \Gamma)(A_i^\varphi) = m_i(A_i^\varphi) = 1$  for  $i = 1, 2$ . This in turn implies that  $\Gamma(\mathbf{X}^\infty \times A_2^\varphi) = \Gamma(A_1^\varphi \times \mathbf{X}^\infty) = 1$ , so that

$\Gamma(A_1^\varphi \times A_2^\varphi) = 1$ . Hence if  $\tilde{\mathcal{D}} = \mathcal{D} \cap (A_1^\varphi \times A_2^\varphi)$ , we have that  $\Gamma(\tilde{\mathcal{D}}) > 0$ . In particular, this implies that  $\tilde{\mathcal{D}}$  is not empty. Observe that for any  $(x^{(1)}, x^{(2)}) \in \tilde{\mathcal{D}}$  we know that for  $i = 1, 2$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \varphi(x_k^{(i)}) = \int_{\mathbf{X}} \varphi(z) \mu_i(dz).$$

On the other hand, since  $(x^{(1)}, x^{(2)}) \in \tilde{\mathcal{D}} \subset \mathcal{D}$ , we know that  $\lim_n d(x_n^{(1)}, x_n^{(2)}) = 0$ . Combining these facts gives

$$\begin{aligned} \left| \int \varphi(x) \mu_1(dx) - \int \varphi(x) \mu_2(dx) \right| &= \left| \lim_{n \rightarrow \infty} \left( \frac{1}{n} \sum_{k=0}^{n-1} \varphi(x_k^{(1)}) - \frac{1}{n} \sum_{k=0}^{n-1} \varphi(x_k^{(2)}) \right) \right| \\ &\leq \lim_{n \rightarrow \infty} \frac{\text{Lip } \varphi}{n} \sum_{k=0}^{n-1} d(x_k^{(1)}, x_k^{(2)}) = 0, \end{aligned}$$

where both the first equality and the second inequality follow from the fact that we choose  $(x^{(1)}, x^{(2)})$  in  $\tilde{\mathcal{D}}$ . Therefore,

$$\int \varphi(x) \mu_1(dx) = \int \varphi(x) \mu_2(dx)$$

for any Lipschitz and bounded  $\varphi$  which implies that  $\mu_1 = \mu_2$ .  $\square$

**Remark 2.1** Note that it is not true in general that the uniqueness of the invariant measure implies that there exist asymptotic couplings for any two starting points! See for example [Lin92, CW00] for a discussion on the relation between coupling, shift coupling, ergodicity, and mixing.

**Corollary 2.2** *Let  $\mathcal{P}$  be a Markov operator on a Polish space. If there exists a measurable set  $A \subset \mathbf{X}$  with the following two properties:*

- $\mu(A) > 0$  for any invariant probability measure  $\mu$  of  $\mathcal{P}$ ,
- there exists a measurable map  $A \times A \ni (x, y) \mapsto \Gamma_{x,y} \in \tilde{\mathcal{C}}(\mathcal{P}_{[\infty]} \delta_x, \mathcal{P}_{[\infty]} \delta_y)$  such that  $\Gamma_{x,y}(\mathcal{D}) > 0$  for every  $x, y \in A$ .

then  $\mathcal{P}$  has at most one invariant probability measure.

**Remark 2.3** The measurability of the map  $\Gamma$  means that the map  $(x, y) \mapsto \int \varphi d\Gamma_{x,y}$  is measurable for every bounded continuous function  $\varphi: \mathbf{X}^\infty \times \mathbf{X}^\infty \rightarrow \mathbf{R}$ .

*Proof of Corollary 2.2.* Assume that there are two invariant measures  $\mu_1$  and  $\mu_2$ . By the ergodic decomposition, we can assume that  $\mu_1$  and  $\mu_2$  are both ergodic since any invariant measure can be decomposed into ergodic invariant measures. We extend the definition of  $\Gamma$  to  $\mathbf{X} \times \mathbf{X}$  by  $\Gamma_{x,y} = \mathcal{P}_{[\infty]} \delta_x \times \mathcal{P}_{[\infty]} \delta_y$  for  $(x, y) \notin A \times A$ . For measurable sets  $B \subset \mathbf{X}^\infty \times \mathbf{X}^\infty$  define the measure  $\Gamma$  by

$$\Gamma(B) = \int_{\mathbf{X} \times \mathbf{X}} \Gamma_{x,y}(B) \mu_1(dx) \mu_2(dy)$$

and notice that  $\Gamma \in \tilde{\mathcal{C}}(\mathcal{P}_{[\infty]} \mu_1, \mathcal{P}_{[\infty]} \mu_2)$  by construction. Furthermore by the assumption that  $\Gamma_{x,y}(\mathcal{D}) > 0$  for  $(x, y) \in A \times A$ , we see that  $\Gamma(\mathcal{D}) > 0$ . Hence Theorem 1.1 implies that  $\mu_1 = \mu_2$ .  $\square$

## 2.2 Convergence of transition probabilities

In this section, we give a simple criterion for the convergence of transition probabilities towards an invariant measure under extremely weak conditions that essentially state that:

1. There exists a point  $x_0 \in \mathbf{X}$  such that the process returns “often” to arbitrarily small neighbourhoods of  $x_0$ .
2. The coupling probability between trajectories starting at  $x$  and  $y$  converges to 1 as  $y \rightarrow x$ .

More precisely, we have the following result:

**Theorem 2.4** *Let  $\mathcal{P}$  be a Markov kernel over a Polish space  $\mathbf{X}$  with metric  $d$  that admits an ergodic invariant measure  $\mu_*$ . Assume that there exist  $B \subset \mathbf{X}$  with  $\mu_*(B) > \frac{1}{2}$  and  $x_0 \in \mathbf{X}$  such that, for every neighbourhood  $U$  of  $x_0$  there exists  $k > 0$  such that  $\inf_{y \in B} \mathcal{P}^k(y, U) > 0$ . Assume furthermore that there exists a measurable map*

$$\begin{aligned} \Gamma: \mathbf{X} \times \mathbf{X} &\rightarrow \mathcal{M}(\mathbf{X}^\infty, \mathbf{X}^\infty) \\ (y, y') &\mapsto \Gamma_{y, y'}, \end{aligned}$$

with the property that  $\Gamma_{y, y'} \in \mathcal{C}(\mathcal{P}_{[\infty]} \delta_y, \mathcal{P}_{[\infty]} \delta_{y'})$  for every  $(y, y')$  and such that, for every  $\varepsilon > 0$  and every  $x \in \mathbf{X}$ , there exists a neighbourhood  $U$  of  $x$  such that  $\inf_{y, y' \in U} \Gamma_{y, y'}(\mathcal{D}) \geq 1 - \varepsilon$ .

Then,  $\mathcal{P}^n(z, \cdot) \rightarrow \mu_*$  weakly as  $n \rightarrow \infty$  for every  $z \in \text{supp } \mu_*$ .

**Remark 2.5** This convergence result is valid even in situations where the invariant measure is not unique. If the process is irreducible however, then  $\text{supp } \mu_* = \mathbf{X}$  and the convergence holds for every  $z \in \mathbf{X}$  (implying in particular that  $\mu_*$  is unique).

*Proof.* We denote by  $\mathcal{D}_N^\varepsilon$  the subset of  $\mathbf{X}^\infty \times \mathbf{X}^\infty$  given by

$$\mathcal{D}_N^\varepsilon = \{(X, Y) : d(X_n, Y_n) \leq \varepsilon \forall n \geq N\},$$

so that  $\mathcal{D} = \bigcap_{\varepsilon > 0} \bigcup_{N > 0} \mathcal{D}_N^\varepsilon$ .

Note first that by the ergodicity of  $\mu_*$  and Birkhoff’s ergodic theorem, there exists a set  $A$  with  $\mu_*(A) = 1$  and such that

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^T \mathbf{1}_B(X_t) = \mu_*(B) > \frac{1}{2},$$

holds almost surely for every process  $X_t$  with  $X_0 \in A$ .

Fix now  $z \in A$  and  $\varepsilon > 0$  and let  $x_0$  be as in the first assumption of the statement. Fix furthermore a neighbourhood  $U$  of  $x_0$  such that  $\Gamma_{y, y'}(\mathcal{D}) \geq 1 - \varepsilon$  for  $y, y' \in U$ . Define the function  $(y, y') \mapsto N_{y, y'}$  by

$$N_{y, y'} = \inf\{N > 0 : \Gamma_{y, y'}(\mathcal{D}_N^\varepsilon) \geq 1 - 2\varepsilon\}. \quad (2.2)$$

It follows from the choice of  $U$  and the fact that  $\mathcal{D} \subset \bigcup_{N > 0} \mathcal{D}_N^\varepsilon$  that one has  $N_{y, y'} < \infty$  for every pair  $y, y' \in U$ . Let now  $Y_t$  be a Markov process generated by  $\mathcal{P}$  with initial distribution  $\mu_*$  and let  $Z_t$  be an independent process with initial condition  $z$ . Since one has the bound  $\mathbf{1}_B(x)\mathbf{1}_B(z) \geq \mathbf{1}_B(x) + \mathbf{1}_B(z) - 1$ , it follows from the definition of  $A$  that

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^T \mathbf{1}_B(Y_t)\mathbf{1}_B(Z_t) \geq 2\mu_*(B) - 1 > 0, \quad (2.3)$$



almost surely. Let now  $\tau = \inf\{t \geq 0 : Y_t \in U \ \& \ Z_t \in U\}$ . Before we proceed, let us argue that  $\tau$  is almost surely finite. We know indeed by assumption that there exists  $T_U > 0$  and  $\alpha > 0$  such that  $\mathcal{P}^{T_U}(y, U) \geq \alpha$  for every  $y \in B$ . Furthermore, setting  $\tau_0 = 0$  and defining  $\tau_k$  recursively by

$$\tau_k = \inf\{t \geq \tau_{k-1} + T_U : Y_t \in B \ \& \ Z_t \in B\},$$

we have the bound

$$\begin{aligned} \mathbf{P}(\tau \geq T) &\leq \mathbf{P}(\tau \geq T \mid \tau_k < T - T_U) \mathbf{P}(\tau_k < T - T_U) + \mathbf{P}(\tau_k \geq T - T_U) \\ &\leq \mathbf{P}(\tau \neq \tau_1 + T_U \ \& \ \dots \ \& \ \tau \neq \tau_k + T_U) + \mathbf{P}(\tau_k \geq T - T_U) \\ &\leq (1 - \alpha)^k + \mathbf{P}(\tau_k \geq T - T_U), \end{aligned}$$

where the last inequality follows from the strong Markov property. Since we know from (2.3) that the  $\tau_k$  are almost surely finite, we can make  $\mathbf{P}(\tau_k \geq T - T_U)$  arbitrarily small for fixed  $k$  by making  $T$  large.

It follows that there exists  $T_0 > 0$  with  $\mathbf{P}(\tau \leq T_0) \geq 1 - \varepsilon$ . Let now  $\mu_0$  be the law of the stopped process at time  $T_0$ , that is  $\mu_0 = \text{Law}(Y_{T_0 \wedge \tau}, Z_{T_0 \wedge \tau})$ . It follows from the definitions of  $T_0$  and  $\tau$  that  $\mu_0(U \times U) \geq 1 - \varepsilon$ .

Since  $N_{y, y'}$  is finite on  $U \times U$ , we can find a sufficiently large value  $T_1 > 0$  such that

$$\mu_0(\{(y, y') \in U \times U : N_{y, y'} \leq T_1\}) \geq 1 - 2\varepsilon.$$

Let now  $\tilde{\Gamma}_{y, z}$  be the coupling between  $\mathcal{P}_{[\infty]} \mu_\star$  and  $\mathcal{P}_{[\infty]} \delta_z$  obtained by first running two independent copies  $(Y_t, Z_t)$  up to the stopping time  $\tau$  and then running them with the coupling  $\Gamma_{Y_\tau, Z_\tau}$ . This is indeed a coupling by the strong Markov property (recall that we are in the discrete time case). Setting  $T = T_0 + T_1$ , it follows immediately from the construction that under the coupling  $\tilde{\Gamma}_{y, z}$ , we have  $\mathbf{P}(d(Y_t, Z_t) \leq \varepsilon) \geq 1 - 4\varepsilon$  for each  $n \geq T$ , thus yielding the convergence of  $\mathcal{P}^n(z, \cdot)$  to  $\mu_\star$  as required.

Let us now extend this argument to more general starting points and fix an arbitrary  $z \in \text{supp } \mu_\star$  and  $\varepsilon > 0$ . Since  $A$  is dense in  $\text{supp } \mu_\star$ , it follows from our assumption on  $\Gamma$  that there exists  $z' \in \mathbf{X}$  such that  $\Gamma_{z, z'}(\mathcal{D}) \geq 1 - \varepsilon$ . In particular, we can find a time  $T_2$  such that  $\Gamma_{z, z'}(\mathcal{D}_{T_2}^\varepsilon) \geq 1 - 2\varepsilon$ . Since on the other hand, we know that  $\mathcal{P}^n(z', \cdot) \rightarrow \mu_\star$  weakly, it immediately follows that we can find  $T_3 \geq T_2$  and for each  $n \geq T_3$  a coupling  $\Gamma$  between  $\mathcal{P}^n(z, \cdot)$  and  $\mu_\star$  such that  $\Gamma(\{(y, y') : d(y, y') \leq 2\varepsilon\}) > 1 - 3\varepsilon$ , thus concluding the proof.  $\square$

### 2.3 More convergence results: rates and convergence for all initial conditions

In this section, we continue to develop the ideas of the previous sections to show how to obtain the convergence of the transition probabilities for *every* initial condition and how to obtain a rate of convergence. We essentially follow the ideas laid out in [Mat02b]. They are sufficient to give exponential convergence, but not a convergence in any operator norm. Here we will not attempt to apply the results to our SDDE setting since Theorem 1.5 provides stronger results in our setting of interest. Nonetheless, the ideas and imagery presented in this section is useful to build intuition. It is also a useful technique in situations where exponential convergence doesn't hold.

Define the 1-Wasserstein distance for two probability measures  $\mu_i$  on  $\mathbf{X}$  by

$$d_1(\mu_1, \mu_2) = \sup_{f \in \text{Lip}_1} \left( \int f(x) \mu_1(dx) - \int f(x) \mu_2(dx) \right),$$

where  $\text{Lip}_1$  are the Lipschitz functions  $f: \mathbf{X} \rightarrow \mathbf{R}$  with Lipschitz constant one.

Let  $\mu_1, \mu_2$  be two probability measures on  $\mathbf{X}$  and  $\Gamma \in \mathcal{C}(\mathcal{P}_{[\infty]\mu_1}, \mathcal{P}_{[\infty]\mu_2})$  and let  $Z \in \mathbf{X}^\infty \times \mathbf{X}^\infty$ , with  $Z_n = (Z_n^{(1)}, Z_n^{(2)})$ , be the stochastic process on  $\mathbf{X} \times \mathbf{X}$  with paths distributed as  $\Gamma$ . Let  $\mathcal{G}_n$  be the  $\sigma$ -algebra generated by  $Z_n$  and  $\Pi_n$  be the projection on  $\mathbf{X}^\infty \times \mathbf{X}^\infty$  defined by  $(\Pi_n x)_k = x_{k+n}$ .

We say that  $\Gamma \in \mathcal{C}(\mathcal{P}_{[\infty]\mu_1}, \mathcal{P}_{[\infty]\mu_2})$  is a *marginally Markovian coupling* if for any stopping time  $\tau$  (adapted to  $\mathcal{G}_n$ ), the conditional distribution  $\text{Law}(\Pi_\tau Z | \mathcal{G}_\tau)$  of  $\Pi_\tau Z$  given  $\mathcal{G}_\tau$  belongs almost surely to  $\mathcal{C}(\mathcal{P}_{[\infty]\delta_{Z_\tau^{(1)}}}, \mathcal{P}_{[\infty]\delta_{Z_\tau^{(2)}}})$ , where  $Z_n = (Z_n^{(1)}, Z_n^{(2)})$ . Notice that this is weaker than assuming that  $Z_n$  is a Markov process.

Let  $\varrho: \mathbf{N} \rightarrow (0, \infty)$  be a strictly decreasing function with  $\lim_n \varrho(n) = 0$ . We define the ‘‘neighborhood’’ of the diagonal

$$\Delta_\varrho = \{(x^{(1)}, x^{(2)}) \in \mathbf{X}^\infty \times \mathbf{X}^\infty : d(x_n^{(1)}, x_n^{(2)}) < \varrho(n)\}.$$

For any stochastic process  $Z$  on  $\mathbf{X} \times \mathbf{X}$  with  $Z_n = (Z_n^{(1)}, Z_n^{(2)})$ , we define the stopping time

$$\tau_\varrho(Z) = \inf\{n \geq 1 : d(Z_n^{(1)}, Z_n^{(2)}) \geq \varrho(n)\} \quad (2.4)$$

and for  $B \in \mathbf{X} \times \mathbf{X}$  we define the hitting time

$$\sigma_B(Z) = \inf\{n \geq 0 : Z_n \in B\}. \quad (2.5)$$

**Theorem 2.6** *Consider a Markov operator  $\mathcal{P}$  as before over a Polish space  $\mathbf{X}$  with distance function  $d \leq 1$ .*

*Fix a strictly decreasing rate function  $\varrho$  and a set  $B \subset \mathbf{X} \times \mathbf{X}$ . Assume that there exists a measurable map  $z_0 \mapsto \Gamma_{z_0} \in \mathcal{C}(\mathcal{P}_{[\infty]\delta_{z_0^{(1)}}}, \mathcal{P}_{[\infty]\delta_{z_0^{(2)}}})$  where  $z_0 = (z_0^{(1)}, z_0^{(2)}) \in \mathbf{X} \times \mathbf{X}$ , which is a marginally Markovian coupling such that, if  $Z$  with  $Z_n = (Z_n^{(1)}, Z_n^{(2)})$ , is distributed as  $\Gamma_{z_0}$  then the following assumptions hold:*

1. *If  $z_0 \notin B$ , then  $\sigma_B(Z)$  is finite almost surely.*
2. *There exists an  $\alpha > 0$  so that if  $z_0 \in B$  then  $\Gamma_{z_0}(\Delta_\varrho) \geq \alpha$ .*

*Then for all  $(z_0^{(1)}, z_0^{(2)}) \in \mathbf{X} \times \mathbf{X}$ ,*

$$d_1(\mathcal{P}_n \delta_{z_0^{(1)}}, \mathcal{P}_n \delta_{z_0^{(2)}}) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

*Proof.* To prove the result, we will construct a new coupling on  $\mathbf{X}^\infty \times \mathbf{X}^\infty$  from the coupling  $\Gamma$ . We will do this by constructing a process on excursions from  $B$ . The state space of our process of excursions will be given by:

$$\mathbb{X} = \bigcup_{k=0}^{\infty} ((\mathbf{X}^k \times \mathbf{X}^k) \times B) \quad \text{and} \quad \overline{\mathbb{X}} = (\mathbf{X}^\infty \times \mathbf{X}^\infty) \cup \mathbb{X}.$$

In words,  $\mathbb{X}$  is the space of finite (but arbitrary) length trajectories taking values in the product space  $\mathbf{X} \times \mathbf{X}$  and ending in  $B$ . The space  $\overline{\mathbb{X}}$  furthermore contains trajectories in  $\mathbf{X} \times \mathbf{X}$  of infinite length.

To build the process on excursions, we begin by constructing a Markov transition kernel  $Q: B \rightarrow \mathcal{M}(\overline{\mathbb{X}})$  from the  $\Gamma$ 's in the following way. For any  $z \in B$ , let  $Z'$  be a  $(\mathbf{X}^\infty \times \mathbf{X}^\infty)$ -valued random variable distributed according to the measure  $\Gamma_z$  and set  $\tau = \tau_\varrho(Z')$  with  $\tau_\varrho$  as in (2.4). If  $\tau = \infty$ , then we set

$Z = Z'$ . If  $\tau < \infty$  and  $Z'_\tau \in B$ , then we set  $Z = (Z'_1, \dots, Z'_\tau)$ . Otherwise, let  $Z''$  be the  $(\mathbf{X}^\infty \times \mathbf{X}^\infty)$ -valued random variable distributed according to the measure  $\Gamma_{Z'_\tau}$ . Since  $\Gamma$  is marginally Markovian and  $\tau$  is a stopping time, we know that the law of  $\Pi_\tau Z'_\tau$  is a coupling of  $\mathcal{P}_{[\infty]} \delta_{Z'_\tau(1)}$ , and  $\mathcal{P}_{[\infty]} \delta_{Z'_\tau(2)}$  almost-surely. Hence we can replace the trajectory of  $Z'$  after time  $\tau$  with the piece of trajectory  $Z''$  and the combined trajectory will still be a coupling starting from the initial data. Setting  $\sigma = \sigma_B(Z'')$  (which is finite almost surely) we define  $Z$  by  $Z = (Z'_1, \dots, Z'_\tau, Z''_1, \dots, Z''_\sigma)$ .

In all cases,  $Z$  is either a trajectory of infinite length contained in  $\Delta_\rho$  or it is a segment of finite length ending in the set  $B$ . Hence  $Z$  is a  $\overline{\mathbb{X}}$ -valued random variable and we define  $Q_z(\cdot)$  to be the distribution of  $Z$ . To extend the definition to a kernel  $Q: \mathbb{X} \rightarrow \mathcal{M}(\overline{\mathbb{X}})$ , we simply define  $Q_z$  for  $z = (z_1, \dots, z_k) \in \mathbb{X}$  by  $Q_z = Q_{z_k}$  since all trajectories in  $\mathbb{X}$  terminate in  $B$ .

We now construct our Markov process on  $\overline{\mathbb{X}}$  which we will denote by  $\mathcal{Z}_n = (Z_n^{(1)}, Z_n^{(2)})$ . We use  $Q$  which was constructed in the preceding paragraph as the Markov transition kernel. If ever the segment drawn is of infinite length, then the process simply stops. The only missing element is the initial condition. If  $z_0 \in B$  then we take  $\mathcal{Z}_0 = z_0$ . If  $z_0 \notin B$ , then we take  $\mathcal{Z}_0 = (z_0, Z'_1, \dots, Z'_\sigma)$  where  $Z'$  is an  $\mathbb{X}$ -valued random variable distributed according to  $\Gamma_{z_0}$  and  $\sigma = \sigma_B(Z')$ .

Now we define  $l_n$  to be the length of  $\mathcal{Z}_n$ . Let  $n^* = \inf\{n : l_n = \infty\}$ . Since for all  $z \in B$ ,  $\Gamma_z(\Delta_\rho) \geq \alpha > 0$  the  $\mathbf{P}(n^* > n) \leq (1 - \alpha)^n$  and thus  $n^*$  is almost surely finite. Lastly we define  $t^* = \sum_{k=0}^{n^*-1} l_k$ . Since  $n^*$ , and the  $l_k$  are all almost surely finite, one sees that  $t^*$  is almost surely finite.

Finally, we are ready to perform the desired calculation. We will denote by  $Z_t = (Z_t^{(1)}, Z_t^{(2)})$  the trajectory in  $\mathbf{X}^\infty \times \mathbf{X}^\infty$  obtained by concatenating together the segments produced by running the Markov chain  $\mathcal{Z}_n$  constructed in the preceding paragraph. For any  $f \in \text{Lip}_1(\mathbf{X})$  one has

$$\begin{aligned} \mathbf{E}f(Z_t^{(1)}) - \mathbf{E}f(Z_t^{(2)}) &\leq \mathbf{E}d(Z_t^{(1)}, Z_t^{(2)})(\mathbf{1}_{t^* > t/2} + \mathbf{1}_{t^* \leq t/2}) \\ &\leq \mathbf{P}(t^* > t/2) + \rho(t/2). \end{aligned} \quad (2.6)$$

Observe that the right hand side is uniform for any  $f \in \text{Lip}_1$ . By assumption  $\rho(t/2) \rightarrow 0$  as  $t \rightarrow \infty$  and since  $t^*$  is almost surely finite  $\mathbf{P}(t^* > t/2) \rightarrow 0$  as  $t \rightarrow \infty$ .  $\square$

The following corollary gives a rate of convergence assuming one can control a appropriate moment of  $t^*$ . As in [Hai02, Mat02b], this is often done by assuming an appropriate Lyapunov structure. This result *does not* prove convergence in any operator norm. In this way, it is inferior to Theorem 1.5 which we prefer. (The fact that the norm does not allow test functions  $f \leq V$  can be rectified with more work under additional assumptions.)

**Corollary 2.7** *In the setting of Theorem 2.6, let  $t^*(z_0)$  be the stopping time defined in the proof of Theorem 2.6 when starting from  $z_0 = (z_0^{(1)}, z_0^{(2)}) \in \mathbf{X} \times \mathbf{X}$ . If  $\mathbf{E}(1/\rho(t^*(z_0))) < \Phi(z_0)$  for some  $\Phi: \mathbf{X} \times \mathbf{X} \rightarrow (0, \infty)$ , then*

$$d_1(\mathcal{P}_t \delta_{z_0^{(1)}}, \mathcal{P}_t \delta_{z_0^{(2)}}) \leq (1 + \Phi(z_0))\rho(t/2)$$

*Proof.* First observe that the Markov inequality implies that  $\mathbf{P}(t^* > t/2) = \mathbf{P}(1/\rho(t^*) > 1/\rho(t/2)) \leq \rho(t/2)\mathbf{E}(1/\rho(t^*)) \leq \rho(t/2)\Phi(z_0)$ . Returning to (2.6), we see that this estimate completes the proof of the desired result.  $\square$

**Remark 2.8** For this result to be useful, one needs control over  $\mathbf{E}(1/\rho(t^*(z_0)))$ . First observe that since  $t^* = \sum_{k=0}^{n^*-1} l_k$  and  $\mathbf{P}(n^* > k) \leq (1 - \alpha)^k$  the main difficulty is controlling the appropriate

moment of  $l_n$ .  $l_n$  consists of two parts. The first is  $\tau$ , the time to exit  $\Delta_\varrho$ , and the second is  $\sigma$  the time to return to  $B$ . The moments of  $\tau$  depend on how quickly  $\Gamma(\Delta_\varrho^n) - \Gamma(\Delta_\varrho)$  goes to zero as a function of  $n$  where  $\Delta_\varrho^n = \{(x^{(i)}, x^{(2)}) \in \mathbf{X}^\infty \times \mathbf{X}^\infty : d(x_k^{(1)}, x_k^{(2)}) < \varrho(k) \text{ for all } k \leq n\}$ . This gives information about how long it takes for  $Z$  to leave  $\Delta_\varrho$  when it is conditioned not to stay in inside  $\Delta_\varrho$  for all times. If this exit time has heavy tails, then it can retard the convergence rate. The return times to  $B$  also influences the convergence rate. Such a return time is often controlled by a Lyapunov function. In the context of obtaining exponential convergence, some of these points are explored in [Mat02b, Hai02]. Subexponential convergence rates via Lyapunov functions has been explored for Harris chains in [MT93, Ver99, BCG08, DFG09, Hai09].

### 3 Application of the uniqueness and convergence criteria to SDDEs

#### 3.1 Application of the uniqueness criterion to SDDEs

Fix  $r > 0$  and let  $\mathcal{C} = \mathcal{C}([-r, 0], \mathbf{R}^d)$  denote the phase space of a general finite-dimensional delay equation with delay  $r$  endowed with the sup-norm  $\|\cdot\|$ . For a function or a process  $X$  defined on  $[t - r, t]$  we write  $X_t(s) := X(t + s)$ ,  $s \in [-r, 0]$ . Consider the following stochastic functional differential equation:

$$\begin{aligned} dX(t) &= f(X_t) dt + g(X_t) dW(t), \\ X_0 &= \eta \in \mathcal{C}, \end{aligned} \tag{3.1}$$

where  $f: \mathcal{C} \rightarrow \mathbf{R}^d$  and  $g: \mathcal{C} \rightarrow \mathbf{R}^m \times \mathbf{R}^d$  and  $W_t = (W_t^{(1)}, \dots, W_t^{(m)})$  is a standard Wiener process.

We will provide conditions on  $f$  and  $g$  which ensure that, for every initial condition  $X_0 = \eta \in \mathcal{C}$ , equation (3.1) has a unique pathwise solution which can then be viewed as a  $\mathcal{C}$ -valued strong Markov process. The problem of existence and/or uniqueness of an invariant measure of such a process (or similar processes) has been addressed by a number of authors, see for example [IN64, Sch84, BM05, RRG06, EGS09].

While existence of an invariant measure has been proven under natural sufficient conditions on the functionals  $f$  and  $g$ , the uniqueness question, as already mentioned in the introduction, has not been answered up to now even in such simple cases as

$$f(x) = -cx(0), \quad g(x) = \psi(x(-r)), \quad d = m = 1,$$

for some  $c > 0$  and  $\psi$  a strictly positive, bounded and strictly increasing function [RRG06]. One difficulty is that the corresponding Markov process on  $\mathcal{C}$  is not strong Feller. Even worse: given the solution  $X_t$  for any  $t > 0$ , the initial condition  $X_0$  can be recovered *with probability one* [Sch05]. Another peculiarity of such equations is that while they do in general generate a Feller semigroup on  $\mathcal{C}$ , they often do not admit a modification which depends continuously on the initial condition – even if  $g$  is linear (see e.g. [Moh86] and [MS97]), so that they do not generate a stochastic flow of homeomorphisms. We do nevertheless have the following uniqueness result:

**Theorem 3.1** *Assume that  $m \geq d > 0$  and that, for every  $\eta \in \mathcal{C}$ ,  $g(\eta)$  admits a right inverse  $g^{-1}(\eta): \mathbf{R}^d \rightarrow \mathbf{R}^m$ . If  $f$  is continuous and bounded on bounded subsets of  $\mathcal{C}$ , and for some  $K \geq 1$ ,  $f$  and  $g$  satisfy*

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

and

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

where  $\|M\|^2 = \text{Tr}(MM^*)$ , then the equation (3.1) has at most one invariant measure.

**Remark 3.2** This assumption is sufficient to ensure the existence of (unique) global solutions (see [RS08] for even weaker hypotheses).

**Remark 3.3** Notice that Theorem 3.1 does not ensure that there exists an invariant measure, only that there exists at most one. Even when there is no invariant (probability) measure, the ideas in Theorem 3.1 can be used to show that solutions with nearby initial conditions behave asymptotically the same with a positive probability.

**Remark 3.4** It is not true in general that one has exponential convergence under the assumptions of Theorem 3.1 (plus the existence of an invariant measure) alone. Consider for example the one-dimensional SDE

$$dx = -\frac{x}{(1+x^2)^\alpha} dt + dW(t),$$

then it is known that for  $\alpha \in (\frac{1}{2}, 1)$  it has a unique invariant measure, but that convergence of transition probabilities is only stretched exponential [Hai09]. However, it does satisfy the one-sided Lipschitz condition of Theorem 3.1.

We will use the following two lemmas which are proven at the end of the section. The first one is similar to Proposition 7.3 in [DZ92].

**Lemma 3.5** *Let  $W(t)$ ,  $t \geq 0$  be a standard Wiener process and fix  $T > 0$  and  $p > 2$ . There exists a function  $\varrho : [0, \infty) \rightarrow [0, \infty)$  satisfying  $\lim_{\lambda \rightarrow \infty} \varrho(\lambda) = 0$  such that the following holds: let  $Y$  satisfy the equation*

$$\begin{aligned} dY(t) &= -\lambda Y(t) dt + h(t) dW(t), \\ Y(0) &= 0, \end{aligned} \tag{3.2}$$

where  $h$  is an adapted process with almost surely càdlàg sample paths. Then for any stopping time  $\tau$  we have

$$\mathbf{E} \left( \sup_{0 \leq t \leq \tau \wedge T} |Y(t)|^p \right) \leq \varrho(\lambda) \mathbf{E} \left( \sup_{0 \leq t \leq \tau \wedge T} |h(t)|^p \right).$$

**Lemma 3.6** *Let  $\lambda > 0$ , consider the coupled set of equations:*

$$\begin{aligned} dX(t) &= f(X_t) dt + g(X_t) dW(t), & X_0 &= \eta, \\ d\tilde{X}(t) &= f(\tilde{X}_t) dt + \lambda(X(t) - \tilde{X}(t)) dt + g(\tilde{X}_t) dW(t), & \tilde{X}_0 &= \tilde{\eta}, \end{aligned} \tag{3.3}$$

and define  $Z(t) := X(t) - \tilde{X}(t)$ ,  $t \geq -r$ . Then, for every  $\gamma_0 > 0$  there exist  $\lambda > 0$  and  $C > 0$  such that the bound  $\mathbf{E}(\sup_{t \geq 0} e^{\gamma_0 t} \|Z_t\|)^8 \leq (C\|Z_0\|)^8$  holds for any pair of initial conditions  $X_0$  and  $\tilde{X}_0$ .

*Proof of Theorem 3.1.* We begin by fixing two initial conditions  $\eta, \tilde{\eta} \in \mathcal{C}$  of (3.3). We furthermore define the ‘‘Girsanov shift’’  $v$  by

$$v(t) = \lambda g(\tilde{X}_t)^{-1}(X(t) - \tilde{X}(t)),$$

where  $\lambda > 0$  is chosen as in Lemma 3.6 for  $\gamma_0 = 1$  (say) and we set  $\tau = \inf\{t \geq 0 : \int_0^t |v(s)|^2 ds \geq \varepsilon^{-1} \|\eta - \tilde{\eta}\|^2\}$ , where  $\varepsilon > 0$  is a small constant to be determined. Thanks to the non-degeneracy

assumption on  $g$  and Lemma 3.6, we obtain  $\lim_{\varepsilon \rightarrow 0} \mathbf{P}\{\tau = \infty\} = 1$  and  $\lim_{t \rightarrow \infty} |X(t) - \tilde{X}(t)| = 0$  almost surely. In particular, there exist some  $\varepsilon > 0$  independent of the initial conditions such that  $\mathbf{P}\{\tau = \infty\} \geq 0$ . We will fix such a value of  $\varepsilon$  from now on.

Setting  $\tilde{W}(t) = W(t) + \int_0^{t \wedge \tau} v(s) ds$ , we observe that the Cameron-Martin-Girsanov Theorem implies that there exists a measure  $\mathbf{Q}$  on  $\Omega := \mathcal{C}([0, \infty), \mathbf{R}^m)$  so that under  $\mathbf{Q}$ ,  $\tilde{W}$  is a standard Wiener process on the time interval  $[0, \infty)$ . Let  $\bar{X}$  be the solution of

$$d\bar{X}(t) = f(\bar{X}_t) dt + g(\bar{X}_t) d\tilde{W}(t) \quad \bar{X}_0 = \tilde{\eta}.$$

Since  $\int_0^\tau v^2(s) ds \leq \varepsilon^{-1} \|\eta - \tilde{\eta}\|^2$  by construction, the law of  $\bar{X}$  is equivalent on  $\mathcal{C}([0, \infty), \mathbf{R}^d)$  to the law of a solution to (3.1) with initial condition  $\tilde{\eta}$ . This means that the law of the pair  $(X, \bar{X})$  has marginals which are equivalent on  $\mathcal{C}([0, \infty), \mathbf{R}^d)$  to solutions to (3.1) starting respectively from  $\eta$  and  $\tilde{\eta}$ . Since  $\bar{X} = \tilde{X}$  on  $\{\tau = \infty\}$ , we have

$$\lim_{t \rightarrow \infty} |\bar{X}(t) - X(t)| = \lim_{t \rightarrow \infty} |\tilde{X}(t) - X(t)| = 0 \quad \text{on } \{\tau = \infty\} \text{ a.s.}$$

Therefore Corollary 2.2 implies that the discrete-time chain  $(X_{rn})_{n \in \mathbf{N}}$  has at most one invariant probability measure and hence the same is true for the Markov process  $(X_t)_{t \geq 0}$ . Since  $\mathcal{M}(\Omega, \Omega)$  endowed with the topology of weak convergence is a Polish space [Vil03] and since all of the constants appearing in our explicit construction can be chosen independently of the initial conditions, the map  $(x, y) \mapsto \Gamma_{x,y}$  is indeed measurable.  $\square$

We now give the proof of Lemma 3.5 which was given at the start of the section.

*Proof of Lemma 3.5.* We begin by noticing that we need only prove the theorem for the supremum over a deterministic time interval  $[0, T]$ . The version over the random time interval follows by considering the function  $\tilde{h}(s) = h(s) \mathbf{1}_{[0, \tau \wedge T)}(s)$ . Observe that  $\tilde{h}(s)$  again almost surely has càdlàg paths and if  $\tilde{Y}(t)$  is the solution to (3.2) with  $h$  replaced by  $\tilde{h}$  then

$$\sup_{t \leq \tau \wedge T} |Y(s)|^p = \sup_{t \leq T} |\tilde{Y}(s)|^p \quad \text{and} \quad \sup_{t \leq \tau \wedge T} |h(s)|^p = \sup_{t \leq T} |\tilde{h}(s)|^p.$$

The second identity is clear from the definition of  $\tilde{h}$ . The first follows from the observation that  $Y(s) = \tilde{Y}(s)$  for  $s \leq \tau \wedge T$  and, for  $s > \tau \wedge T$ ,  $|\tilde{Y}(s)|$  only decreases since  $\tilde{h}$  is identically zero. Hence it is enough to prove the lemma over a deterministic time interval.

We begin by observing that the solution  $Y$  can be represented in the form

$$Y(t) = e^{-\lambda t} \int_0^t e^{\lambda s} h(s) dW(s). \quad (3.4)$$

Therefore, using Burkholder's inequality and abbreviating  $h^* := \sup_{0 \leq t \leq T} |h(t)|$ , we obtain

$$\begin{aligned} \mathbf{E}|Y(t)|^p &= e^{-\lambda t p} \mathbf{E} \left| \int_0^t e^{\lambda s} h(s) dW(s) \right|^p \\ &\leq C_p e^{-\lambda t p} \mathbf{E} \left| \int_0^t e^{2\lambda s} h^2(s) ds \right|^{p/2} \\ &\leq C_p \mathbf{E}(h^*)^p (2\lambda)^{-p/2}. \end{aligned} \quad (3.5)$$

Let  $N \in \mathbf{N}$  and define  $t_k := t_k(N) := kT/N$  for  $k = 0, \dots, N$  and

$$I_k(t) := \int_{t_k}^t h(s) dW(s), \quad t_k \leq t \leq t_{k+1}, \quad k = 0, \dots, N-1.$$

Notice that  $I_k(t)$  is a local martingale with respect to the filtration it generates. Integrating (3.4) by parts, we get

$$Y(t) = Y(t_k) e^{-\lambda(t-t_k)} + I_k(t) - \lambda \int_{t_k}^t e^{-\lambda(t-s)} I_k(s) ds, \quad t_k \leq t \leq t_{k+1}.$$

Hence,

$$\begin{aligned} \sup_{0 \leq t \leq T} |Y(t)|^p &= \max_{k=0, \dots, N-1} \sup_{t_k \leq t \leq t_{k+1}} |Y(t)|^p \\ &\leq 2^{p-1} \max_{k=0, \dots, N-1} \left( |Y(t_k)|^p + 2 \sup_{t_k \leq t \leq t_{k+1}} |I_k(t)|^p \right). \end{aligned}$$

Using Burkholder's inequality and (3.5), we get

$$\begin{aligned} \mathbf{E} \sup_{0 \leq t \leq T} |Y(t)|^p &\leq 2^{p-1} N C_p \mathbf{E}(h^*)^p (2\lambda)^{-p/2} + 2^p N \max_{k=0, \dots, N-1} \mathbf{E} \sup_{t_k \leq t \leq t_{k+1}} |I_k(t)|^p \\ &\leq 2^{p-1} C_p \mathbf{E}(h^*)^p \left( N(2\lambda)^{-p/2} + 2T^{p/2} N^{1-\frac{p}{2}} \right). \end{aligned}$$

For each  $\varepsilon > 0$ , we can choose  $N$  large enough such that the coefficient of  $\mathbf{E}(h^*)^p$  becomes smaller than  $\varepsilon$  for all sufficiently large  $\lambda$ .  $\square$

We conclude with the proof of Lemma 3.6:

*Proof of Lemma 3.6.* First observe that the pair of equations (3.3) admits a unique global solution (see [RS08]). Setting  $Z(t) = X(t) - \tilde{X}(t)$ , we see that

$$\begin{aligned} d|Z(t)|^2 &= 2\langle f(X_t) - f(\tilde{X}_t), Z(t) \rangle dt + \|g(X_t) - g(\tilde{X}_t)\|^2 dt - 2\lambda|Z(t)|^2 dt + dM(t) \\ &\leq K\|Z_t\|^2 dt - 2\lambda|Z(t)|^2 dt + dM(t) \end{aligned}$$

where  $M(0) = 0$  and  $dM(t) = 2\langle Z(t), (g(X_t) - g(\tilde{X}_t)) dW(t) \rangle$ . Define now  $Y(t) = e^{\alpha t} |Z(t)|^2$  for a constant  $\alpha > 0$  to be determined later. Then

$$\begin{aligned} dY(t) &= \alpha Y(t) dt + e^{\alpha t} d|Z(t)|^2 \\ &\leq (\alpha - 2\lambda)Y(t) dt + K e^{\alpha t} \|Z_t\|^2 dt + e^{\alpha t} dM(t) \\ &\leq (\alpha - 2\lambda)Y(t) dt + K e^{\alpha t} \|Y_t\| dt + e^{\alpha t} dM(t). \end{aligned}$$

Setting  $N(t) = \int_0^t e^{-\kappa(t-s)} e^{\alpha s} dM(s)$  and  $\kappa = 2\lambda - \alpha$ , the variation of constants formula thus yields

$$\begin{aligned} Y(t) &\leq e^{-\kappa t} Y(0) + K e^{\alpha t} \int_0^t e^{-\kappa(t-s)} \|Y_s\| ds + N(t) \\ &\leq e^{-\kappa t} Y(0) + \frac{K e^{\alpha t}}{\kappa} \sup_{s \in [0, t]} \|Y_s\| + N(t). \end{aligned}$$

For  $\varepsilon > 0$ , let now  $\tau_\varepsilon$  be the stopping time defined by  $\tau_\varepsilon = 2r \wedge \inf\{t \geq 0 : \|Y_t\| \geq \varepsilon^{-1}\}$ . It follows that there exists a constant  $\bar{K}$  independent of  $\alpha$ ,  $\lambda$  and  $\varepsilon$  such that

$$\mathbf{E} \sup_{s \in [0, \tau_\varepsilon]} \|Y_s\|^4 \leq \bar{K} \|Y_0\|^4 + \frac{\bar{K} e^{4\alpha r}}{\kappa^4} \mathbf{E} \sup_{s \in [0, \tau_\varepsilon]} \|Y_s\|^4 + \bar{K} \mathbf{E} \sup_{s \in [0, \tau_\varepsilon]} |N(s)|^4.$$

Now observe that by Lemma 3.5, we have for  $N$  the bound

$$\mathbf{E} \sup_{s \in [0, \tau_\varepsilon]} |N(s)|^4 \leq \varrho(\kappa) 2^4 \mathbf{E} \sup_{s \in [0, \tau_\varepsilon]} (e^{4\alpha s} |Z(s)|^4 \|g(X_s) - g(\tilde{X}_s)\|^4) \leq C e^{2\alpha r} \varrho(\kappa) \mathbf{E} \sup_{s \in [0, \tau_\varepsilon]} \|Y_s\|^4,$$

for a constant  $C$  independent of  $\alpha$  and  $\lambda$ . This shows that we can find a function  $\alpha \mapsto \Lambda(\alpha)$  such that both  $\bar{K} e^{4\alpha r} / (\Lambda(\alpha) - \alpha)^4 \leq \frac{1}{4}$  and  $\bar{K} C e^{\alpha r} \varrho(2\Lambda(\alpha) - \alpha) \leq \frac{1}{4}$ , thus obtaining

$$\mathbf{E} \sup_{s \in [0, \tau_\varepsilon]} \|Y_s\|^4 \leq \bar{K} \|Y_0\|^4 + \frac{1}{2} \mathbf{E} \sup_{s \in [0, \tau_\varepsilon]} \|Y_s\|^4,$$

provided that we choose  $\lambda = \Lambda(\alpha)$ , so that  $\mathbf{E} \sup_{s \in [0, \tau_\varepsilon]} \|Y_s\|^4 \leq 2\bar{K} \|Y_0\|^4$ . Since this bound is independent of  $\varepsilon > 0$ , we can take the limit  $\varepsilon \rightarrow 0$ , so that the monotone convergence theorem yields  $\mathbf{E} \sup_{s \in [0, 2r]} \|Y_s\|^4 \leq 2\bar{K} \|Y_0\|^4$ . In terms of our original process  $Z$ , we conclude that

$$\mathbf{E} \|Z_r\|^8 \leq 2\bar{K} \|Z_0\|^8 \quad \text{and} \quad \mathbf{E} \|Z_{2r}\|^8 \leq 2\bar{K} e^{-4\alpha r} \|Z_0\|^8. \quad (3.6)$$

Since  $\bar{K}$  is independent of  $\alpha$ , we can ensure that  $2\bar{K} e^{-4\alpha r} \leq e^{-18r\gamma_0}$  by taking  $\alpha$  (and therefore also  $\lambda$ ) sufficiently large. Iterating (3.6), we then obtain

$$\mathbf{E} \|Z_{2nr}\|^8 \leq e^{-18r\gamma_0 n} \|Z_0\|^8 \quad \text{and} \quad \mathbf{E} \|Z_{(2n+1)r}\|^8 \leq 2\bar{K} e^{-18r\gamma_0 n} \|Z_0\|^8. \quad (3.7)$$

Note now that if  $t \in [nr, (n+1)r]$ , then  $\|Z_t\| \leq \|Z_{nr}\| + \|Z_{(n+1)r}\|$ . Therefore, there exists a constant  $C > 0$  such that

$$\sup_{t \geq 0} e^{8\gamma_0 t} \|Z_t\|^8 \leq C \sum_{n=0}^{\infty} e^{8\gamma_0 r n} \|Z_{rn}\|^8.$$

Hence using (3.7), we have for a different constant  $C > 0$

$$\mathbf{E} \sup_{t \geq 0} e^{8\gamma_0 t} \|Z_t\|^8 \leq C \|Z_0\|^8 \sum_{n=0}^{\infty} e^{-\gamma_0 r n}.$$

Since the sum on the right hand side converges, the proof is complete.  $\square$

This shows the uniqueness of the invariant measure for a large class of stochastic delay equations. It turns out that under exactly the same conditions, we can ensure that the invariant measure is not only unique, but that transition probabilities converge to it.

### 3.2 Convergence of transition probabilities of SDDEs

In this section we will apply the abstract results of Section 2.2 to the  $\mathcal{C}$ -valued Markov process  $(X_t)$  which we introduced in the previous section. We will denote its transition probabilities by  $\mathcal{P}_t(\eta, \cdot)$ . We will prove the following result:



**Theorem 3.7** *Let the assumptions of Theorem 3.1 be satisfied. If the Markov process  $X_t, t \geq 0$  admits an invariant probability measure  $\mu$ , then for each  $\eta \in \mathcal{C}$  we have  $\mathcal{P}_t(\eta, \cdot) \rightarrow \mu$  weakly.*

We start with the following lemma (which we will also need in Section 5):

**Lemma 3.8** *Let the assumptions of Theorem 3.1 be satisfied and denote by  $\mathbf{B}_R$  the closed ball in  $\mathcal{C}$  with radius  $R$  and center 0. Then for each  $R, \delta > 0$  and each  $t_* \geq 2r$ ,*

$$\inf_{\|\eta\| \leq R} \mathcal{P}_{t_*}(\eta, \mathbf{B}_\delta) > 0.$$

*Proof.* The proof resembles that of Lemma 2.4 in [SS02]. Fix  $R \geq \delta > 0$ . For each  $y \in \mathbf{R}^d, |y| \leq 3R/2$ , let  $h = h_y : [0, t_*] \rightarrow \mathbf{R}^d$  be continuously differentiable with Lipschitz constant at most  $2R/r$  and satisfy  $h = 0$  on  $[r, t_*]$ ,  $h(0) = y$ . Define

$$D(t) := |X(t) - h(t)|^2 - (\delta/2)^2,$$

where  $X$  solves SDDE (3.1) with initial condition  $\eta \in \mathcal{C}, \|\eta\| \leq R, y := \eta(0) - (\delta/2, 0, \dots, 0)^T$  and  $h = h_y$  is defined as above. Then

$$dD(t) = 2\langle X(t) - h(t), f(X_t) - h'(t) \rangle dt + 2\langle X(t) - h(t), g(X_t) dW(t) \rangle + \sum_{i,j} g_{ij}^2(X_t) dt,$$

while  $D(0) = 0$ . Let  $\tau := \inf\{t \geq 0 : |D(t)| > (\delta/4)^2\}$ . Let now  $W_1$  be a Wiener process that is independent of  $W$  and set

$$Y(t) := D(t \wedge \tau) + (W_1(t) - W_1(\tau))\mathbf{1}_{t \geq \tau}.$$

This is a semimartingale with  $Y(0) = 0$  which fulfills the conditions of Lemma I.8.3 of [Bas98] (with  $(\delta/4)^2$  in place of  $\varepsilon$ ). Therefore, there exists  $p > 0$  such that for all  $\|\eta\| \leq R$  we have

$$\mathcal{P}_{t_*}(\eta, \mathbf{B}_\delta) \geq \mathbf{P}\left(\sup_{0 \leq t \leq t_*} |Y(t)| \leq (\delta/4)^2 \mid X_0 = \eta\right) \geq p,$$

thus concluding the proof.  $\square$

*Proof of Theorem 3.7.* Theorem 2.4 is formulated for discrete time, so we first show that the two conditions are satisfied for the Markov kernel  $\mathcal{P}_t$  for some  $t > 0$ . The previous lemma immediately implies that the first condition of Theorem 2.4 is satisfied for any  $t > 0$  and any sufficiently large  $k$ . The second condition follows from the fact that there exists (a small value)  $\delta > 0$  such that the metric  $d(x, y) := 1 \wedge \delta^{-1}\|x - y\|$  on  $\mathcal{C}$  is contracting for  $\mathcal{P}_t$  (see Definition 4.6) for any sufficiently large  $t > 0$  (which is proved in Section 5.1) and Proposition 4.12. Since the support of  $\mu$  equals  $\mathcal{C}$ , it therefore follows from Theorem 2.4 that for some suitable  $t$ , all transition probabilities of the chain associated to  $\mathcal{P}_t$  converge to  $\mu$  weakly. To show that even all transition probabilities of the continuous-time Markov process  $(X_t)$  converge to  $\mu$  weakly, it suffices to observe that (by Proposition 5.4) there exists a constant  $C$  such that  $d(\mathcal{P}_\tau \nu, \mathcal{P}_\tau \tilde{\nu}) \leq C d(\nu, \tilde{\nu})$  for all  $\tau \in [0, t]$  and all  $\nu, \tilde{\nu}$ .  $\square$

## 4 A general form of Harris' theorem

In this section, we show that under very mild additional assumptions on the dynamic of (3.1), the uniqueness result for an invariant measure obtained in the previous section can be strengthened to an exponential convergence result in a type of weighted Wasserstein distance. Our main ingredient will be the existence of a *Lyapunov* function for our system. Recall that a Lyapunov function for a Markov semigroup  $\{\mathcal{P}_t\}_{t \geq 0}$  over a Polish space  $\mathbf{X}$  is a function  $V: \mathbf{X} \rightarrow [0, \infty]$  such that  $V$  is integrable with respect to  $\mathcal{P}_t(x, \cdot)$  for every  $x \in \mathbf{X}$  and  $t \geq 0$  and such that there exist constants  $C_V, \gamma, K_V > 0$  such that the bound

$$\int V(y) \mathcal{P}_t(x, dy) \leq C_V e^{-\gamma t} V(x) + K_V, \quad (4.1)$$

holds for every  $x \in \mathbf{X}$  and  $t \geq 0$ .

**Remark 4.1** If the generator  $\mathcal{L}$  of  $\mathcal{P}_t$  is explicitly known, then a sufficient (but by no means necessary) condition for a function  $V$  in the domain of  $\mathcal{L}$  to be a Lyapunov function is the existence of positive constants  $\lambda$  and  $K$  such that  $\mathcal{L}V(x) \leq K - \lambda V(x)$  for every  $x \in \mathbf{X}$ .

In the usual theory of stability for Markov processes, the notion of a “small set” plays an equally important role. We say that a set  $A \subset \mathbf{X}$  is *small* if there exists a time  $t > 0$  and a constant  $\varepsilon > 0$  such that

$$\|\mathcal{P}_t(x, \cdot) - \mathcal{P}_t(y, \cdot)\|_{\text{TV}} \leq 1 - \varepsilon, \quad (4.2)$$

for every  $x, y \in A$ . Recall that the total variation distance between two probability measures is equal to 1 if and only if the two measures are mutually singular. A set is therefore small if the transition probabilities starting from any two points in the set have a “common part” of mass at least  $\varepsilon$ . The classical Harris theorem [MT93, HM08b] then states that:

**Theorem 4.2 (Harris)** *Let  $\mathcal{P}_t$  be a Markov semigroup over a Polish space  $\mathbf{X}$  such that there exists a Lyapunov function  $V$  with the additional property that the level sets  $\{x : V(x) \leq C\}$  are small for every  $C > 0$ . Then,  $\mathcal{P}_t$  has a unique invariant measure  $\mu_*$  and  $\|\mathcal{P}_t(x, \cdot) - \mu_*\|_{\text{TV}} \leq C e^{-\gamma_* t} (1 + V(x))$  for some positive constants  $C$  and  $\gamma_*$ .*

The proof of Harris' theorem is based on the fact that a semigroup satisfying these assumptions has a spectral gap in a modified total variation distance, where the variation is weighted by the Lyapunov function  $V$ . This theorem can clearly not be applied to Markov semigroups generated by stochastic delay equations in general. As already mentioned earlier, it is indeed known that even in simple cases where the diffusion coefficient  $g$  only depends on the past of the process, the initial condition can be recovered exactly from the solution at any subsequent time. This implies that in such a case

$$\|\mathcal{P}_t(x, \cdot) - \mathcal{P}_t(y, \cdot)\|_{\text{TV}} = 1$$

for every  $x \neq y$  and every  $t > 0$ , so that (4.2) fails. We would therefore like to replace the notion of a small set (4.2) by a notion of “closedness” between transition probabilities that reflects the topology of the underlying space  $\mathbf{X}$ . Before we state our modified notion of a  $d$ -small set, we introduce another notation: given a positive function  $d: \mathbf{X} \times \mathbf{X} \rightarrow \mathbf{R}_+$ , we extend it to a positive function  $d: \mathcal{M}_1(\mathbf{X}) \times \mathcal{M}_1(\mathbf{X}) \rightarrow \mathbf{R}_+$ , where  $\mathcal{M}_1(\mathbf{X})$  stands for the set of (Borel) probability measures on  $\mathbf{X}$ , by

$$d(\mu, \nu) = \inf_{\pi \in \mathcal{C}(\mu, \nu)} \int_{\mathbf{X}^2} d(x, y) \pi(dx, dy). \quad (4.3)$$

If  $d$  is a metric, then its extension to  $\mathcal{M}_1(\mathbf{X})$  is simply the corresponding Wasserstein-1 distance. In this section, we will be considering functions  $d: \mathbf{X} \times \mathbf{X} \rightarrow \mathbf{R}_+$  that are not necessarily metrics but that are “distance-like” in the following sense:

**Definition 4.3** Given a Polish space  $\mathbf{X}$ , a function  $d: \mathbf{X} \times \mathbf{X} \rightarrow \mathbf{R}_+$  is distance-like if it is symmetric, lower semi-continuous, and such that  $d(x, y) = 0 \Leftrightarrow x = y$ .

Even though we think of  $d$  as being a kind of metric, it need not satisfy the triangle inequality. However, when lifted to the space of probability measures,  $d$  provides a reasonable way of measuring distances between measures in the sense that  $d(\mu, \nu) \geq 0$  and  $d(\mu, \nu) = 0 \Leftrightarrow \mu = \nu$ , the latter property being a consequence of the lower semi-continuity of  $d$ . The lower semicontinuity of  $d$  also ensures that the infimum in (4.3) is always reached by some coupling  $\pi$ . With this notation at hand, we set:

**Definition 4.4** Let  $\mathcal{P}$  be a Markov operator over a Polish space  $\mathbf{X}$  endowed with a distance-like function  $d: \mathbf{X} \times \mathbf{X} \rightarrow [0, 1]$ . A set  $A \subset \mathbf{X}$  is said to be  $d$ -small if there exists  $\varepsilon > 0$  such that

$$d(\mathcal{P}(x, \cdot), \mathcal{P}(y, \cdot)) \leq 1 - \varepsilon, \quad (4.4)$$

for every  $x, y \in A$ .

**Remark 4.5** If  $d(x, y) = d_{\text{TV}}(x, y) := \mathbf{1}_{x \neq y}$ , then the notion of a  $d$ -small set coincides with the notion of a small set given in the introduction, since  $\|\mu - \nu\|_{\text{TV}} = d_{\text{TV}}(\mu, \nu)$ .

In general, it is clear that having a Lyapunov function  $V$  with  $d$ -small level sets cannot be sufficient to imply the unique ergodicity of a Markov semigroup. A simple example is given by the Glauber dynamic of the 2D Ising model which exhibits two distinct ergodic invariant measures at low temperatures, but for which every set is  $d$ -small if  $d$  is a distance function that metrises the product topology on the state space  $\{0, 1\}^{\mathbb{Z}^2}$ , for example  $d(\sigma, \sigma') = \sum_{k \in \mathbb{Z}^2} \frac{1}{2^{|k|}} |\sigma_k - \sigma'_k|$ .

This shows that if we wish to make use of the notion of a  $d$ -small set, we should impose additional assumptions on the function  $d$ . One feature that distinguishes the total variation distance  $d_{\text{TV}}$  among other distance-like functions is that, for any Markov operator  $\mathcal{P}$ , one *always* has the contraction property

$$d_{\text{TV}}(\mathcal{P}\mu, \mathcal{P}\nu) \leq d_{\text{TV}}(\mu, \nu).$$

It is therefore natural to look for distance-like functions with a similar property. This motivates the following definition:

**Definition 4.6** Let  $\mathcal{P}$  be a Markov operator over a Polish space  $\mathbf{X}$  endowed with a distance-like function  $d: \mathbf{X} \times \mathbf{X} \rightarrow [0, 1]$ . The function  $d$  is said to be contracting for  $\mathcal{P}$  if there exists  $\alpha < 1$  such that the bound

$$d(\mathcal{P}(x, \cdot), \mathcal{P}(y, \cdot)) \leq \alpha d(x, y) \quad (4.5)$$

holds for every pair  $x, y \in \mathbf{X}$  with  $d(x, y) < 1$ .

**Remark 4.7** The assumption that  $d$  takes values in  $[0, 1]$  is not a restriction at all. One can indeed check that if an unbounded function  $d$  is contracting for a Markov operator  $\mathcal{P}$  and  $A$  is a  $d$ -small set, then the same statements are true for  $d$  replaced by  $d \wedge 1$ .

It may seem at first sight that (4.5) alone is already sufficient to guarantee the convergence of transition probabilities toward a unique invariant measure. A little more thought shows that this is not the case, since the total variation distance  $d_{TV}$  is contracting for *every* Markov semigroup. The point here is that (4.5) says nothing about the pairs  $(x, y)$  with  $d(x, y) = 1$ , and this set may be very large. However, combined with the existence of a Lyapunov function  $V$  that has  $d$ -small level sets, it turns out that this contraction property is sufficient not only for the existence and uniqueness of the invariant measure  $\mu_*$ , but even for having exponential convergence of transition probabilities to  $\mu_*$  in a type of Wasserstein distance:

**Theorem 4.8** *Let  $\mathcal{P}_t$  be a Markov semigroup over a Polish space  $\mathbf{X}$  admitting a continuous Lyapunov function  $V$ . Suppose furthermore that there exists  $t_* > 0$  and a distance-like function  $d: \mathbf{X} \times \mathbf{X} \rightarrow [0, 1]$  which is contracting for  $\mathcal{P}_{t_*}$  and such that the level set  $\{x \in \mathbf{X} : V(x) \leq 4K_V\}$  is  $d$ -small for  $\mathcal{P}_{t_*}$ . (Here  $K_V$  is as in (4.1).)*

*Then,  $\mathcal{P}_t$  can have at most one invariant probability measure  $\mu_*$ . Furthermore, defining  $\tilde{d}(x, y) = \sqrt{d(x, y)(1 + V(x) + V(y))}$ , there exists  $t > 0$  such that*

$$\tilde{d}(\mathcal{P}_t\mu, \mathcal{P}_t\nu) \leq \frac{1}{2}\tilde{d}(\mu, \nu), \quad (4.6)$$

*for all pairs of probability measures  $\mu$  and  $\nu$  on  $\mathbf{X}$ .*

**Remark 4.9** In the special case  $d = d_{TV}$ , we simply recover Harris' theorem, as stated for example in [HM08b], so that this is a genuinely stronger statement. It is in this sense that Theorem 4.8 is a “weak” version of Harris' theorem where the notion of a “small set” has been replaced by the notion of a  $d$ -small set for a contracting distance-like function  $d$ . The only small difference is that Harris' theorem tells us that the Markov semigroup  $\mathcal{P}_t$  exhibits a spectral gap in a total variation norm weighted by  $1 + V$ , whereas we obtain a spectral gap in a total variation norm weighted by  $1 + \sqrt{V}$ . This is because the proof of Harris' theorem does not require the “close to each other” step (since if  $d(x, y) < 1$ , one has  $x = y$  and the estimate is trivial), so that we never need to apply the Cauchy-Schwarz inequality.

*Proof of Theorem 4.8.* Before we start the proof itself, we note that we can assume without loss of generality that  $t_* > \log(8C_V)/\gamma$ , so that

$$\mathcal{P}_{t_*}V \leq \frac{1}{8}V + K_V. \quad (4.7)$$

This is a simple consequence of the following two facts that can be checked in a straightforward way from the definitions:

- If  $d$  is contracting for two Markov operators  $\mathcal{P}$  and  $\mathcal{Q}$ , then it is also contracting for the product  $\mathcal{P}\mathcal{Q}$ . (Actually it is sufficient for  $d$  to be contracting for  $\mathcal{P}$  and to have (4.5) with  $\alpha = 1$  for  $\mathcal{Q}$ .)
- If a set  $A$  is  $d$ -small for  $\mathcal{Q}$  and  $d$  is contracting for  $\mathcal{P}$ , then  $A$  is also  $d$ -small for  $\mathcal{P}\mathcal{Q}$ .

Note also that the function  $\tilde{d}: \mathcal{M}_1(\mathbf{X}) \times \mathcal{M}_1(\mathbf{X}) \rightarrow \mathbf{R}_+$  is convex in each of its arguments, so that the bound

$$\tilde{d}(\mathcal{P}_t\mu, \mathcal{P}_t\nu) \leq \int_{\mathbf{X} \times \mathbf{X}} \tilde{d}(\mathcal{P}_t(x, \cdot), \mathcal{P}_t(y, \cdot)) \pi(dx, dy),$$

is valid for any coupling  $\pi \in \mathcal{C}(\mu, \nu)$ . As a consequence, in order to show (4.6), it is sufficient to show that it holds in the particular case where  $\mu$  and  $\nu$  are Dirac measures. In other words, it is sufficient to

show that there exists  $t > 0$  and  $\alpha' < 1$  such that

$$\tilde{d}(\mathcal{P}_t(x, \cdot), \mathcal{P}_t(y, \cdot)) \leq \alpha' \tilde{d}(x, y), \quad (4.8)$$

for every  $x, y \in \mathbf{X}$ . Note also that (4.6) is sufficient to conclude that  $\mathcal{P}_t$  can have at most one invariant measure by the following argument. Since  $V$  is a Lyapunov function for  $\mathcal{P}_t$ , it is integrable with respect to any invariant measure so that, if  $\mu$  and  $\nu$  are any two such measures, one has  $\tilde{d}(\mu, \nu) < \infty$ . It then follows immediately from (4.6) and from the invariance of  $\mu, \nu$ , that  $\tilde{d}(\mu, \nu) = 0$ . It follows from the lower semicontinuity of  $\tilde{d}$  that  $\mu = \nu$  as required.

In order to show that (4.8) holds, we make use of a trick similar to the one used in [HM08a]. For  $\beta > 0$  a (small) parameter to be determined later, we set

$$\tilde{d}_\beta(x, y) = \sqrt{d(x, y)(1 + \beta V(x) + \beta V(y))}.$$

Note that, because of the positivity of  $V$ , there exist constants  $c$  and  $C$  (depending on  $\beta$  of course) such that  $c\tilde{d}(x, y) \leq \tilde{d}_\beta(x, y) \leq C\tilde{d}(x, y)$ . As a consequence, if we can show (4.8) for  $\tilde{d}_\beta$  with some value of the parameter  $\beta$ , then it also holds for  $\tilde{d}$  by possibly considering a larger time  $t$ . Just as in [HM08a], we now proceed by showing that  $\beta$  can be tuned in such a way that (4.8) holds, whether  $x$  and  $y$  are “close to each other,” “far from the origin” or “close to the origin.”

**Close to each other.** This is the situation where  $d(x, y) < 1$ , so that

$$\tilde{d}_\beta^2(x, y) = d(x, y)(1 + \beta V(x) + \beta V(y)).$$

In this situation, we make use of the contractivity of  $d$ , the fact that  $V$  is a Lyapunov function, and the Cauchy-Schwarz inequality to obtain

$$\begin{aligned} (\tilde{d}_\beta(\mathcal{P}_{t_*}(x, \cdot), \mathcal{P}_{t_*}(y, \cdot)))^2 &\leq \inf_{\pi} \int d(x', y') \pi(dx', dy') \int (1 + \beta V(x') + \beta V(y')) \pi(dx', dy') \\ &\leq \alpha d(x, y) \left(1 + \frac{\beta}{8}(V(x) + V(y)) + 2\beta K_V\right), \end{aligned}$$

where the infimum runs over all  $\pi \in \mathcal{C}(\mathcal{P}_{t_*}(x, \cdot), \mathcal{P}_{t_*}(y, \cdot))$ . For any given  $\alpha_1 \in (\alpha, 1)$ , we can furthermore choose  $\beta$  sufficiently small such that  $\alpha(1 + 2\beta K_V) \leq \alpha_1$ , so that

$$(\tilde{d}_\beta(\mathcal{P}_{t_*}(x, \cdot), \mathcal{P}_{t_*}(y, \cdot)))^2 \leq \alpha_1 \tilde{d}_\beta^2(x, y).$$

**Far from the origin.** This is the situation where  $d(x, y) \geq 1$  and  $V(x) + V(y) \geq 4K_V$ , so that

$$\tilde{d}_\beta^2(x, y) = 1 + \beta(V(x) + V(y)) \geq 1 + 3\beta K_V + \frac{\beta}{4}(V(x) + V(y)).$$

Using the Lyapunov structure (4.1), we thus get

$$\begin{aligned} (\tilde{d}_\beta(\mathcal{P}_{t_*}(x, \cdot), \mathcal{P}_{t_*}(y, \cdot)))^2 &\leq 1 + 2\beta K_V + C_V \beta e^{-\gamma t_*} (V(x) + V(y)) \leq 1 + 2\beta K_V + \frac{\beta}{8}(V(x) + V(y)) \\ &\leq \max\left\{\frac{1 + 2\beta K_V}{1 + 3\beta K_V}, \frac{1}{2}\right\} \tilde{d}_\beta^2(x, y) =: \alpha_2 \tilde{d}_\beta^2(x, y), \end{aligned}$$

where we made use again of (4.7). While  $\alpha_2$  depends on the choice of  $\beta$ , we see that for any fixed  $\beta > 0$ , one has  $\alpha_2 < 1$ .

**Close to the origin.** This is the final situation where  $d(x, y) = 1$  and  $V(x) + V(y) \leq 4K_V$ , so that  $\tilde{d}(x, y) \geq 1$ . In this case, we make use of the fact that the level set  $\{x : V(x) \leq 4K_V\}$  is assumed to be small to conclude that there exists a coupling  $\pi$  for  $\mathcal{P}_{t_*}(x, \cdot)$  and  $\mathcal{P}_{t_*}(y, \cdot)$  and a constant  $\varepsilon > 0$  such that  $\int d \, d\pi \leq 1 - \varepsilon$ , so that

$$\begin{aligned} (\tilde{d}_\beta(\mathcal{P}_{t_*}(x, \cdot), \mathcal{P}_{t_*}(y, \cdot)))^2 &\leq \int d(x', y') \pi(dx', dy') \int (1 + \beta V(x') + \beta V(y')) \pi(dx', dy') \\ &\leq (1 - \varepsilon)(1 + 2\beta K_V + 2\beta C_V e^{-\gamma t_*}) \leq (1 - \varepsilon)(1 + 4\beta K_V) \tilde{d}(x, y), \end{aligned}$$

where we made again use of (4.7). Here,  $\varepsilon > 0$  is independent of  $\beta$ . Therefore, choosing  $\beta$  sufficiently small (for example  $\beta = \varepsilon/(4K_V)$ ), we can again make sure that the constant appearing in this expression is strictly smaller than 1, thus concluding the proof of Theorem 4.8.  $\square$

**Remark 4.10** If the assumptions of the theorem hold uniformly for  $t_*$  belonging to an open interval of times, then one can check that Theorem 4.8 implies that there exists  $r > 0$  and  $t_0 > 0$  such that the bound

$$\tilde{d}(\mathcal{P}_t \mu, \mathcal{P}_t \nu) \leq e^{-rt} \tilde{d}(\mu, \nu),$$

holds for all  $t > t_0$ , instead of multiples of  $t_*$  only.

If  $d$  is somewhat comparable to a metric, it turns out that we can even infer the *existence* of an invariant measure from the assumptions of Theorem 4.8, just like in the case of Harris' theorem:

**Corollary 4.11** *If there exists a complete metric  $d_0$  on  $\mathbf{X}$  such that  $d_0 \leq \sqrt{d}$  and such that  $\mathcal{P}_t$  is Feller on  $\mathbf{X}$ , then under the assumptions of Theorem 4.8, there exists a unique invariant measure  $\mu_*$  for  $\mathcal{P}_t$ .*

*Proof.* It only remains to show that an invariant measure exists for  $\mathcal{P}_t$ . Fix an arbitrary probability measure  $\mu$  on  $\mathbf{X}$  such that  $\int V \, d\mu < \infty$  and let  $t$  be the time obtained from Theorem 4.8. Since  $\tilde{d} \geq \sqrt{d} \geq d_0$  by assumption and since  $\tilde{d}(\mu, \mathcal{P}_t \mu) < \infty$  by (4.1), it then follows from (4.8) that

$$d_0(\mathcal{P}_{nt} \mu, \mathcal{P}_{(n+1)t} \mu) \leq \tilde{d}(\mathcal{P}_{nt} \mu, \mathcal{P}_{(n+1)t} \mu) \leq \frac{\tilde{d}(\mu, \mathcal{P}_t \mu)}{2^n},$$

so that the sequence  $\{\mathcal{P}_{nt} \mu\}_{n \geq 0}$  is Cauchy in the space of probability measures on  $\mathbf{X}$  endowed with the Wasserstein-1 distance associated to  $d_0$ . Since this space is complete [Vil03], there exists  $\mu_\infty$  such that  $\mathcal{P}_{nt} \mu \rightarrow \mu_\infty$  weakly. In particular, the Feller property of  $\mathcal{P}_t$  implies  $\mathcal{P}_t \mu_\infty = \mu_\infty$  so that, defining  $\mu_*$  by

$$\mu_*(A) = \frac{1}{t} \int_0^t (\mathcal{P}_s \mu_\infty)(A) \, ds,$$

one can check that  $\mathcal{P}_r \mu_* = \mu_*$  for every  $r > 0$  as required.  $\square$

The following result (which we already used in the previous section) relates the contraction property to the conditions in our main result on the convergence of transition probabilities.

**Proposition 4.12** *Let  $(\mathcal{P}_t)$  be a Feller Markov semigroup on  $\mathbf{X}$  and assume that there exists a continuous metric  $d$  which generates the topology of  $\mathbf{X}$  and which is contracting for  $\mathcal{P}_t$  for some  $t > 0$ . Then the second condition in Theorem 2.4 is satisfied for the Markov kernel  $\mathcal{P}_t$ , i.e. there exists a coupling  $\Gamma$  giving positive mass to the “diagonal at infinity”  $\mathcal{D}$ .*

Before we turn to the proof of Proposition 4.12, we give the following result which is essential to settle the measurability questions arising in the proof:

**Lemma 4.13** *Let  $\mathcal{Q}$  be a Feller Markov operator on a Polish space  $\mathbf{X}$  and let  $d$  be a  $[0, 1]$ -valued distance-like function on  $\mathbf{X} \times \mathbf{X}$  which is contracting for  $\mathcal{Q}$ . Then there exists  $\tilde{\alpha} < 1$  and a Markov operator  $\mathcal{T}$  on  $\mathbf{X} \times \mathbf{X}$  such that transition probabilities of  $\mathcal{T}$  are couplings of the transition probabilities for  $\mathcal{Q}$  and such that the inequality*

$$(\mathcal{T}d)(x, y) \leq \tilde{\alpha}d(x, y),$$

holds for every  $(x, y)$  such that  $d(x, y) < 1$ .

*Proof.* Denote as before by  $\mathcal{M}(\mathbf{X} \times \mathbf{X})$  the set of probability measures on  $\mathbf{X} \times \mathbf{X}$  endowed with the topology of weak convergence, so that it is again a Polish space. Let  $\tilde{\alpha} \in (\alpha, 1)$ , where  $\alpha$  is as in Definition 4.6. For every  $(x, y) \in \mathbf{X} \times \mathbf{X}$ , denote

$$F(x, y) = \{\Gamma \in \mathcal{C}(\mathcal{Q}(x, \cdot), \mathcal{Q}(y, \cdot)) : \Gamma(d) \leq \tilde{\alpha}d(x, y)\},$$

and denote by  $\Delta$  the closure in  $\mathbf{X} \times \mathbf{X}$  of the set  $\{d(x, y) < 1\}$ . We know that  $F(x, y)$  is non-empty by assumption whenever  $d(x, y) < 1$ . The Feller property of  $\mathcal{Q}$  then ensures that this is also true for  $(x, y) \in \Delta$ .

The proof of the statement is complete as soon as we can show that there exists a measurable map  $\hat{\mathcal{T}}: \mathbf{X} \times \mathbf{X} \rightarrow \mathcal{M}(\mathbf{X} \times \mathbf{X})$  such that  $\hat{\mathcal{T}}(x, y) \in F(x, y)$  for every  $x, y \in \Delta$ , since it then suffices to set for example

$$\mathcal{T}(x, y; \cdot) = \begin{cases} \hat{\mathcal{T}}(x, y) & \text{if } (x, y) \in \Delta, \\ \mathcal{Q}(x, \cdot) \otimes \mathcal{Q}(y, \cdot) & \text{otherwise.} \end{cases}$$

Since the set  $F(x, y)$  is closed for every pair  $(x, y)$  by the continuity of  $d$ , this follows from the Kuratowski, Ryll-Nardzewski selection theorem [KR65, Wag77] provided we can show that, for every open set  $U \subset \mathcal{M}(\mathbf{X} \times \mathbf{X})$ , the set  $F^{-1}(U) = \{(x, y) : F(x, y) \cap U \neq \emptyset\}$  is measurable.

Since on a Polish space every open set is a countable union of closed sets and since  $F^{-1}(U \cup V) = F^{-1}(U) \cup F^{-1}(V)$  (the same is not true in general for intersections!), the claim follows if we can show that  $F^{-1}(U)$  is measurable for every closed set  $U$ . Under our assumptions,  $F^{-1}(U)$  actually turns out to be closed if  $U$  is closed. To see this, take a convergent sequence  $(x_n, y_n) \in F^{-1}(U)$ . The definition of  $F$  implies that there exist couplings  $\Gamma_n \in \mathcal{C}(\mathcal{Q}(x_n, \cdot), \mathcal{Q}(y_n, \cdot))$  with  $\Gamma_n(d) \leq \tilde{\alpha}d(x_n, y_n)$ . Since  $\mathcal{Q}$  is Feller, the sequence  $\{\Gamma_n\}$  is tight, so that there exists a subsequence converging to a limit  $\Gamma$ . Since  $\Gamma$  belongs to  $\mathcal{C}(\mathcal{Q}(x, \cdot), \mathcal{Q}(y, \cdot))$  and since, by the continuity of  $d$ , we have  $\Gamma(d) \leq \tilde{\alpha}d(x, y)$ ,  $(x, y) \in F^{-1}(U)$  as claimed.  $\square$

*Proof of Proposition 4.12.* By assumption, there exists some  $\alpha \in (0, 1)$  such that  $d(\mathcal{P}_t(x, \cdot), \mathcal{P}_t(y, \cdot)) \leq \alpha d(x, y)$  for all  $x, y \in \mathbf{X}$  which satisfy  $d(x, y) < 1$ . Consider the Markov operator  $\mathcal{Q} := \mathcal{P}_t$ . Let  $\mathcal{T}$  be the Markov operator from Lemma 4.13 so that if, for fixed  $x, y \in \mathbf{X}$ , we denote the corresponding chain starting at  $(x, y)$  by  $(X_n, Y_n)$ , then we have  $\mathbf{E}d(X_1, Y_1) \leq \tilde{\alpha}d(x, y)$  whenever  $d(x, y) < 1$ . Let  $\Gamma_{x,y}$  be the law of  $(X_n, Y_n), n \in \mathbf{N}_0$ . Now we define

$$V_n := \tilde{\alpha}^{-n}d(X_n, Y_n)$$

and  $\tau := \inf\{n \in \mathbf{N} : d(X_n, Y_n) \geq 1\}$ . Then  $V_{n \wedge \tau}, n \in \mathbf{N}_0$  is a non-negative supermartingale and therefore

$$\mathbf{P}\{\tau < \infty\} \leq \mathbf{P}\{\sup_{n \geq 0} V_{n \wedge \tau} \geq 1\} \leq d(x, y),$$

i.e.

$$\Gamma_{x,y}\{d(X_n, Y_n) \leq \tilde{\alpha}^n \text{ for all } n \in \mathbf{N}_0\} \geq 1 - d(x, y).$$

This shows that the second assumption in Theorem 2.4 is satisfied for the chain associated to  $\mathcal{Q}$ .  $\square$

#### 4.1 A consequence of spectral gap results: stability of invariant measures

One may wonder why we made all this effort to ensure that a ‘‘spectral gap’’ result like Theorem 4.8 holds, rather than being content with having simply exponential convergence of transition probabilities to the invariant measure. One payoff is that it allows to very easily obtain stability results of the invariant measure with respect to small perturbations in the Markov semigroup, as we will show in this section.

Assume for the sake of the argument that  $\tilde{d}$  satisfies the triangle inequality (in general it doesn't; see below) and that we have a sequence of ‘‘approximating semigroups’’  $\mathcal{P}_t^\varepsilon$  such that the bound

$$\tilde{d}(\mathcal{P}_t^\varepsilon(x, \cdot), \mathcal{P}_t(x, \cdot)) \leq \varepsilon C(t) \tilde{V}(x),$$

holds, where  $C$  is a function that is bounded on bounded subsets of  $\mathbf{R}$  and  $\tilde{V}$  is some positive function  $\tilde{V}: \mathbf{X} \rightarrow \mathbf{R}_+$ .

Let now  $\mu_*$  denote the invariant measure for  $\mathcal{P}_t$  and  $\mu_*^\varepsilon$  an invariant measure for  $\mathcal{P}_t^\varepsilon$  (which need not be unique). Choosing  $t$  as in (4.6), one then has the bound

$$\begin{aligned} \tilde{d}(\mu_*, \mu_*^\varepsilon) &= \tilde{d}(\mathcal{P}_t \mu_*, \mathcal{P}_t^\varepsilon \mu_*^\varepsilon) \leq \tilde{d}(\mathcal{P}_t \mu_*, \mathcal{P}_t \mu_*^\varepsilon) + \tilde{d}(\mathcal{P}_t \mu_*^\varepsilon, \mathcal{P}_t^\varepsilon \mu_*^\varepsilon) \\ &\leq \frac{1}{2} \tilde{d}(\mu_*, \mu_*^\varepsilon) + \varepsilon C(t) \int_{\mathbf{X}} \tilde{V}(x) \mu_*^\varepsilon(dx), \end{aligned}$$

from which we deduce that  $\tilde{d}(\mu_*, \mu_*^\varepsilon) \leq 2\varepsilon C(t) \int_{\mathbf{X}} \tilde{V}(x) \mu_*^\varepsilon(dx)$ . If one can obtain an *a priori* bound on  $\mu_*^\varepsilon$  that ensures that  $\int_{\mathbf{X}} \tilde{V}(x) \mu_*^\varepsilon(dx)$  is bounded independently of  $\varepsilon$ , this shows that the distance between the invariant measures for  $\mathcal{P}_t$  and  $\mathcal{P}_t^\varepsilon$  is comparable to the distance between the transition probabilities for some fixed time  $t$ .

This argument is still valid if the ‘‘distance function’’  $\tilde{d}$  satisfies a weak form of the triangle inequality, i.e. if there exists a positive constant  $K > 0$  such that

$$\tilde{d}(x, y) \leq K(\tilde{d}(x, z) + \tilde{d}(z, y)), \quad (4.9)$$

for every  $x, y, z \in \mathbf{X}$ . This turns out to be often satisfied in practice, due to the following result:

**Lemma 4.14** *Let  $d: \mathbf{X} \times \mathbf{X} \rightarrow [0, 1]$  be a distance-like function and assume that there exists a constant  $K_d$  such that*

$$d(x, y) \leq K_d(d(x, z) + d(z, y)), \quad (4.10)$$

*holds for every  $x, y, z \in \mathbf{X}$ . Assume furthermore that  $V: \mathbf{X} \rightarrow \mathbf{R}_+$  is such that there exist constants  $c, C$  such that the implication*

$$d(x, z) \leq c \quad \Rightarrow \quad V(x) \leq CV(z) \quad (4.11)$$

*holds. Then, there exists a constant  $K$  such that (4.9) holds for  $\tilde{d}$  defined as in Theorem 4.8.*

*Proof.* Note first that it is sufficient to show that there exists a constant  $K$  such that

$$d(x, y)(1 + V(x) + V(y)) \leq K(d(x, z)(1 + V(x) + V(z)) + d(z, y)(1 + V(z) + V(y))). \quad (4.12)$$



Since  $\tilde{d}$  is symmetric, we can assume without loss of generality that  $V(x) \geq V(y)$ . We consider the following two cases

If  $d(x, z) \geq c$ , then the boundedness of  $d$  implies the existence of a constant  $\tilde{C}$  such that

$$d(x, y)(1+V(x)+V(y)) \leq \tilde{C}(1+V(x)+V(y)) \leq \frac{\tilde{C}}{c}d(x, z)(1+V(x)+V(y)) \leq \frac{2\tilde{C}}{c}d(x, z)(1+V(x)),$$

from which (4.12) follows with  $K = 2\tilde{C}/c$ .

If  $d(x, z) \leq c$ , we make use of our assumptions (4.10) and (4.11) to deduce that

$$\begin{aligned} d(x, y)(1+V(x)+V(y)) &\leq K_d(d(x, z) + d(z, y))(1+V(x)+V(y)) \\ &\leq 2K_d d(x, z)(1+V(x)) + 2CK_d d(z, y)(1+V(z)), \end{aligned}$$

from which (4.12) follows with  $K = 2K_d(1 \vee C)$ .  $\square$

**Remark 4.15** If  $\mathbf{X}$  is a Banach space and  $d(x, y) = 1 \wedge \|x - y\|$ , then Lemma 4.14 essentially states that  $\tilde{d}$  satisfies (4.9) provided that  $V(x)$  grows at most exponentially with  $\|x\|$ .

## 5 Application of the spectral gap result to SDDEs

In this section, we apply the abstract results from the previous section to the problem of exponential convergence to an invariant measure for the type of stochastic delay equations considered earlier. The main problem will turn out to be to find a distance-like function  $d$  which is contracting. In order to obtain an exponential convergence result, we will have to assume, just like in the case of Harris chains [MT93, HM08b] some Lyapunov structure. We therefore introduce the following assumption:

**Assumption 5.1** *There exists a continuous function  $V: \mathcal{C} \rightarrow \mathbf{R}_+$  such that  $\lim_{\|X\| \rightarrow \infty} V(X) = +\infty$  and such that there exist strictly positive constants  $C_V, \gamma$  and  $K_V$  such that the bound*

$$\mathbf{E}V(X_t) \leq C_V e^{-\gamma t} V(X_0) + K_V,$$

*holds for solutions to (3.1) with arbitrary initial conditions  $X_0 \in \mathcal{C}$ .*

**Remark 5.2** If  $g$  is bounded, then Assumption 5.1 is satisfied if  $\langle f(x), x(0) \rangle \leq C - c|x(0)|^2$  for some  $c > 0$ , by setting  $V(X) = |X(0)|^2$ . There are however many examples (see for example [RBT02, Hai09] in the case without delay), where finding an appropriate Lyapunov function is an art by itself that requires a thorough understanding of the dynamic and has no ‘‘one fits all’’ recipe.

The distance-like function  $d$  that we are going to use in this section is given by

$$d(X, Y) = 1 \wedge \delta^{-1} \|X - Y\|, \tag{5.1}$$

for a suitable (small) constant  $\delta$  to be determined later. We start by verifying that bounded sets are  $d$ -small for every value of  $\delta$  and we will then proceed to showing that under suitable assumptions, it is possible to find  $\delta > 0$  such that  $d$  is also contracting.

### 5.1 Bounded sets are $d$ -small

**Proposition 5.3** *Let the assumptions of Theorem 3.1 be satisfied, let  $d$  be as in (5.1) and let  $t \geq 2r$  be arbitrary. Then every bounded set is  $d$ -small for  $\mathcal{P}_t$ .*

*Proof.* Fix  $t \geq 2r$ . We show that every closed ball  $\mathbf{B}_R \subset \mathcal{C}$  with center 0 and radius  $R$  is  $d$ -small for  $\mathcal{P}_t$ . By Lemma 3.8, we know that  $p := \inf_{x \in \mathbf{B}_R} \mathcal{P}_t(x, \mathbf{B}_{\delta/4}) > 0$ . Let  $x, y \in \mathbf{B}_R$  and let  $X$  and  $Y$  be solutions of (3.1) with initial conditions  $x$  and  $y$  respectively. We couple  $X$  and  $Y$  independently. Then

$$\begin{aligned} d(\mathcal{P}_t(x, \cdot), \mathcal{P}_t(y, \cdot)) &\leq \mathbf{E}(1 \wedge (\delta^{-1} \|X_t - Y_t\|)) \\ &\leq \mathbf{P}(\{X_t \notin \mathbf{B}_{\delta/4}\} \cup \{Y_t \notin \mathbf{B}_{\delta/4}\}) + \frac{1}{2} \mathbf{P}\{X_t \in \mathbf{B}_{\delta/4}, Y_t \in \mathbf{B}_{\delta/4}\} \leq 1 - \frac{1}{2} p^2, \end{aligned}$$

for all  $x, y \in \mathbf{B}_R$ , so the claim follows.  $\square$

### 5.2 The distance $d$ is contracting

Before we start, we give the following a priori estimate that shows that trajectories of (3.1) driven by the same realisation of the noise cannot separate too rapidly. More precisely, we have:

**Proposition 5.4** *Let the assumptions of Theorem 3.1 be satisfied. There exists  $\kappa > 0$  such that the bound*

$$\mathbf{E}\|X_t - \tilde{X}_t\|^4 \leq e^{\kappa(1+t)^2} \|X_0 - \tilde{X}_0\|^4,$$

holds for all  $t \geq 0$  and any pair of initial conditions  $X_0, \tilde{X}_0 \in \mathcal{C}$ .

*Proof.* The proof is similar to the argument used in the proof of Theorem 3.1. Setting  $Z(t) = X(t) - \tilde{X}(t)$ , we have the bound

$$d|Z(t)|^2 = 2\langle f(X_t) - f(\tilde{X}_t), Z(t) \rangle dt + \|g(X_t) - g(\tilde{X}_t)\|^2 dt + dM(t) \leq K\|Z_t\|^2 dt + dM(t),$$

where  $M$  is a martingale with quadratic variation process bounded by  $C \int_0^t \|Z_s\|^4 ds$ . Defining  $M^*(t) = \sup_{s \leq t} M(s)$ , we thus obtain the bound

$$\|Z_t\|^2 \leq \|Z_0\|^2 + K \int_0^t \|Z_s\|^2 ds + M^*(t),$$

so that

$$\begin{aligned} \mathbf{E}\|Z_t\|^4 &\leq 3\mathbf{E}\left(\|Z_0\|^4 + K^2 \left(\int_0^t \|Z_s\|^2 ds\right)^2 + (M^*(t))^2\right) \\ &\leq 3\left(\|Z_0\|^4 + K^2 t \int_0^t \mathbf{E}\|Z_s\|^4 ds + C \int_0^t \mathbf{E}\|Z_s\|^4 ds\right), \end{aligned}$$

where we used the Burkholder-Davis-Gundy inequality [RY91] in order to bound the expectation of  $(M^*)^2$ . The claim follows from Gronwall's lemma.  $\square$

In this subsection, we show one possible way of verifying that  $d$  is contracting that is suited to our problem. This is by far not the only one. One can check for example that the procedure followed in [HM08a] allows to construct a contracting distance for the degenerate 2D stochastic Navier-Stokes

equations by using a gradient bound on the semigroup. A general version of this argument is presented in Section 5.3 below. For the problem at hand, it seems however more appropriate and technically straightforward to consider a “binding construction” in the terminology of [MY02, Hai02, Mat02b].

We fix two initial conditions  $X_0, \tilde{X}_0 \in \mathcal{C}$  and consider the construction from Section 3. We fix some  $\gamma_0 > 0$  and choose  $\lambda$  sufficiently large so that the conclusion of Lemma 3.6 holds. As in the proof of Theorem 3.1, we also introduce the stopping time  $\tau = \inf\{t > 0 : \int_0^t |v(s)|^2 ds \geq \varepsilon^{-1} \|X_0 - \tilde{X}_0\|^2\}$ , where  $v$  is as in the proof of Theorem 3.1. (Note that the value of  $\varepsilon$  is not necessarily that from Section 3, but will be determined later.) We also define  $\tilde{v}$  by  $\tilde{v}(s) = v(s)\mathbf{1}_{\tau > s}$ .

This defines a map  $\Psi$  from  $\Omega := \mathcal{C}([0, \infty), \mathbf{R}^m)$  to itself by  $\Psi(w) = w + \int_0^\cdot \tilde{v}(s) ds$  (the map  $\Psi$  furthermore depends on the initial conditions  $X_0$  and  $\tilde{X}_0$ , but we suppress this dependence from the notation). The image  $\tilde{\mathbf{P}}$  of Wiener measure  $\mathbf{P}$  under  $\Psi$  has a density  $\mathcal{D}(w) = d\tilde{\mathbf{P}}/d\mathbf{P}$ .

The aim of introducing the cutoff is that if we define  $\tilde{\mathcal{D}}(w) = 1/\mathcal{D}(w)$ , we obtain “for free” bounds of the type

$$\int (1 - \mathcal{D}(w))^2 \mathbf{P}(dw) \leq C\varepsilon^{-1} \|X_0 - \tilde{X}_0\|^2, \quad \int (1 - \tilde{\mathcal{D}}(w))^2 \tilde{\mathbf{P}}(dw) \leq C\varepsilon^{-1} \|X_0 - \tilde{X}_0\|^2,$$

for some constant  $C > 0$ , provided that we restrict ourselves to pairs of initial conditions such that

$$\|X_0 - \tilde{X}_0\|^2 \leq \varepsilon. \quad (5.2)$$

Had we not introduced the cut-off, we would need to get exponential integrability of  $v$  first.

The map  $\Psi$  allows to construct, for any two initial conditions  $X_0$  and  $\tilde{X}_0$ , a coupling for  $\mathbf{P}$  with itself in the following way. Define the map  $\tilde{\Psi}: \Omega \rightarrow \Omega \times \Omega$  by  $\tilde{\Psi}(w) = (w, \Psi(w))$ , denote by  $\pi^i$  the projection onto the  $i$ th component of  $\Omega \times \Omega$ , and set

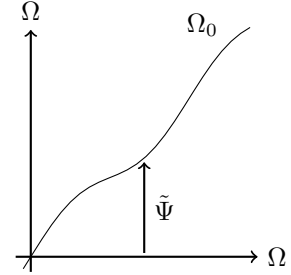
$$\begin{aligned} \Pi_0(dw_1, dw_2) &= (1 \wedge \tilde{\mathcal{D}}(w_2))(\tilde{\Psi}_\# \mathbf{P})(dw_1, dw_2), \\ \Pi(dw_1, dw_2) &= \Pi_0(dw_1, dw_2) + Z^{-1}(\mathbf{P} - \pi_\#^1 \Pi_0)(dw_1)(\mathbf{P} - \pi_\#^2 \Pi_0)(dw_2), \end{aligned}$$

where  $Z = 1 - \Pi_0(\Omega \times \Omega) = \frac{1}{2} \|\mathbf{P} - \tilde{\mathbf{P}}\|_{\text{TV}}$  is a suitable constant. One can check that  $\Pi$  as defined above is a coupling for  $\mathbf{P}$  and  $\tilde{\mathbf{P}}$ . Furthermore, it is designed in such a way that it maximises the mass of the set  $\Omega_0 = \{(w, w') : w' = \Psi(w)\}$ . We claim that this coupling is designed in such a way that its image under the product solution map of (3.1) allows to verify that  $d$  as in (5.1) is contracting for some sufficiently small value of  $\delta > 0$  to be determined later. Note that, since the bound (4.6) only needs to be checked for pairs of initial conditions with  $d(X_0, \tilde{X}_0) < 1$ , the constraint (5.2) is satisfied provided that we make sure that  $\delta^2 \leq \varepsilon$ .

In order to see that this is indeed the case, we fix a terminal time  $t$  and we break the space  $\Omega \times \Omega = \Omega_1 \cup \Omega_2 \cup \Omega_3$  into three parts:

$$\begin{aligned} \Omega_1 &= \{(w, w') : w' = \Psi(w) \ \& \ \tau(w) \geq t\}, \\ \Omega_2 &= \{(w, w') : w' = \Psi(w) \ \& \ \tau(w) < t\}, \\ \Omega_3 &= \{(w, w') : w' \neq \Psi(w)\}. \end{aligned}$$

Here, we made use of the stopping time  $\tau$  defined at the beginning of this section. We consider the set  $\Omega_1$  as being the set of “good” realisations and we will show that  $\Omega_1$  has high probability. The contributions from the other two sets will be considered as error terms.



Consider now a pair  $(X_0, \tilde{X}_0)$  of initial conditions such that  $d(X_0, \tilde{X}_0) < 1$ , which is to say that  $\|X_0 - \tilde{X}_0\| < \delta$ . Denote by  $X_t$  and  $\tilde{X}_t$  the solutions to (3.1) driven by the noise realisations  $w$  and  $w'$  respectively. We then have

$$\begin{aligned} \int_{\Omega_1} d(X_t(w), \tilde{X}_t(w')) \Pi(dw, dw') &\leq \delta^{-1} \int_{\Omega_1} \|X_t(w) - \tilde{X}_t(w')\| \Pi(dw, dw') \\ &\leq \delta^{-1} \int_{\Omega} \|X_t(w) - \tilde{X}_t(\Psi(w))\| \mathbf{P}(dw) \leq C\delta^{-1} e^{-\gamma_0 t} \|X_0 - \tilde{X}_0\| \\ &= C e^{-\gamma_0 t} d(X_0, \tilde{X}_0), \end{aligned}$$

where we made use of the bounds obtained in Lemma 3.6. Regarding the integral over  $\Omega_2$ , we combine Lemma 3.6, Proposition 5.4 and the strong Markov property to conclude that

$$\begin{aligned} \int_{\Omega_2} d(X_t(w), \tilde{X}_t(w')) \Pi(dw, dw') &\leq \delta^{-1} \int_{\tau < t} \|X_t(w) - \tilde{X}_t(\Psi(w))\| \mathbf{P}(dw) \\ &\leq \delta^{-1} \left( \int_{\tau < t} \|X_t(w) - \tilde{X}_t(\Psi(w))\|^2 \mathbf{P}(dw) \right)^{1/2} \sqrt{\mathbf{P}(\tau < t)} \\ &\leq \delta^{-1} \mathbf{E} \left( C e^{-\gamma_0 \tau} e^{\kappa(1+t-\tau)^2} \right) \|X_0 - \tilde{X}_0\| \sqrt{\mathbf{P}(\tau < t)} \\ &\leq C e^{\kappa(1+t)^2} d(X_0, \tilde{X}_0) \sqrt{\mathbf{P}(\tau < t)}. \end{aligned}$$

At this stage, we combine Lemma 3.6 with Chebychev to conclude that

$$\mathbf{P}(\tau < t) \leq \mathbf{P} \left( \int_0^\infty |v(s)|^2 ds \geq \varepsilon^{-1} \|X_0 - \tilde{X}_0\|^2 \right) \leq C\varepsilon,$$

for some constant  $C$  independent of  $t$  and the pair  $(X_0, \tilde{X}_0)$ . Finally, we obtain the bound

$$\begin{aligned} \int_{\Omega_3} d(X_t(w), \tilde{X}_t(w')) \Pi(dw, dw') &\leq \Pi(\Omega_3) = \int_{\Omega} (1 - 1 \wedge \tilde{D}(w)) \tilde{\mathbf{P}}(dw) \\ &= \int_{\Omega} (0 \vee (1 - \tilde{D}(w))) \tilde{\mathbf{P}}(dw) \leq \left( \int_{\Omega} (1 - \tilde{D}(w))^2 \tilde{\mathbf{P}}(dw) \right)^{1/2} \\ &\leq C\varepsilon^{-1/2} \|X_0 - \tilde{X}_0\| \leq C\delta\varepsilon^{-1/2} d(X_0, \tilde{X}_0). \end{aligned}$$

The required bound follows by first taking  $\varepsilon$  small enough and then taking  $\delta$  small enough.

### 5.3 Construction of contracting distances for SPDEs

Finally, we want to show how the existence of a contracting distance for a Markov semigroup  $\mathcal{P}_t$  can be verified in the case of stochastic PDEs. This is very similar to the calculation performed in [HM08a], but it has the advantage of not being specific to the Navier-Stokes equations. Recall that [HM08c] yields conditions under which the Markov semigroup (over some separable Hilbert space  $\mathcal{H}$ ) generated by a class of stochastic PDEs satisfies the following gradient bound for every function  $\varphi \in \mathcal{C}^1(\mathcal{H}, \mathbf{R})$ :

$$\|D\mathcal{P}_t\varphi(X)\| \leq W(X) \left( e^{-\tilde{\gamma}t} \sqrt{(\mathcal{P}_t\|D\varphi\|^2)(X)} + C\|\varphi\|_\infty \right). \quad (5.3)$$

Here,  $W: \mathcal{H} \rightarrow \mathbf{R}_+$  is some continuous function that controls the regularising properties of  $\mathcal{P}_t$  and  $\tilde{\gamma}, C$  are some strictly positive constants. It turns out that if the semigroup  $\mathcal{P}_t$  has sufficiently good dissipativity properties with respect to  $W$ , then one can find a contracting distance function for it.

Before we state the result, let us define a family of “weighted metrics”  $\varrho_p$  on  $\mathcal{H}$  by

$$\varrho_p(X, Y) = \inf_{\gamma: X \rightarrow Y} \int_0^1 W^p(\gamma(t)) \|\dot{\gamma}(t)\| dt ,$$

where the infimum runs over all Lipschitz continuous paths  $\gamma: [0, 1] \rightarrow \mathcal{H}$  connecting  $X$  to  $Y$ . With this notation at hand, we have:

**Proposition 5.5** *Let  $\{\mathcal{P}_t\}_{t \geq 0}$  be a Markov semigroup over a separable Hilbert space  $\mathcal{H}$  satisfying the bound (5.3) for some continuous function  $W: \mathcal{H} \rightarrow [1, \infty)$ . Assume furthermore that there exists  $p > 1$ , a time  $t_* > 0$  and a constant  $\tilde{C} > 0$  such that the bound*

$$\mathcal{P}_t W^{2p} \leq \tilde{C} W^{2p-2} , \quad (5.4)$$

*holds for every  $t \geq t_*$ . (In other words,  $W$  is a kind of super-Lyapunov function for  $\mathcal{P}_t$ .) Then, there exists  $\delta > 0$  and  $T > 0$  such that the metric  $d(X, Y) = 1 \wedge \delta^{-1} \varrho_p(X, Y)$  is contracting for  $\mathcal{P}_T$ .*

*Proof.* By Monge-Kantorowitch duality, it is sufficient to show that there exist  $T > 0$  and  $\delta > 0$  such that the bound

$$|\mathcal{P}_T \varphi(X) - \mathcal{P}_T \varphi(Y)| \leq \frac{\varrho_p(X, Y)}{2\delta} , \quad (5.5)$$

holds for every  $\mathcal{C}^1$  function  $\varphi: \mathcal{H} \rightarrow \mathbf{R}$  which has Lipschitz constant 1 with respect to  $d$ . Note now that such a function  $\varphi$  satisfies

$$|\varphi(X)| \leq \frac{1}{2} , \quad \|D\varphi(X)\| \leq \frac{W^p(X)}{\delta} .$$

In particular, it follows from the gradient bound (5.3) combined with (5.4) that for  $T \geq t_*$ , one has

$$\|D\mathcal{P}_T \varphi(X)\| \leq W(X) \left( \delta^{-1} e^{-\tilde{\gamma}T} \sqrt{\tilde{C}} W^{p-1}(X) + \frac{C}{2} \right) .$$

Choosing  $T$  sufficiently large and  $\delta$  sufficiently small, we see that it is possible to ensure that

$$\|D\mathcal{P}_T \varphi(X)\| \leq \frac{W^p(X)}{2\delta} .$$

Since, on the other hand, for any path  $\gamma$  connecting  $X$  to  $Y$  we have

$$|\mathcal{P}_T \varphi(X) - \mathcal{P}_T \varphi(Y)| \leq \int_0^1 \|D\mathcal{P}_T \varphi(\gamma(s))\| \|\dot{\gamma}(s)\| ds ,$$

the requested bound (5.5) follows at once.  $\square$

## References

- [BCG08] Bakry, D., Cattiaux, P. and Guillin, A. (2008) Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.* **254**, 3, 727–759.
- [BM05] Bakhtin, Y., and Mattingly, J. (2005). Stationary solutions of stochastic differential equations with memory and stochastic partial differential equations, *Communications in Contemporary Mathematics*, **7**, 553–582, World Scientific.

- [Bas98] Bass, R. (1998). *Diffusions and Elliptic Operators*. Springer, New York.
- [BKL01] Bricmont, J., Kupiainen, A., and Lefevere, R. (2001). Ergodicity of the 2D Navier-Stokes equations with random forcing, *Comm. Math. Phys.* **224**, 1, 65–81.
- [CW00] Cranston, M., and Wang, F.-Y. (2002). A condition for the equivalence of coupling and shift coupling, *Annals of Probability*, **28**, 4, 1666–1679.
- [DFG09] Douc, R., Fort, G. and Guillin, A. (2009) Subgeometric rates of convergence of  $f$ -ergodic strong Markov processes. *Stoch. Proc. Appl.*, **119**, 3, 897–923.
- [DZ92] Da Prato, G., and Zabczyk, J. (1992). *Stochastic Equations in Infinite Dimensions*. Cambridge University Press, Cambridge.
- [EGS09] Es-Sarhir, A., v. Gaans, O., and Scheutzow, M. (2009). Invariant measures for stochastic functional differential equations with superlinear drift term, *Diff. and Int. Equ.* To appear.
- [EH01] Eckmann, J.-P. and Hairer, M. (2001). Uniqueness of the invariant measure for a stochastic PDE driven by degenerate noise, *Comm. Math. Phys.* **219**, 3 523–565.
- [EMS01] E, W., Mattingly, J. C., and Sinai, Y. (2001). Gibbsian dynamics and ergodicity for the stochastically forced Navier-Stokes equation, *Comm. Math. Phys.* **224**, 1, 83–106.
- [Hai02] Hairer, M. (2002). Exponential mixing properties of stochastic PDEs through asymptotic coupling, *Prob. Theory Rel. Fields* **124**, 345–380.
- [Hai09] Hairer, M. (2009). How hot can a heat bath get?, *Comm. Math. Phys.* **292**, 1, 131–177.
- [HM06] Hairer, M., and Mattingly, J. (2006). Ergodicity of the 2D Navier-Stokes equations with degenerate stochastic forcing, *Annals of Mathematics* **164**, 993–1032.
- [HM08a] Hairer, M., and Mattingly, J. (2008). Spectral gaps in Wasserstein distances and the 2D stochastic Navier-Stokes equations, *Annals of Probability*, **36**, 6, 2050–2091.
- [HM08b] Hairer, M., and Mattingly, J. (2008). Yet another look at Harris' ergodic theorem for Markov chains, *Submitted*.
- [HM08c] Hairer, M., and Mattingly, J. (2008). A theory of hypoellipticity and unique ergodicity for semilinear stochastic PDEs, *Submitted*.
- [IN64] Itô, K., and Nisio, M. (1964). On stationary solutions of a stochastic differential equation, *J. Math. Kyoto Univ.* **4**, 1–75.
- [KR65] Kuratowski, K. and Ryll-Nardzewski, C. (1965). A general theorem on selectors, *Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys.* **13**, 397–403.
- [KS02] Kuksin, S. and Shirikyan, A. (2002). Coupling approach to white-forced nonlinear PDEs, *J. Math. Pures Appl. (9)* **81**, 6, 567–602.
- [Lin92] Lindvall, T. (1992). *Lectures on the coupling method*. John Wiley & Sons Inc., New York.
- [MY02] Masmoudi, N. and Young, L.-S. (2002). Ergodic theory of infinite dimensional systems with applications to dissipative parabolic PDEs. *Comm. Math. Phys.* **227**, 3, 461–481.
- [Mat02b] Mattingly, J. C. Exponential convergence for the stochastically forced Navier-Stokes equations and other partially dissipative dynamics. *Comm. Math. Phys.* **230**, 3, (2002), 421–462.
- [Mat07] Mattingly, J. C. *Saint Flour Lectures*. In preparation, Summer 2007 version from summer school.
- [Moh86] Mohammed, S.E.A. (1986). Non-linear flows of stochastic linear delay equations. *Stochastics* **17**, 207–213.
- [MS97] Mohammed, S.E.A., and Scheutzow, M. (1997). Lyapunov exponents of linear stochastic functional differential equations driven by semimartingales. Part II: examples and case studies, *Annals of Probability* **25**, 1210–1240.

- [MT93] Meyn, S.P. and Tweedie, R.L. (1993). *Markov chains and stochastic stability*. Springer, London.
- [RRG06] Reiß, M., Riedle, M., and van Gaans, O. (2006). Delay differential equations driven by Lévy processes: stationarity and Feller properties, *Stochastic Processes and their Applications* **116**, 1409–1432.
- [RS08] v. Renesse, M. and Scheutzow, M. (2008). Existence and uniqueness of solutions of stochastic functional differential equations, *Submitted*.
- [RY91] Revuz, D. and Yor, M. (1991). *Continuous Martingales and Brownian Motion*. Grundlehren der mathematischen Wissenschaften, 293. Springer-Verlag, Berlin.
- [RBT02] Rey-Bellet, L. and Thomas, L. (2002). Exponential convergence to non-equilibrium stationary states in classical statistical mechanics, *Comm. Math. Phys.* **225**, 2, 305–329.
- [Sch84] Scheutzow, M. (1984). Qualitative behaviour of stochastic delay equations with a bounded memory, *Stochastics* **12**, 41–80.
- [Sch05] Scheutzow, M. (2005). Exponential growth rates for stochastic delay differential equations, *Stochastics and Dynamics* **5**, 163–174.
- [SS02] Scheutzow, M., and Steinsaltz, D. (2002). Chasing balls through martingale fields, *Annals of Probability*, **30**, 2046–2080.
- [Ver99] Veretennikov, A. Yu. (1999) On polynomial mixing and the rate of convergence for stochastic differential and difference equations, *Teor. Veroyatnost. i Primenen.*, **44**, 312–327
- [Vil03] Villani, C. (2003). *Topics in optimal transportation*. Graduate Studies in Mathematics, 58. American Mathematical Society, Providence, RI.
- [Wag77] Wagner, D. H. (1977). Survey of measurable selection theorems, *SIAM J. Control Optimization* **15**, 5, 859–903.