Noise induced resonance in bistable systems caused by delay feedback*

Markus Fischer
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Mohrenstr. 39
10117 Berlin
Germany

Peter Imkeller
Institut für Mathematik
Humboldt-Universität zu Berlin
Unter den Linden 6
10099 Berlin
Germany

August 7, 2005

Abstract

The subject of the present paper is a simplified model for a symmetric bistable system with memory or delay, the reference model, which in the presence of noise exhibits a phenomenon similar to what is known as stochastic resonance. The reference model is given by a one dimensional parametrized stochastic differential equation with point delay, basic properties whereof we check.

With a view to capturing the effective dynamics and, in particular, the resonance-like behavior of the reference model we construct a simplified or reduced model, the two state model, first in discrete time, then in the limit of discrete time tending to continuous time. The main advantage of the reduced model is that it enables us to explicitly calculate the distribution of residence times which in turn can be used to characterize the phenomenon of noise-induced resonance.

Drawing on what has been proposed in the physics literature, we outline a heuristic method for establishing the link between the two state model and the reference model. The resonance characteristics developed for the reduced model can thus be applied to the original model.

2000 AMS subject classifications: primary 34K50, 60H10; secondary 60G17, 34K11, 34K13, 34K18, 60G10.

Key words and phrases: stochastic differential equation; delay differential equation; stochastic resonance; effective dynamics; Markov chain; stationary process; stochastic synchronization.

1 Introduction

Stochastic resonance in a narrower sense is the random amplification of a weak periodic signal induced by the presence of noise of low intensity such that the signal amplification is maximal at a certain optimal non-zero level of noise. In addition to weak additive noise and a weak periodic input signal there is a third ingredient in systems where stochastic resonance can occur, namely a threshold or a barrier that induces several (in our case two) macroscopic states in the output signal.

Consider a basic, yet fundamental example. Let V be a symmetric one dimensional double well potential. A common choice for V is the standard quartic potential, see Fig. 1 a). The barrier mentioned

*This work was partially supported by the DFG research center Matheon (FZT 86) in Berlin and the DFG Sonderforschungsbereich 649 "Economic Risk".

above is in this case the potential barrier of V separating the two local minima. Assume that the periodic input signal is sinusoidal and the noise white. The output of such a system is described by the stochastic differential equation (SDE)

$$dX(t) \;=\; -\Big(V'\big(X(t)\big) + a\cdot\sin\big(\tfrac{2\pi}{T}t\big)\Big)dt \;+\; \sigma\cdot dW(t), \qquad t\geq 0, \tag{1}$$

where W is a standard one dimensional Wiener process, $\sigma \geq 0$ a noise parameter, V' the first order derivative of the double well potential V, $a \geq 0$ the amplitude and $T > 0$ the period of the input signal.

As an alternative to the system view, Equation (1) can be understood as describing the overdamped motion of a small particle in the potential landscape V in the presence of noise and under the influence of an exterior periodic force. It was originally proposed by Benzi et al. [1, 2] and Nicolis [3] as an energy balance model designed to explain the succession of ice and warm ages in paleoclimatic records as a phenomenon of quasi periodicity in the average global temperature on Earth.

If $a = 0$, i. e. in the absence of a periodic signal, Equation (1) reduces to an autonomous SDE which has two metastable states corresponding to the two local minima of V. With $\sigma > 0$ sufficiently small, the diffusion will spend most of its time near the positions of these minima. In the presence of weak noise, there are two distinct time scales, a short one corresponding to the quadratic variation of the Wiener process, and a long one proportional to the average time it takes the diffusion to travel from one of the metastable states to the other.

The fact that the time scale induced by the noise process is small in comparison with the mean residence time as σ tends to zero should allow us to disregard small intrawell fluctuations when we are interested in the interwell transition behavior.

Suppose $a > 0$ small enough so that there are no interwell transitions in case $\sigma = 0$, i. e. in the deterministic case. The input signal then slightly and periodically tilts the double well potential V. We now have two different mean residence times, namely the average time the particle stays in the shallow well and the average time of residence in the deep well. Of course, both time scales also depend on the noise intensity.

Notice that deep and shallow well change roles every half period $\frac{T}{2}$. Given a sufficiently long period T, the noise intensity can now be tuned in such a way as to render the occurrence

of transitions from the shallow to the deep well probable within one half period, while this time span is too short for the occurrence of transitions in the opposite direction. At a certain noise level the output signal will exhibit quasi periodic transition behavior, thereby inducing an amplification of the input signal.

For a more comprehensive description of stochastic resonance and its wide field of applications in many fields of science and engineering see Gammaitoni et al. [4] or Anishchenko et al. [5]. Very recently it has been used in economics related models designed for the explanation of non-linear market phenomena such as crashes and bubbles, see Krawiecki and Holyst [6]. In their model, the external periodic force corresponds to a weak external information carrying signal. What models that exhibit stochastic resonance have in common is the quasi periodicity of the output at a certain non-zero noise level. More generally, stochastic resonance is an instance of noise-induced order.

In view of the fact that the system given by Equation (1) can work as a random amplifier it seems natural to take the frequency spectrum of the output signal as basis for a measure of resonance. The most common measure of this kind is the spectral power amplification (SPA) coefficient. Another measure of resonance based on the frequency spectrum is the signal to noise ratio (SNR). For a detailed analysis see Pavlyukevich [7].

In general, when measuring stochastic resonance, it is assumed that the solution is in a "stationary regime". Since Equation (1) is time dependent for $a > 0$ we cannot expect $(X(t))_{t\geq 0}$ to be a stationary process. Transforming the non-autonomous SDE (1) into an autonomous SDE with state space $\mathbb{R} \times S^1$, one can recover the time homogeneous Markov property and a unique invariant probability measure exists, cf. Imkeller and Pavlyukevich [8]. In Section 3 we will make use of the same idea of appropriately enlarging the state space in order to regain a time homogeneous Markov model.

A different starting point for a measure of resonance – the one that will be adopted here – is the distribution of intrawell residence times. Observe that the roles of the two potential wells are interchangeable.

A third class of measures of resonance is provided by methods of quantifying (un-)certainty, in particular by the entropy of a distribution. This agrees well with the view of stochastic resonance as an instance of noise-induced order.

The fact that with $\sigma > 0$ and $a > 0$ small a typical solution to (1) spends most of its time near the positions of the two minima of the double well potential V suggests to identify the two potential wells with their respective minima. The state space $\mathbb{R}$ of the non-autonomous SDE thus gets reduced to two states, say -1 and 1, corresponding to the left and the right well, respectively.

According to an idea of McNamara and Wiesenfeld [9] the effective dynamics of Equation (1) can be captured in the two state model by constructing a $\{-1, 1\}$-valued time inhomogeneous Markov chain with certain (time dependent) transition rates. These rates are determined as the rates of escape from the potential well of the tilted double well potential which corresponds to the reduced state in question. An approximation of the rate of escape from a parabolic potential well is given in the limit of small noise by the Kramers formula, cf. Section 5.

In the physics literature, a standard ansatz for calculating the two state process given time dependent transition rates is to solve an associated differential equation for the probabilities of occupying state ± 1 at time t, a so-called master equation [cf. 4].

An advantage of the reduced model is its simplicity. It should be especially useful in systems with more than two meta-stable states. Although it is intuitively plausible to apply a two state filter, there is possibly a problem with the measure of resonance, for it might happen that with the same notion of tuning stochastic resonance would be detected in the two state model, while no optimal noise level, i. e. no point of stochastic resonance, exists in the continuous case. This is, indeed, a problem for the SPA coefficient and related measures, see Pavlyukevich [7]. The reason is that in passing to the reduced model small intrawell fluctuations are "filtered out", while they decisively contribute to the SPA coefficient in the original model.

Measures of resonance based on the distribution of intrawell residence times, however, do not have this limitation, that is they are robust under model reduction as Herrmann et al. [10] show.

In Equation (1) replace the term that represents the periodic input signal with a term that corresponds to a force dependent on the state of the solution path a fixed amount of time into the past, that is replace the periodic signal with a point delay. This yields what will be our reference model, see Equation (2).

The idea to study such equations with regard to noise-induced resonance seems to originate with Ohira and Sato [11]. Their analysis, though, is of limited use, because they make too strong assumptions on independence between the components of the reduced model which they consider in discrete time only.

A better analysis of the reduced model for an important special choice of Equation (2) can be found in Tsimring and Pikovsky [12]. The same model is the object of recent studies by Masoller [13], Houlihan et al. [14] and Curtin et al. [15], and it will be our standard example, too.

While the measure of resonance applied by Tsimring and Pikovsky [12] is essentially the first peak in the frequency spectrum, in the other articles focus is laid on the residence time distribution in the reduced model, which is compared with numerical simulations of the original dynamics. Under certain simplifying assumptions, approximative analytical results are obtained via a master equation approach, where the master equation is a DDE instead of an ODE.

In Section 5 we follow Tsimring and Pikovsky [12] in establishing the link between the reduced and the reference model. Results by Masoller [13] show that the density of the residence time distribution has a characteristic jump. She proposes to take the height of this jump as a measure of resonance, and we will follow her proposal, supplementing it by an alternative.

Our approach is different, though, in that we do not use any kind of master equation. Instead, we construct a reduced model with enlarged state space, which has the Markov property and which allows us to explicitly calculate the stationary distributions as well as the residence time distributions.

2 The reference model

Consider the one dimensional motion of a small particle in the presence of large friction and additive white noise subject to the influence of two additional forces: one dependent on the current position of the particle and corresponding to a symmetric double well potential V, the other dependent on the position of the particle a certain amount of time r in the past and corresponding to a symmetric single well potential U, where the position of the extremum of U coincides with the position of the saddle point of V.

Without loss of generality we may assume that the saddle point of the potential V is at the origin and the extrema are located at $(-1,-L)$ and $(1,-L)$ respectively, where $L>0$ is the height of the potential barrier. A standard choice for V is the quartic potential $x \mapsto L(x^4-2x^2)$.

Instead of U we will consider $\beta \cdot U$, where β is a scalar, that serves to "adjust" explicitly the strength of the delay force. An admissible function for U is the parabola $x \mapsto \frac{1}{2}x^2$. In fact, with this choice of U and taking as potential V the quartic potential with $L=\frac{1}{4}$ we find ourselves in the setting that was studied by Tsimring and Pikovsky [12].[1] Another reasonable choice for U would be a function whose first derivative equals the sign function outside a small symmetric interval around 0 and is smoothly continued on this interval (see Fig. 1).

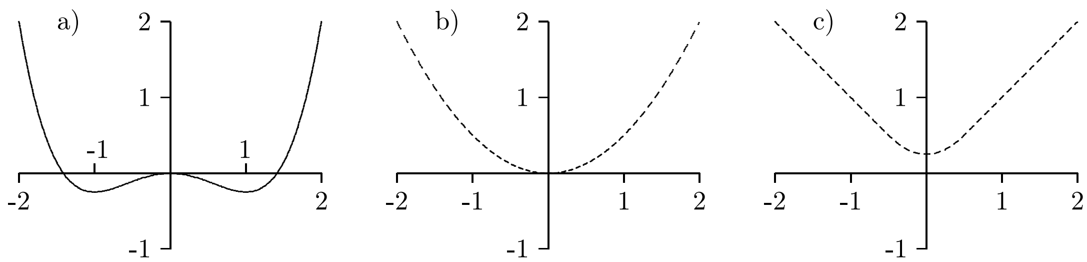

Figure 1: Graphs on the interval $[-2,2]$ of a) quartic double well potential V, b) quadratic delay potential $U : x \mapsto \frac{1}{2}x^2$, c) absolute value delay potential $U : x \mapsto |x|, x \in \mathbb{R} \setminus (-\delta,\delta)$, smoothly continued on $(-\delta,\delta)$.

[1] Our notation is slightly different from that of Equation (1) in Tsimring and Pikovsky [12, p. 1]. In particular, their parameter ϵ, indicating the "strength of the feedback", corresponds to $-\beta$, here.

2.1 The underlying SDDE

The dynamics that govern the motion of a Brownian particle as described above can be expressed by the following stochastic delay differential equation determining our reference model:

$$dX(t) \;=\; -\big(V'\big(X(t)\big) + \beta \cdot U'\big(X(t-r)\big)\big)\, dt \;+\; \sigma \cdot dW(t), \qquad t \geq 0, \tag{2}$$

where $W(.)$ is a standard one dimensional Wiener process on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ adapted to a filtration $(\mathcal{F}_t)_{t\geq 0}$ satisfying the usual conditions, $r > 0$ is the time delay, V', U' are the first derivatives of V and U, respectively, $\beta \in \mathbb{R}$ is a parameter regulating the intensity of the delay force and $\sigma \geq 0$ a noise parameter. In the special case $\sigma = 0$ Equation (2) becomes a DDE, while in case $\beta = 0$ we have an ordinary SDE.

As initial condition at time zero one prescribes an $\mathcal{F}_0$-measurable $\mathbf{C}([-r,0])$-valued random variable ξ such that $\mathbf{E}(\|\xi\|_\infty^2) < \infty$ and $X_0 = \xi$ P-almost surely. Here, X_0 denotes the segment of X at time zero. More generally,

$$X_t := \big[s \mapsto X(t+s),\ s \in [-r,0]\big]$$

is the *segment* at time $t \geq 0$, and $(X_t)_{t\geq 0}$ the *segment process* associated with X, provided a solution $(X(t))_{t\geq -r}$ to Equation (2) exists. The initial segment may, of course, be deterministic, i. e. $X_0 = f$ for some function $f \in \mathbf{C}([-r,0])$.

The above description of the two potentials is compatible with the following conditions on V and U:

$$\begin{aligned}
&\text{(3a)} && V, U \in \mathbf{C}^2(\mathbb{R}), && \\
&\text{(3b)} && V(x) \;=\; V(-x) \quad \text{for all } x \in \mathbb{R}, && U(x) \;=\; U(-x) \quad \text{for all } x \in \mathbb{R},\\
&\text{(3c)} && V'(x) \;=\; 0 \quad \text{iff} \quad x \in \{-1,0,1\}, && U'(x) \;=\; 0 \quad \text{iff} \quad x = 0,\\
&\text{(3d)} && V''(-1) \;=\; V''(1) > 0, && \\
&\text{(3e)} && \sup\{V'(x) \mid x \in (-\infty,-1) \cup (0,1)\} \leq 0, && \sup\{U'(x) \mid x \in (-\infty,0)\} \leq 0.
\end{aligned}$$

If V' and U' are bounded or satisfy a linear growth condition, then results from the literature ensure existence of a unique strong solution for every $\mathcal{F}_0$-measurable $C([-r,0])$-valued square integrable random variable and a (weakly) unique weak solution for every probability measure on $\mathcal{B}(\mathbf{C}([-r,0]))$ as initial condition. The segment processes associated with those solutions enjoy the strong Markov property, see Mohammed [16, 17].

Let $R > 0$ and let V_R, U_R be functions such that V_R and U_R have bounded derivatives, while agreeing with V and U, respectively, on the interval $[-R,R]$. By considering Equation (2) with V, U replaced by V_R, U_R we see that unique solutions exist up to an explosion time.

In order to prevent explosion of solutions to Equation (2) we need growth conditions on V and U. We choose them in a way such as to give us control over the impact of the delay potential U in terms of the potential function V. In addition to conditions (3) let us assume that for some positive constants $\hat{R}_0$, $\hat{\eta}$ and δ we have

$$\begin{aligned}
&\text{(4a)} && U'(x) \;>\; 0 \quad \text{for all } x \in (0,\infty), && \\
&\text{(4b)} && U''(x) \;\geq\; 0 \quad \text{for all } x \in \mathbb{R}, && V''(x) \;\geq\; 0 \quad \text{for all } x \in \mathbb{R} \setminus [-\hat{R}_0, \hat{R}_0],\\
&\text{(4c)} && \frac{V'(x)}{U'(3x)} \;\geq\; \hat{\eta} \cdot |x|^{1+\delta} \quad \text{for all } x \in \mathbb{R} \setminus [-\hat{R}_0, \hat{R}_0]. &&
\end{aligned}$$

Without loss of generality we may assume $\delta \in (0,1]$. Henceforth, whenever the reference model is concerned, we will suppose that conditions (3) and (4) are satisfied. These conditions not only guarantee existence and uniqueness of solutions, but also the existence of a stationary distribution as will be shown next.

Both, non-explosion as well as stationarity, can be checked by the "one step method". Let $b\colon [0,r]\times\mathbb{R}\to\mathbb{R}$ be a continuous function, *locally* Lipschitz in the second variable (uniformly in the first) and such that for some positive constants R_0, η we have

$$x\cdot b(t,x) \;\leq\; -\eta\cdot|x| \quad \text{for all } t\in[0,r],\ x\in\mathbb{R}\setminus[-R_0,R_0]. \tag{5}$$

Growth condition (5) guarantees existence of a unique strong solution to the non-autonomous SDE

$$dY(t) \;=\; b\big(t,Y(t)\big)dt \;+\; \sigma\cdot dW(t), \qquad t\in[0,r], \tag{6}$$

for every $\mathcal{F}_0$-measurable real valued square integrable random variable as initial condition for $Y(0)$, see Durrett [18, pp. 190-192].[2] Moreover, $\mathbf{E}(|Y(0)|^2)<\infty$ implies $\mathbf{E}(\|Y_r\|_\infty^2)<\infty$.

Let $y\in\mathbb{R}$ and let Y^y be a solution to Equation (6) with deterministic initial condition $Y(0)=y$. For $R>0$ denote by τ_R the time of first exit of Y^y from the interval $(-R,R)$, that is we set

$$\tau_R := \inf\big\{t\in[0,r]\;\big|\;|Y^y(t)|\geq R\big\},$$

with $\inf\emptyset=\infty$ by convention. A one dimensional version of Proposition 1.4 in Herrmann et al. [20] yields the following estimate for the probability that Y^y leaves the interval $(-R,R)$ within time r provided R is big enough. For any $\lambda>1$ it holds that

$$\mathrm{P}\big(\tau_R\leq r\big) \;\leq\; 6\left(\frac{r}{\sigma^2}\eta^2+2\right)\exp\left(-\frac{2(\lambda-1)}{\lambda\sigma^2}\,\eta\,R\right) \qquad \text{for all} \quad R\;\geq\;\lambda(R_0\vee|y|). \tag{7}$$

For $f\in\mathbf{C}([-r,0])$ define the drift coefficient b_f by

$$b_f(t,x) := \;-V'(x)-\beta\cdot U'\big(f(t-r)\big), \qquad (t,x)\in[0,r]\times\mathbb{R}.$$

With $(t,x)\in[0,r]\times\mathbb{R}$ we have

$$\begin{aligned}
x\cdot b_f(t,x) \;&=\; -\,x\,V'(x) \;-\; \beta\,x\,U'\big(f(t-r)\big) && \big|\ \text{apply (3b)}\\
&=\; -\,|x|\,U'\big(3|x|\big)\,\frac{V'(x)}{U'(3x)} \;-\; \beta\,x\,U'\big(f(t-r)\big) && \big|\ \text{apply (4b)}\\
&\leq\; -\,|x|\,U'\big(3|x|\big)\,\frac{V'(x)}{U'(3x)} \;+\; |\beta|\,|x|\,U'\big(\|f\|_\infty\big).
\end{aligned}$$

Set $K_V := \sup\{-x\cdot V'(x)\mid x\in[-\hat{R}_0,\hat{R}_0]\}$. Because of (4c) we find that

$$x\cdot b_f(t,x) \quad\leq\quad \begin{cases} K_V+|\beta|\,\hat{R}_0\,U'\big(\|f\|_\infty\big) & \text{if} \quad |x|<\hat{R}_0,\\ -\hat{\eta}\,U'\big(3|x|\big)\,|x|^{2+\delta} \;+\; |\beta|\,|x|\,U'\big(\|f\|_\infty\big) & \text{if} \quad |x|\geq\hat{R}_0.\end{cases}$$

If $\|f\|_\infty\geq 3\big(\hat{R}_0\vee\frac{2|\beta|}{\hat{\eta}}\vee 1\big)$, then

$$x\cdot b_f(t,x) \quad\leq\quad \begin{cases} K_V+|\beta|\,\hat{R}_0\,U'\big(\|f\|_\infty\big) & \text{if} \quad |x|<\hat{R}_0,\\ \frac{1}{3}\,|\beta|\,\|f\|_\infty\,U'\big(\|f\|_\infty\big) & \text{if} \quad \hat{R}_0\leq|x|<\frac{1}{3}\|f\|_\infty,\\ -\frac{\hat{\eta}}{18}\,\|f\|_\infty^{1+\delta}\,U'\big(\|f\|_\infty\big)\,|x| & \text{if} \quad |x|\geq\frac{1}{3}\|f\|_\infty.\end{cases} \tag{8}$$

[2] Alternatively, we could invoke Theorem 10.2.2 in Stroock and Varadhan [19] and the fact that pathwise uniqueness holds.

In particular, provided that $\|f\|_\infty \geq 3\big(\hat{R}_0 \vee \frac{2|\beta|}{\hat{\eta}} \vee 1\big)$, we have

$$(9) \qquad x \cdot b_f(t,x) \;\leq\; -\eta_f\,|x| \qquad \text{for all} \quad t \in [0,r],\ |x| \geq \tfrac{1}{3}\|f\|_\infty,$$

where $\eta_f := \frac{\hat{\eta}}{18}\,\|f\|_\infty^{1+\delta}\,U'(\|f\|_\infty)$.

Clearly, inequality (9) implies growth condition (5) if we replace b with b_f and R_0 with $\frac{1}{3}\|f\|_\infty$. Notice that (5) is also satisfied for b_f when $\|f\|_\infty$ is small. In this case we would estimate the product $x \cdot b_f(t,x)$ by applying (4c) and observing that $U'(x)$ is greater than some positive constant for all positive x big enough.

Let $(X(t)_{t\in[-r,\hat{\tau})}$ be a strong solution to Equation (2) for some admissible initial condition ξ up to explosion time $\hat{\tau}$. We have to show that $\hat{\tau} = \infty$ P-a. s.

Suppose $\hat{\tau} > n \cdot r$ P-almost surely for some $n \in \mathbb{N}_0$. This will certainly be the case if we take $n = 0$. Set $s := n \cdot r$. By hypothesis, $(X(t))_{t\in[-r,s]}$ is well defined as a solution to Equation (2) up to time s. Let $\omega \in \Omega$, and set $f := X_s(\omega)$. Observe that $X_s(\tilde{\omega}) = f$ for almost all $\tilde{\omega} \in \Omega$ with respect to the probability measure $\mathrm{P}(.|\mathcal{F}_s)(\omega)$.

In Equation (6) replace the drift b with b_f and prescribe $f(0)$ as initial condition. Notice that b_f fulfills a growth condition like (5). The fact that there is a unique strong solution to Equation (6) now implies that, $\mathrm{P}(.|\mathcal{F}_s)(\omega)$-almost surely, $(X(s+t))_{t\in[0,r]}$ exists and satisfies (6). Consequently, $\hat{\tau} > s + r$ with probability one under $\mathrm{P}(.|\mathcal{F}_s)(\omega)$.

Since $\mathrm{P}(A) = \mathbf{E}(\mathrm{P}(A|\mathcal{F}_s))$ for any event $A \in \mathcal{F}$, we see that $\hat{\tau} > n \cdot r$ entails $\hat{\tau} > (n+1)r$. Proceeding by induction, we conclude that solutions to our reference Equation (2) cannot explode in finite time.

Stationarity is studied in Scheutzow [21, 22] for equations of a special form. Let F be a real valued Borel measurable and locally bounded functional on $\mathbf{C}([-1,0])$, where *locally bounded* means bounded on bounded subsets. Consider the SDDE

$$(10) \qquad d\tilde{X}(t) \;=\; F(\tilde{X}_t)dt \;+\; dW(t), \qquad t \geq 0.$$

By appropriately scaling time and space one can bring SDDEs with delay length $r \neq 1$ or noise parameter $\sigma \neq 1$, as long as both are positive, into the form of Equation (10). Let us specify F as

$$(11) \qquad F(g) := -\frac{\sqrt{r}}{\sigma} \cdot \Big(V'\big(\sigma\sqrt{r} \cdot g(0)\big) + \beta \cdot U'\big(\sigma\sqrt{r} \cdot g(-1)\big)\Big), \quad g \in \mathbf{C}([-1,0]).$$

Since the coordinate projections are measurable and V', U' are locally bounded continuous functions because of (3a), F is Borel measurable and locally bounded.

Let $g \in \mathbf{C}([-1,0])$ and assume that $\tilde{X}$ together with a Wiener process $(\tilde{W}, (\tilde{\mathcal{F}}_t)_{t\geq 0})$ on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathrm{P}})$ is a weak solution to (10) with F as just defined and $\tilde{X}_0 = g$ $\tilde{\mathrm{P}}$-almost surely. Set $f(t) := \sigma\sqrt{r} \cdot g(\frac{t}{r})$, $t \in [-r, 0]$. Let X together with a Wiener process $(W, (\mathcal{F}_t)_{t\geq 0})$ on $(\Omega, \mathcal{F}, \mathrm{P})$ be a weak solution of (2) such that $\mathrm{P}(X_0 = f) = 1$. Then the processes $(X(t))_{t\geq -r}$ and $(\sigma\sqrt{r} \cdot \tilde{X}(\frac{t}{r}))_{t\geq -r}$ have the same distribution [cf. 21, p. 31]. Therefore, if $\sigma > 0$, then our reference equation can be transformed into an instance of Equation (10).

Theorem 1 cites part of Theorem 3 and Theorem 4 from Scheutzow [22, pp. 47-48, 54]. Strong existence and pathwise uniqueness clearly imply weak existence and uniqueness in distribution. Notice that in order to have a weak solution for any initial distribution it is sufficient to check existence of solutions for all *deterministic* initial conditions.

Theorem 1 (Scheutzow). *Let F be a Borel measurable and locally bounded functional on* $\mathbf{C}([-1,0])$. *Assume that weak existence and uniqueness hold for Equation* (10). *Let $\mathcal{L}\colon \mathbf{C}([-1,0]) \to [0,\infty)$ be Borel measurable, and set for $R \geq 0$*

$$A_R := \,-\, \sup\big\{\tilde{\mathbf{E}}_g\big(\mathcal{L}(\tilde{X}_1)\big) - \mathcal{L}(g) \;\big|\; g \in \mathbf{C}([-1,0]),\ \|g\|_\infty \geq R\big\}.$$

Here, $\tilde{\mathbf{E}}_g$ means expectation with respect to the probability measure of a weak solution for (10) with deterministic initial condition g.

1. *Let $((\tilde{X},\tilde{W}),(\tilde{\Omega},\tilde{\mathcal{F}},\tilde{\mathrm{P}}),(\tilde{\mathcal{F}}_t))$ be a weak solution to (10) with arbitrary initial distribution ν. There is at most one invariant probability measure π for Equation (10). If an invariant probability measure π exists, then*
$$\tilde{\mathrm{P}}_{\tilde{X}_t} \overset{t\to\infty}{\to} \pi \quad \textit{in total variation.}$$
2. *If $A_0 \vee 0 < \infty$ and $0 \leq \lim_{R\to\infty} \frac{A_0\vee 0}{A_R} < 1$, then (10) possesses an invariant probability measure.*

As Lyapunov functional $\mathcal{L}$ we choose

$$\mathcal{L}(g) := \|g\|_\infty + |g(0)|, \qquad g \in \mathbf{C}([-1,0]).$$

Notice that beside conditions (4) only properties (3a) and (3b) are needed in order to derive growth estimates (8) and (9). For $g \in \mathbf{C}([-1,0])$ define the rescaled drift condition $\tilde{b}_g$ by

$$\tilde{b}_g(t,x) := \tfrac{\sqrt{r}}{\sigma}\cdot b_{f_g}\big(r\cdot t, \sigma\sqrt{r}\cdot x\big), \quad (t,x)\in[0,1]\times\mathbb{R}, \quad \text{where} \quad f_g(t) := \sigma\sqrt{r}\cdot g\big(\tfrac{t}{r}\big), \quad t\in[-r,0].$$

In particular, $\|f_g\|_\infty = \sigma\sqrt{r}\cdot\|g\|_\infty$. Set $\tilde{R}_0 := \frac{3}{\sigma\sqrt{r}}\big(\hat{R}_0 \vee \frac{2|\beta|}{\hat{\eta}} \vee 1\big)$. If $\|g\|_\infty \geq \tilde{R}_0$ then inequality (9) entails that

$$x\cdot\tilde{b}_g(t,x) \;\leq\; -\tilde{\eta}_g\,|x| \qquad \text{for all} \quad t\in[0,1],\ |x| \geq \tfrac{1}{3}\|g\|_\infty, \tag{12}$$

where

$$\tilde{\eta}_g := \tfrac{\sqrt{r}}{\sigma}\,\eta_{f_g} \;=\; \tfrac{\hat{\eta}}{18}\cdot\sigma^\delta\cdot r^{1+\frac{\delta}{2}}\cdot\|g\|_\infty^{1+\delta}\cdot U'\big(\sigma\sqrt{r}\cdot\|g\|_\infty\big).$$

In analogy to Equation (6), we consider non-autonomous SDEs of the form

$$d\tilde{Y}(t) \;=\; \tilde{b}\big(t,\tilde{Y}(t)\big)dt \;+\; dW(t), \qquad t\in[0,1]. \tag{13}$$

The drift coefficient $\tilde{b}$ has to be chosen in accordance with the deterministic initial condition of Equation (10). For $g\in\mathbf{C}([-1,0])$ denote by $\tilde{Y}^g$ the (strongly) unique solution to Equation (13) with $\tilde{b}_g$ in place of $\tilde{b}$ and initial condition $g(0)$. Let $\tilde{\tau}_R$ denote the time of first exit of $\tilde{Y}^g$ from the interval $(-R,R)$, where $R>0$, that is we set

$$\tilde{\tau}_R := \inf\big\{t\in[0,1] \;\big|\; |\tilde{Y}^g(t)| \geq R\big\},$$

with $\inf\emptyset = \infty$ by convention. As a consequence of (7), provided that $\|g\|_\infty \geq \tilde{R}_0$, with $\lambda > 1$ we have

$$\mathrm{P}\big(\tilde{\tau}_R \leq 1\big) \;\leq\; 6\big(\tilde{\eta}_g^2+2\big)\exp\left(-\frac{2(\lambda-1)}{\lambda}\,\tilde{\eta}_g\,R\right) \qquad \text{for all} \quad R \;\geq\; \lambda\big(\tfrac{1}{3}\|g\|_\infty \vee |g(0)|\big). \tag{14}$$

First, we turn to estimating the expected supremum of $|\tilde{Y}^g|$. Set $y_g := \frac{1}{3}\|g\|_\infty \vee |g(0)|$. For $\|g\|_\infty \geq \tilde{R}_0$ and any $\lambda > 1$ it holds that

$$\begin{aligned}
\mathbf{E}\big(\|\tilde{Y}_1^g\|_\infty\big) \;&=\; \int_0^\infty \mathrm{P}\left(\sup_{t\in[0,1]}|\tilde{Y}^g(t)| > R\right)dR \;\;\leq\;\; \lambda\,y_g + \int_{\lambda y_g}^\infty \mathrm{P}\left(\tilde{\tau}_R \leq 1\right)dR \qquad \big|\ \text{apply (14)}\\
&\leq\; \lambda\,y_g \;+\; 6\big(\tilde{\eta}_g^2+2\big)\cdot\int_{\lambda y_g}^\infty \exp\left(-\frac{2(\lambda-1)}{\lambda}\,\tilde{\eta}_g\,R\right)dR\\
&=\; \lambda\,y_g \;+\; \frac{3\lambda(\tilde{\eta}_g^2+2)}{(\lambda-1)\tilde{\eta}_g}\cdot\exp\left(-\frac{2(\lambda-1)}{\lambda}\,\tilde{\eta}_g\,y_g\right)\\
&\leq\; \lambda\big(\tfrac{1}{3}\|g\|_\infty \vee |g(0)|\big) \;+\; \frac{3\lambda}{\lambda-1}\left(\tilde{\eta}_g + \frac{2}{\tilde{\eta}_g}\right)\cdot\exp\left(-\frac{2(\lambda-1)}{3}\,\tilde{\eta}_g\,\|g\|_\infty\right).
\end{aligned}$$

Therefore, given $a > 1$ we find $R(a) > 0$ such that for all $g \in \mathbf{C}([-1,0])$ with $\|g\|_\infty \geq R(a)$ it holds that

$$\mathbf{E}\big(\|\tilde{Y}_1^g\|_\infty\big) \quad \leq \quad \begin{cases} \frac{a}{3}\,\|g\|_\infty & \text{if} \quad |g(0)| \leq \frac{1}{3}\|g\|_\infty, \\ a|g(0)| & \text{if} \quad |g(0)| > \frac{1}{3}\|g\|_\infty. \end{cases} \tag{15}$$

Second, we try to find an upper bound for the expected value of $|\tilde{Y}^g|$. Clearly, if $|g(0)| \leq \frac{1}{3}\|g\|_\infty$, then inequality (15) implies $\mathbf{E}(|Y^g(1)|) \leq \frac{a}{3}\|g\|_\infty$. Assume that $|g(0)| > \frac{1}{3}\|g\|_\infty$. Girsanov's theorem implies that the stochastic equation

$$Z(t) \;=\; x \;-\; \int_0^t \theta \cdot \mathrm{sgn}\big(Z(s)\big)\,ds \;+\; \tilde{W}(t), \qquad t \in [0,1], \tag{16}$$

where $\tilde{W}$ is a standard Wiener process, possesses a (weakly) unique weak solution for any choice of the parameters $x \in \mathbb{R}$, $\theta > 0$.

Set $x := |g(0)| - \frac{1}{3}\|g\|_\infty$, $\theta := \tilde{\eta}_g$, and let Z^g be a solution process satisfying (16) with x, θ on an appropriate stochastic basis. By symmetry and because of growth inequality (12), we find that

$$\mathbf{E}\big(|\tilde{Y}^g(t)|\big) \quad \leq \quad \tfrac{1}{3}\|g\|_\infty \;+\; \tilde{\mathbf{E}}\big(|Z^g(t)|\big) \qquad \text{for all } t \in [0,1].$$

Notice that Z^g is a Brownian motion with two-valued drift started at $x = |g(0)| - \frac{1}{3}\|g\|_\infty$, and $|Z^g|$ is a reflected Brownian motion with drift $-\tilde{\eta}_g$. The transition probabilities of Z^g can be computed explicitly, see Karatzas and Shreve [23, pp. 437-441]. It holds that

$$\begin{aligned} \tilde{\mathbf{E}}\big(\|Z^g(1)\|_\infty\big) \quad &= \quad \int_0^\infty \frac{z}{\sqrt{2\pi}} \left(\exp\left(-\frac{1}{2}(x-z-\theta)^2\right) + \exp\left(2\theta x - \frac{1}{2}(x+z+\theta)^2\right)\right) dz \\ &\quad + \int_0^\infty 2z\theta\, \exp\big(-2\theta z\big) \frac{1}{\sqrt{2\pi}} \int_{x+z}^\infty \exp\left(-\frac{1}{2}(v-\theta)^2\right) dv\, dz. \end{aligned}$$

Observe that $\frac{1}{3}\|g\|_\infty \leq x \leq \|g\|_\infty$, that is x is of order $\|g\|_\infty$, while $\theta = \tilde{\eta}_g$ is of order at least $\|g\|_\infty^{1+\delta}$. Therefore, $\tilde{\mathbf{E}}(\|Z^g(1)\|_\infty)$ goes to zero as $\|g\|_\infty$ tends to infinity. We conclude that, given $a > 1$, there is $\tilde{R}(a) > 0$ such that for all $g \in \mathbf{C}([-1,0])$ with $\|g\|_\infty \geq \tilde{R}(a)$ it holds that

$$\mathbf{E}\big(|\tilde{Y}^g(1)|\big) \quad \leq \quad \tfrac{a}{3}\,\|g\|_\infty. \tag{17}$$

Observe that A_0 as defined in Theorem 1 is finite. Estimates (15) and (17) imply that A_R tends to infinity as R goes to infinity. Theorem 1 thus guarantees existence of an invariant probability measure. Let us summarize our findings.

Proposition 1. *Suppose that V, U satisfy conditions* (3) *and* (4)*. Let $\sigma \geq 0$, $\beta \in \mathbb{R}$ be given. Then the following holds for Equation* (2)*:*

1. *For every $\mathcal{F}_0$-measurable $\mathbf{C}([-r,0])$-valued random variable as initial condition there is a (pathwise) unique strong solution.*
2. *The segment solution processes enjoy the strong Markov property.*
3. *If $\sigma > 0$, then there is a unique invariant probability measure π for the segment process, which converges in total variation to π for every initial distribution.*

The additional conditions (4) are rather crude and could be varied in many ways. For example, one might relax condition (4c) by requiring a ratio between V' and U' of order one instead of order $1+\delta$ provided the growth of U' is of polynomial order δ.

A different restriction on the geometry of V and U would be the following: Assume that a constant R_{pot} greater than the positive root of V exists such that V and U are linear on $\mathbb{R}\setminus[-R_{pot}, R_{pot}]$. Clearly, this condition would not allow V to be a parabolic or quartic potential. In this setting, however, the delay parameter β becomes important. By appealing to Theorem 5 in Scheutzow [22, pp. 55-56] we find that a stationary distribution exists if $\sigma > 0$ and $\beta > -\frac{V'(R_{pot})}{U'(R_{pot})}$, while no invariant measure exists in case $\beta < -\frac{V'(R_{pot})}{U'(R_{pot})}$.

2.2 Basic parameter settings

Let us have a look at basic parameter settings for Equation (2). The simplest and least interesting choice of parameters is $\sigma = 0$ and $\beta = 0$, i. e. no noise and no delay. In this case, (2) reduces to a one dimensional ordinary differential equation with two stable solutions, namely -1 and 1, and an instable trivial solution.

The dynamics of the general deterministic delay equation, i. e. $\sigma = 0$, $\beta \neq 0$, is not obvious for all combinations $\beta \in \mathbb{R}$, $r > 0$. In Redmonda et al. [24] stabilization of the trivial solution and the corresponding bifurcation points are studied. The parameter region such that the zero solution is stable is contained in $\beta \geq 1$, $r \in [0, 1]$.[3] This is not the parameter region we are interested in, here. Recall from Section 1 that stochastic resonance is a phenomenon concerned with an increase of order in the presence of weak non-zero noise. For large $|\beta|$ the delay force would be predominant. Similarly, with r small the noise would not have enough time to influence the dynamics.

Indeed, we must be careful in our choice of β lest we end up with a randomly perturbed deterministic oscillator. Solutions to Equation (2) exhibit periodic behavior even for $\beta > 0$ comparatively small.

If $\beta = 0$ and $\sigma > 0$, then our SDDE (2) reduces to an ordinary SDE. Of interest is again the case of small noise. A Brownian particle moving along a solution trajectory spends most of its time fluctuating near the position of the minimum of one or the other potential well, while interwell transitions only occasionally occur.

Now, let $\sigma > 0$ and $|\beta|$ be small enough so that the corresponding deterministic system does not exhibit oscillations. Let us suppose first that β is positive. Then the effect of the delay force should be that of favouring interwell transitions whenever the Brownian particle is currently in the same potential well it was in r units of time in the past, while transitions should become less likely whenever the particle is currently in the well opposite to the one it was in before. Notice that the influence of the delay force alone is insufficient to trigger interwell transitions. In fact, with $\sigma > 0$ not too big, transitions are rare and a typical solution trajectory will still be found near the position of one or the other minimum of V with high probability.

Consider what happens if the noise intensity increases. Of course, interwell transitions become more frequent, while at the same time the intrawell fluctuations increase in strength. But there is an additional effect: As we let the noise grow stronger interwell transitions occur at time intervals of approximately the same length, namely at intervals between r and $2r$, with high probability. The solution trajectories exhibit quasi-periodic switching behavior at a non-zero noise level. This is what we may call an instance of stochastic resonance.

[3]Equation (1.3) in Redmonda et al. [24] is our standard example with V the quartic potential, where $-\alpha$ corresponds to our parameter β.

Further increasing the noise intensity leads to ever growing intrawell fluctuations which eventually destroy the quasi-periodicity of the interwell transitions. When the noise is too strong, the potential barrier of V has no substantial impact anymore and random fluctuations easily crossing the barrier are predominant.

Suppose β is negative. The effect of the delay force, now, is that of pushing the Brownian particle out of the potential well it is currently in whenever the particle's current position is on the side of the potential barrier opposite to the one remembered in the past. Sojourns of duration longer than r, on the other hand, become prolonged due to the influence of the delay which in this case renders transitions less likely.

In order to obtain some kind of regular transition behavior a higher noise level as compared to the case of positive β is necessary. Of course, one could change time scales by increasing the delay time r, thereby allowing for lower noise intensities. In Section 5 we will state more precisely what regular transition behavior means in case $\beta < 0$, yet we will not subsume it under the heading of stochastic resonance.

3 The two state model in discrete time

Applying the ideas sketched in Section 1, we develop a reduced model with the aim of capturing the effective dynamics of the reference model. To simplify things further we start with discrete time. As the segment process associated with the unique solution to (2), the reference model equation, enjoys the strong Markov property, it is reasonable to approximate the transition behavior of that solution by a sequence of Markov chains. One unit of time in the discrete case corresponds to r/M time units in the original model, where the delay interval $[-r, 0]$ is divided into $M \in \mathbb{N}$ equally spaced subintervals.

After defining the approximating Markov chains we obtain an explicit formula for their stationary distributions which will be useful in calculating, for each $M \in \mathbb{N}$, the residence time distribution in the stationary regime and deriving its density function in the limit of discrete time tending to continuous time. Finally, based on the residence time distributions, we introduce two simple measures of resonance.

The results on Markov chains we need are elementary and can be found, for example, in Brémaud [25], which will be our standard reference.

3.1 A sequence of Markov chains and stationary distributions

Let $M \in \mathbb{N}$ be the discretization degree, that is the number of subintervals of $[-r, 0]$. The current state of the process we have to construct can attain only two values, say -1 and 1, corresponding to the positions of the two minima of the double-well potential V. Now, there are $M+1$ lattice points in $[-r, 0]$ that delimit the M equally spaced subintervals, giving rise to 2^{M+1} possible states in the enlarged state space.

Let $\mathrm{S}_M := \{-1, 1\}^{M+1}$ denote the state space of the Markov chain with time unit r/M. Elements of S_M will be written as $(M+1)$-tuples having $\{-1, 1\}$-valued entries indexed (from left to right) from $-M$ to 0. This choice of the index range serves as a mnemonic device to recall how we have discretized the delay interval $[-r, 0]$. Thus, $l \in \{-M, \ldots, 0\}$ corresponds to the point $l \cdot r/M$ in continuous time.

To embed the discrete into the time continuous model, let α, γ be positive real numbers. If $X(.)$ is the unique solution to (2) in the case of "interesting" noise parameter σ and delay parameter β, one may think of α as the escape rate of $X(.)$ from one of the two potential wells under the condition $X(t) \approx X(t-r)$ and of γ as the escape rate of $X(.)$ under the condition $X(t) \approx -X(t-r)$. All of the parameters of the reference model, including the delay length and the geometry of the potentials U and V, will enter the discrete model through the transition rates α and γ, cf. Section 5.

In the discrete model of degree M, instead of two different transition rates we have two different transition probabilities α_M and γ_M with $\alpha_M = R_{sc}(\alpha, M)$, $\gamma_M = R_{sc}(\gamma, M)$, where R_{sc} is an appropriate scaling function. In analogy to the time discretization of a Markov process we set

$$R_{sc} : \{\alpha, \gamma\} \times \mathbb{N} \ni (\eta, N) \mapsto \tfrac{\eta}{\alpha+\gamma} \cdot (1 - e^{-\frac{\alpha+\gamma}{N}}) \in (0, 1). \tag{18}$$

Let $Z = (Z^{(-M)}, \dots, Z^{(0)})$, $\tilde{Z} = (\tilde{Z}^{(-M)}, \dots, \tilde{Z}^{(0)})$ be elements of S_M. A transition from Z to $\tilde{Z}$ shall have positive probability only if the following *shift condition* holds:

$$\forall\, l \in \{-M, \dots, -1\} : \quad \tilde{Z}^{(l)} = Z^{(l+1)}. \tag{19}$$

Example. Take the element $(-1, 1, -1) \in \mathrm{S}_2$. According to the shift condition, starting from $(-1, \mathbf{1}, -\mathbf{1})$ there are at most two transitions with positive probability, namely to the elements $(\mathbf{1}, -\mathbf{1}, 1)$ and $(\mathbf{1}, -\mathbf{1}, -1)$. $\Diamond$

If (19) holds for Z and $\tilde{Z}$ then there are two cases to distinguish which correspond to the conditions $X(t) \approx X(t - r)$ and $X(t) \approx -X(t - r)$, respectively. Denote by $p^M_{Z\tilde{Z}}$ the probability to get from state Z to state $\tilde{Z}$. Under condition (19) we must have

$$\begin{aligned} Z^{(0)} = Z^{(-M)} \quad &\text{then} \quad p^M_{Z\tilde{Z}} = \begin{cases} \alpha_M & \text{if} \quad \tilde{Z}^{(0)} \neq Z^{(0)}, \\ 1 - \alpha_M & \text{if} \quad \tilde{Z}^{(0)} = Z^{(0)}, \end{cases} \\ Z^{(0)} \neq Z^{(-M)} \quad &\text{then} \quad p^M_{Z\tilde{Z}} = \begin{cases} \gamma_M & \text{if} \quad \tilde{Z}^{(0)} \neq Z^{(0)}, \\ 1 - \gamma_M & \text{if} \quad \tilde{Z}^{(0)} = Z^{(0)}. \end{cases} \end{aligned} \tag{20}$$

The fact that – because of (18) – we always have $\alpha_M, \gamma_M \in (0, 1)$, implies

$$p^M_{Z\tilde{Z}} \neq 0 \quad \text{iff} \quad \text{shift condition (19) is satisfied.} \tag{21}$$

Set $\mathbf{P}_M := (p^M_{Z\tilde{Z}})_{Z,\tilde{Z}\in \mathrm{S}_M}$. Clearly, $\mathbf{P}_M$ is a $2^{M+1} \times 2^{M+1}$ transition matrix. For every $M \in \mathbb{N}$ choose an S_M-valued discrete process $(X^M_n)_{n\in\mathbb{N}_0}$ on some measurable space $(\Omega_M, \mathcal{F}_M)$ and probability measures P^M_Z, $Z \in \mathrm{S}_M$, on $\mathcal{F}_M$ such that under P^M_Z the process X^M is a homogeneous Markov chain with transition matrix $\mathbf{P}_M$ and initial condition $\mathrm{P}^M_Z(X^M_0 = Z) = 1$.

If ν is a probability measure on the power set $\wp(\mathrm{S}_M)$ then, as usual, let P^M_ν denote the probability measure on $\mathcal{F}_M$ such that X^M is a Markov chain with transition matrix $\mathbf{P}_M$ and initial distribution ν with respect to P^M_ν. Write P_ν instead of P^M_ν, when there is no ambiguity about the measure P^M.

From relation (21), characterizing the non-zero entries of $\mathbf{P}_M$, it follows that $\mathbf{P}_M$ and the associated Markov chains are irreducible. They are also aperiodic, because the time of residence in state $(-1, \dots, -1)$, for example, has positive probability for any finite number of steps. Since the state space S_M is finite, irreducibility implies positive recurrence, and these two properties together are equivalent to the existence of a uniquely determined stationary distribution on the state space, cf. Brémaud [25, pp. 104-105]. Therefore, for every $M \in \mathbb{N}$, we have a uniquely determined probability measure π_M on $\wp(\mathrm{S}_M)$ such that[4]

$$\pi_M(\tilde{Z}) \quad = \quad \sum_{Z\in\mathrm{S}_M} \pi_M(Z)\, p^M_{Z\tilde{Z}} \qquad \text{for all } \tilde{Z} \in \mathrm{S}_M. \tag{22}$$

[4]For the probability of a singleton $\{Z\}$ under a discrete measure ν we just write $\nu(Z)$.

There is a simple characterization of the stationary distribution π_M in terms of the number of "jumps" of the elements of S_M.[5] Let $Z = (Z^{(-M)}, \ldots, Z^{(0)})$ be an element of S_M, and define the number of jumps of Z as

$$\mathcal{J}(Z) := \#\Big\{ j \in \{-M+1, \ldots, 0\} \mid Z^{(j)} \neq Z^{(j-1)} \Big\}.$$

The global balance equations (22) then lead to

Proposition 2 (Number of jumps formula). *Let $M \in \mathbb{N}$. Set $\tilde{\alpha}_M := \frac{\alpha_M}{(1-\gamma_M)}$, $\tilde{\gamma}_M := \frac{\gamma_M}{(1-\alpha_M)}$, $\tilde{\eta}_M := \tilde{\alpha}_M \cdot \tilde{\gamma}_M$. Then for all $Z \in \mathrm{S}_M$ the following formula holds*

$$\pi_M(Z) \;=\; \frac{1}{c_M}\, \tilde{\alpha}_M^{\lfloor \frac{\mathcal{J}(Z)+1}{2} \rfloor}\, \tilde{\gamma}_M^{\lfloor \frac{\mathcal{J}(Z)}{2} \rfloor} \;=\; \frac{1}{c_M}\, \tilde{\alpha}_M^{\mathcal{J}(Z) \bmod 2}\, \tilde{\eta}_M^{\lfloor \frac{\mathcal{J}(Z)}{2} \rfloor}, \tag{23}$$

where $c_M := \; 2 \cdot \sum\limits_{j=0}^{M} \binom{M}{j}\, \tilde{\alpha}_M^{j \bmod 2}\, \tilde{\eta}_M^{\lfloor \frac{j}{2} \rfloor}$.

Proof. The right-hand part of Equation (23) is just a rearrangement of the middle part. For $Z \in \mathrm{S}_M$ define

$$\psi_M(Z) := \; \tilde{\alpha}_M^{\lfloor \frac{\mathcal{J}(Z)+1}{2} \rfloor}\, \tilde{\gamma}_M^{\lfloor \frac{\mathcal{J}(Z)}{2} \rfloor}.$$

We then have $c_M = \sum_{Z \in \mathrm{S}_M} \psi_M(Z)$, because $\mathcal{J}(Z) \in \{0, \ldots, M\}$ for every $Z \in \mathrm{S}_M$, and with $j \in \{0, \ldots, M\}$ there are exactly $2 \cdot \binom{M}{j}$ elements in S_M having j jumps.
Let $Z = (Z^{(-M)}, \ldots, Z^{(0)})$ be an element of S_M. Define elements $\tilde{Z}$, $\hat{Z}$ of S_M as

$$\tilde{Z} := \big(Z^{(-1)}, Z^{(-M)}, \ldots, Z^{(-1)}\big), \qquad\qquad \hat{Z} := \big(-Z^{(-1)}, Z^{(-M)}, \ldots, Z^{(-1)}\big).$$

Because of (20) and (21) the global balance Equations (22) reduce to

$$\pi_M(Z) \quad = \quad \begin{cases} (1-\alpha_M) \cdot \pi_M(\tilde{Z}) + (1-\gamma_M) \cdot \pi_M(\hat{Z}) & \text{if} \quad Z^{(-1)} = Z^{(0)}, \\ \alpha_M \cdot \pi_M(\tilde{Z}) + \gamma_M \cdot \pi_M(\hat{Z}) & \text{if} \quad Z^{(-1)} = -Z^{(0)}. \end{cases} \tag{24}$$

Equations (24) determine π_M up to a multiplicative constant. Of course, $\sum_{Z \in \mathrm{S}_M} \pi_M(Z) = 1$, and we have already seen that $\frac{1}{c_M} \sum_{Z \in \mathrm{S}_M} \psi_M(Z) = 1$. It is therefore sufficient to show that $\psi_M(Z)$, $Z \in \mathrm{S}_M$, satisfy (24). Let Z, $\tilde{Z}$,$\hat{Z}$ be elements of S_M as above. Then

$$\mathcal{J}(\tilde{Z}) \quad = \quad \begin{cases} \mathcal{J}(Z) & \text{if} \quad Z^{(-M)} = Z^{(0)}, \\ \mathcal{J}(Z)+1 & \text{if} \quad Z^{(-M)} = -Z^{(-1)} = -Z^{(0)}, \\ \mathcal{J}(Z)-1 & \text{if} \quad Z^{(-M)} = Z^{(-1)} = -Z^{(0)}, \end{cases}$$

$$\mathcal{J}(\hat{Z}) \quad = \quad \begin{cases} \mathcal{J}(Z) & \text{if} \quad Z^{(-M)} = -Z^{(0)}, \\ \mathcal{J}(Z)+1 & \text{if} \quad Z^{(-M)} = Z^{(-1)} = Z^{(0)}, \\ \mathcal{J}(Z)-1 & \text{if} \quad Z^{(-M)} = Z^{(0)} = -Z^{(-1)}, \end{cases}$$

and $\psi_M(\tilde{Z})$, $\psi_M(\hat{Z})$ can now be calculated. This yields the assertion. □

Let c_M be the normalizing constant from Proposition 2. By splitting up the sum in the binomial formula we see that

$$c_M \quad = \quad \Big(1 + \sqrt{\tfrac{\tilde{\alpha}_M}{\tilde{\gamma}_M}}\Big)\big(1 + \sqrt{\tilde{\eta}_M}\big)^M + \Big(1 - \sqrt{\tfrac{\tilde{\alpha}_M}{\tilde{\gamma}_M}}\Big)\big(1 - \sqrt{\tilde{\eta}_M}\big)^M. \tag{25}$$

[5] At the moment, "number of changes of sign" would be a label more precise for $\mathcal{J}(Z)$, but cf. Section 4.

3.2 Residence time distributions

Let Y^M be the $\{-1,1\}$-valued sequence of current states of X^M, that is [6]

$$Y_n^M := \begin{cases} (X_n^M)^{(0)} & \text{if} \quad n \in \mathbb{N}, \\ (X_0^M)^{(n)} & \text{if} \quad n \in \{-M, \dots, 0\}. \end{cases}$$

Denote by $L_M(k)$ the probability to remain exactly k units of time in the same state conditional on the occurrence of a jump, that is

$$(26) \qquad L_M(k) \;=\; \mathrm{P}_{\pi_M}(Y_n^M = 1, \dots, Y_{n+k-1}^M = 1, Y_{n+k}^M = -1 \mid Y_{n-1}^M = -1, Y_n^M = 1), \quad k \in \mathbb{N},$$

where $n \in \mathbb{N}$ is arbitrary. The above conditional probability is well defined, because

$$\mathrm{P}_{\pi_M}\big(Y_{-1}^M = -1, Y_0^M = 1\big) \;=\; \pi_M\big(\{(*, \dots, *, -1, 1)\}\big) \;>\; 0.$$

Here, $\{(*, \dots, *, -1, 1)\}$ denotes the set $\{Z \in \mathrm{S}_M \mid Z^{(-1)} = -1, Z^{(0)} = 1\}$. By symmetry the roles of -1 and 1 in (26) are interchangeable. Under P_{π_M} not only X^M is a stationary process, but – as a coordinate projection – Y^M is stationary, too, although it does not, in general, enjoy the Markov property. We note that $L_M(k)$, $k \in \mathbb{N}$, gives the residence time distribution of the sequence of current states of X^M in the stationary regime.

Observe that $L_M(.)$ has a "geometric tail". To make this statement precise set

$$(27) \qquad K_M := \mathrm{P}_{\pi_M}\big(Y_0^M = -1, Y_1^M = 1, \dots, Y_M^M = 1 \;\big|\; Y_0^M = -1, Y_1^M = 1\big).$$

In view of the "extended Markov property" of Y^M, that is the Markov property of the segment chain X^M, we have

$$(28) \qquad L_M(k) \quad = \quad (1 - \gamma_M) \cdot K_M \cdot \alpha_M \cdot (1 - \alpha_M)^{k-M-1}, \qquad k \geq M + 1,$$

where $(1 - \gamma_M) \cdot K_M$ is the probability mass of the geometric tail. Stationarity of P_{π_M} implies

$$K_M \quad = \quad \frac{\pi_M\big((-1, 1, \dots, 1)\big)}{\pi_M\big(\{(*, \dots, *, -1, 1)\}\big)}.$$

From Proposition 2 we see that

$$\pi_M\big((-1, 1, \dots, 1)\big) \;=\; \frac{\tilde{\alpha}_M}{c_M},$$

and arranging the elements of $\{(*, \dots, *, -1, 1)\}$ according to their number of jumps we obtain

$$\pi_M\big(\{(*, \dots, *, -1, 1)\}\big) \quad = \quad \frac{\tilde{\alpha}_M}{2c_M} \cdot \Big(\Big(1 + \sqrt{\tfrac{\tilde{\gamma}_M}{\tilde{\alpha}_M}}\Big)\big(1 + \sqrt{\tilde{\eta}_M}\big)^{M-1} + \Big(1 - \sqrt{\tfrac{\tilde{\gamma}_M}{\tilde{\alpha}_M}}\Big)\big(1 - \sqrt{\tilde{\eta}_M}\big)^{M-1}\Big).$$

We therefore have

$$(29) \qquad K_M \quad = \quad \frac{2}{\Big(1 + \sqrt{\tfrac{\tilde{\gamma}_M}{\tilde{\alpha}_M}}\Big)\big(1 + \sqrt{\tilde{\eta}_M}\big)^{M-1} + \Big(1 - \sqrt{\tfrac{\tilde{\gamma}_M}{\tilde{\alpha}_M}}\Big)\big(1 - \sqrt{\tilde{\eta}_M}\big)^{M-1}}.$$

In a similar fashion we can calculate $L_M(k)$ for $k \in \{1, \dots, M\}$. We obtain

$$(30a) \qquad L_M(M) \quad = \quad \frac{\mathrm{P}_{\pi_M}\big(Y_0^M = -1, Y_1^M = 1, \dots, Y_M^M = 1, Y_{M+1}^M = -1\big)}{\pi_M\big(\{(*, \dots, *, -1, 1)\}\big)} \quad = \quad \gamma_M \cdot K_M,$$

[6] Recall the tuple notation for elements of S_M.

and for $k \in \{1, \ldots, M-1\}$

$$L_M(k) \quad = \quad \frac{\sqrt{\tilde{\gamma}_M}}{2} \cdot K_M \cdot \Big(\sqrt{\tilde{\gamma}_M}\, \big((1+\sqrt{\tilde{\eta}_M})^{M-1-k} + (1-\sqrt{\tilde{\eta}_M})^{M-1-k}\big) + \sqrt{\tilde{\alpha}_M}\, \big((1+\sqrt{\tilde{\eta}_M})^{M-1-k} - (1-\sqrt{\tilde{\eta}_M})^{M-1-k}\big)\Big). \tag{30b}$$

More interesting than the residence time distribution in the case of discrete time is to know the behavior of this distribution in the limit of discretization degree M tending to infinity.

Recall the definition of scaling function R_{sc} according to Equation (18) for some numbers $\alpha, \gamma > 0$. If $\alpha_M = R_{sc}(\alpha, M)$ and $\gamma_M = R_{sc}(\gamma, M)$ for all $M \in \mathbb{N}$, then – with the usual notation $\mathcal{O}(.)$ for the order of convergence – we have

$$\alpha_M = \frac{\alpha}{M} + \mathcal{O}(\frac{1}{M^2}), \qquad\qquad \gamma_M = \frac{\gamma}{M} + \mathcal{O}(\frac{1}{M^2}). \tag{31}$$

Indeed, if condition (31) holds between the transition probabilities α_M, γ_M, $M \in \mathbb{N}$, and some positive transition rates α, γ, then we can calculate the normalizing constant c_M, the "tail constant" K_M and the density function of the residence time distribution in the limit $M \to \infty$.

Proposition 3. *Let $\alpha_M, \gamma_M \in (0,1)$, $M \in \mathbb{N}$. Suppose that the sequences $(\alpha_M)_{M\in\mathbb{N}}$, $(\gamma_M)_{M\in\mathbb{N}}$ satisfy relation (31) for some positive real numbers α, γ. Then c_M and K_M converge to c_∞ and K_∞, respectively, as $M \to \infty$, where*

$$c_\infty := \lim_{M\to\infty} c_M \;=\; \Big(1+\sqrt{\tfrac{\alpha}{\gamma}}\Big)e^{\sqrt{\alpha\gamma}} + \Big(1-\sqrt{\tfrac{\alpha}{\gamma}}\Big)e^{-\sqrt{\alpha\gamma}} \;=\; 2\cdot\sum_{k=0}^{\infty}\frac{1}{k!}\,\alpha^{k \bmod 2}\,(\alpha\gamma)^{\lfloor\frac{k}{2}\rfloor}, \tag{32}$$

$$K_\infty := \lim_{M\to\infty} K_M \;=\; \frac{2}{(1+\sqrt{\frac{\gamma}{\alpha}})e^{\sqrt{\alpha\gamma}} + (1-\sqrt{\frac{\gamma}{\alpha}})e^{-\sqrt{\alpha\gamma}}} \;=\; \frac{\sqrt{\alpha}}{\sqrt{\alpha}\cosh(\sqrt{\alpha\gamma}) + \sqrt{\gamma}\sinh(\sqrt{\alpha\gamma})}. \tag{33}$$

Define a function $f_L : (0,\infty) \mapsto \mathbb{R}$ by $q \mapsto f_L(q) := \lim_{M\to\infty} M \cdot L_M\big(\lfloor qM \rfloor\big)$. Then

$$f_L(q) = \begin{cases} \sqrt{\gamma}\cdot K_\infty \cdot \Big(\sqrt{\gamma}\,\cosh\big(\sqrt{\alpha\gamma}(1-q)\big) + \sqrt{\alpha}\,\sinh\big(\sqrt{\alpha\gamma}(1-q)\big)\Big) & \textit{if} \quad q \in (0,1], \\ K_\infty \cdot \alpha \cdot \exp\big(-\alpha(q-1)\big) & \textit{if} \quad q > 1. \end{cases} \tag{34}$$

Proof. If relation (31) holds, then in order to derive (32) and (33) from (25) and (29), respectively, it is sufficient to observe that $(1 + \frac{a}{N} + \mathcal{O}(\frac{1}{N^2}))^N \overset{N\to\infty}{\to} e^a$ for every $a \in \mathbb{R}$. The last part of (32) is obtained by series expansion. Similarly, expression (34) for f_L follows from Equations (30a), (30b) and (28). ☐

Observe that f_L as defined in Proposition 3 is indeed the density of a probability measure on $(0,\infty)$. In case $\alpha = \gamma$ this probability measure is just an exponential distribution with parameter α $(= \gamma)$. If $\alpha \neq \gamma$ then f_L has a discontinuity at position 1, where the height of the jump is

$$f_L(1+) - f_L(1-) \quad = \quad K_\infty \cdot (\alpha - \gamma). \tag{35}$$

Clearly, the restrictions of f_L to $(0,1]$ and $(1,\infty)$, respectively, are still strictly decreasing functions, and $f_L(q)$, $q \in (1,\infty)$, is again the density of an exponential distribution, this time with parameter α $(\neq \gamma)$ and total probability mass K_∞. The function $f_L(q)$, $q \in (0,1)$, is the density of a mixture of two "hyperbolic" distributions with the geometric mean $\sqrt{\alpha\gamma}$ of α and γ as parameter and total probability mass $1 - K_\infty$. The ratio between the hyperbolic cosine and the hyperbolic sine density is $\sqrt{\gamma}$ to $\sqrt{\alpha}$.

Recall how at the beginning of this section we interpreted the discretization degree M as the number of subintervals of $[-r,0]$, where $r>0$ is the length of the delay that appears in Equation (2). Let us assume that the numbers α, γ are functions of the parameters of our reference model, in particular of the noise parameter σ and the length of the delay r. Then we should interpret the density f_L as being defined on normalized time, that is one unit of time corresponds to r units of time in the reference model. The density of the residence time distribution for the two state model in continuous time should therefore read

$$\tilde{f}_L(t) := \tfrac{1}{r} f_L\big(\tfrac{t}{r}\big), \qquad t \in (0,\infty). \tag{36}$$

Before we may call $\tilde{f}_L$ the density of a residence time distribution, we have to justify the passage to the limit $M \to \infty$ at the level of distributions of the Markov chains X^M, which underlie the definition of L_M. We return to this issue in Section 4.

3.3 Two measures of resonance

Drawing on the residence time distribution of the Markov chain X^M we introduce simple characteristics that provide us with a notion of quality of tuning for the reduced model in discrete time.

We consider X^M and the resonance characteristics to be defined in the stationary regime only, because by doing so we can guarantee that an eventual resonance behavior of the trajectories of X^M is independent of transitory behavior. We know that $\mathbf{P}_M$ is a positive recurrent, irreducible and aperiodic transition matrix and, therefore, the distribution of X_n^M converges to

π_M in total variation as $n \to \infty$ for every initial distribution of X_0 [25, p. 130]. In Section 2 we saw an analogous result for the segment process of a solution to Equation (2).

Assume that the transition probabilities α_M, γ_M are related to some transition rates α, γ by means of a smooth scaling function like (18), for example, such that condition (31) is satisfied. Under this assumption we let the discretization degree M tend to infinity. Assume further that α, γ are functions of the parameters of the reference model, in particular, that $\alpha = \alpha(\sigma)$, $\gamma = \gamma(\sigma)$ are $\mathbf{C}^2$-functions of the noise parameter $\sigma \in (0,\infty)$. The resonance characteristics can then be understood as functions of σ.

Recall that the residence time distribution L_M has a geometric tail in the sense that $L_M(k)$, $k \geq M+1$, renormalized by the factor $(1-\gamma_M)\cdot K_M$ is equivalent to a geometric distribution on $\mathbb{N}\setminus\{1,\ldots,M\}$ with K_M as defined by (27). The distribution which L_M induces on $\{1,\ldots,M\}$ is given – up to a renormalizing factor – by Equations (30b) and (30a). A natural characteristic seems to be the jump in the density of the residence time distribution f_L obtained above. In discrete time, i. e. with discretization degree $M \in \mathbb{N}$, we set

$$\upsilon_M := M\cdot\big(L_M(M+1) - L_M(M)\big). \tag{37}$$

Because of (28), (30a) and (35) we have

$$\upsilon_M \;=\; M \cdot K_M \cdot \big((1-\gamma_M)\cdot\alpha_M - \gamma_M\big), \qquad \upsilon_\infty := \lim_{M\to\infty} \upsilon_M \;=\; K_\infty \cdot (\alpha - \gamma). \tag{38}$$

To consider the height of the discontinuity of f_L as a measure of resonance has already been proposed by Masoller [13]. Following her proposal we define what stochastic resonance means according to the jump characteristic.

Definition 1. Let $M \in \mathbb{N}\cup\{\infty\}$, and suppose that the following conditions hold:

(i) υ_M as a function of the noise parameter σ is twice continuously differentiable,

(ii) $\lim_{\sigma \downarrow 0} \upsilon_M(\sigma) = 0$,

(iii) υ'_M has a smallest root $\sigma_{opt} \in (0, \infty)$.

If υ_M has a global maximum at σ_{opt}, then let us say that the Markov chain X^M or, in case $M = \infty$, the reduced model defined by the family $(X^N)_{N \in \mathbb{N}}$ exhibits *stochastic resonance* and call σ_{opt} the *resonance point*. If υ_M has a global minimum at σ_{opt}, then let us say that the Markov chain X^M (or, in case $M = \infty$, the reduced model) exhibits *pseudo-resonance* and call σ_{opt} the *pseudo-resonance point*.

Alternatively, we may take the probability of transitions in a certain time window as characteristic of the resonance effect. For $M \in \mathbb{N}$ and $q \in (0, 1]$ define

$$\hat{\kappa}_M := \sum_{k=1}^{M} L_M(k), \qquad \kappa_M^{(q)} := \sum_{k=M+1}^{\lfloor (q+1)M \rfloor} L_M(k). \tag{39}$$

By summation over k we see from (28) that

$$\hat{\kappa}_M = 1 - (1 - \gamma_M) \cdot K_M, \qquad \kappa_M^{(q)} = (1 - \gamma_M) \cdot K_M \cdot \left(1 - (1 - \alpha_M)^{\lfloor qM \rfloor}\right),$$

and letting M tend to infinity we get

$$\hat{\kappa}_\infty := \lim_{M \to \infty} \hat{\kappa}_M = 1 - K_\infty, \qquad \kappa_\infty^{(q)} := \lim_{M \to \infty} \kappa_M^{(q)} = K_\infty \cdot (1 - e^{-q \cdot \alpha}). \tag{40}$$

Recall that M steps in time of the chain X^M or the $\{-1, 1\}$-valued process Y^M correspond to an amount of time r in the reference model. Thus, $\hat{\kappa}_M$ corresponds to the probability of remaining at most time r in one and the same state, while $\kappa_M^{(q)}$ approximates the probability of state transitions occurring in a time window corresponding to $(r, (q+1)r]$ of length $q \cdot r$ given a transition at time zero.

In (39) we could have allowed for a "window width" $q > 1$. The interesting case, however, is a small time window, because then $\kappa_M^{(q)}$ measures the probability of transitions within the second delay interval. For $q = 1$ the two components of our resonance measure correspond to time windows of equal length, that is $\hat{\kappa}_M$ gives the probability of transitions within the first delay interval, while $\kappa_M^{(1)}$ is the probability of hopping events occurring in the second delay interval. Since L_M is geometrically distributed on $\mathbb{N} \setminus \{1, \ldots, M\}$, $\kappa_M^{(1)}$ majorizes the transition probability for all time windows of the same length starting after the end of the first delay interval. Let us write κ_M for $\kappa_M^{(1)}$.

The idea of the following definition is to maximize quasi-periodicity by finding a noise level such that sojourns in the same state become neither too long nor too short. Here, short sojourns are those that last less than the length of one delay interval, long sojourns those that last longer than the length of two delay intervals. Observe that if the current state of X^M remains the same for more than M steps in discrete time, then the influence of the delay will be constant until a transition occurs.

Definition 2. Let $M \in \mathbb{N} \cup \{\infty\}$, and suppose that the following conditions hold:

(i) κ_M as a function of the noise parameter σ is twice continuously differentiable with values in the unit interval,

(ii) $\lim_{\sigma \downarrow 0} (\hat{\kappa}_M + \kappa_M)(\sigma) = 0$,

(iii) κ_M has a unique global maximum at $\sigma_{opt} \in (0, \infty)$.

If $\kappa_M(\sigma_{opt}) > \hat{\kappa}_M$, then let us say that the Markov chain X^M or, in case $M = \infty$, the reduced model defined by the family $(X^N)_{N\in\mathbb{N}}$ exhibits *stochastic resonance* of strength $\kappa_M(\sigma_{opt})$, and call σ_{opt} the *resonance point*, else let us speak of *pseudo-resonance* and call σ_{opt} the *pseudo-resonance point*.

In the above definition we might have taken a shorter time window than the second delay interval. A natural choice would have been the probability of transitions occurring in a time window corresponding to $(r,(1+q)r]$ normalized by the window width. In the limit $M \to \infty$ we obtain

$$\lim_{q\downarrow 0} \frac{1}{q} \cdot \kappa_\infty^{(q)} \quad = \quad K_\infty \cdot \alpha \quad = \quad f_L(1+). \tag{41}$$

Here, f_L is the density of the residence time distribution from Proposition 3 and $f_L(1+)$ is the right-hand limit appearing in Equation (35), which gives the height of the discontinuity of f_L.

Of course, Definition 2 could be modified in other ways, most importantly by allowing the time window that corresponds to κ_M to float. This would be necessary for a distributed delay. Suppose that

in the reference model instead of the point delay we had a delay supported on $[-r,-\delta]$ for some $\delta > 0$. Then a reasonable starting point for a measure of resonance could be a time window of length r with its left boundary floating from δ to r. Notice that a distributed delay (in the reference or in the reduced model) can be chosen in such a way as to render continuous the density f_L.

4 The two state model in continuous time

Our aim in this section is to justify the passage from time discretization degree M to the limit $M \to \infty$ as undertaken in Sections 3.2 and 3.3. To this end we will look for a process in continuous time that is the limit in distribution of the Markov chains X^M, $M \in \mathbb{N}$, in the stationary regime. We can then consider the distribution of residence times for this new process and show that it coincides with the limit of the residence time distributions in discrete time which was calculated in Section 3.2. Since the measures of resonance introduced in Section 3.3 were defined over the (discrete) residence time distributions, we may conclude that in this case, too, the passage to the limit $M \to \infty$ is admissible.

For $M \in \mathbb{N}$ the Markov chain X^M takes its values in the finite space S_M with cardinality 2^{M+1}. The first thing to be done, therefore, is to choose a common state space for the Markov chains. This will be $D_0 := D_{\{-1,1\}}([-r,0])$, the space of all $\{-1,1\}$-valued càdlàg functions, i.e. right-continuous functions with left limits, on the interval $[-r,0]$, endowed with the Skorokhod topology. This simplest of all Skorokhod spaces is introduced in detail in Appendix A.3, while in Appendix A.5 we present $D_\infty := D_{\{-1,1\}}([-r,\infty))$, the space of all $\{-1,1\}$-valued càdlàg functions on the infinite interval $[-r,\infty)$.

Recall how in Section 3.1 we partitioned the delay interval $[-r,0]$. Time step $n \in \{-M,-M+1,\ldots\}$ with respect to the chain X^M was said to correspond to point $n \cdot \frac{r}{M}$ in continuous time. Keeping in mind this correspondence, in Section 4.1 we embed the spaces S_M, $M \in \mathbb{N}$, into D_0, which allows us to look upon the stationary distributions π_M as being probability measures on $\mathcal{B}(D_0)$ and to view the random sequences X^M as being D_0-valued Markov chains.

Now, because of shift condition (19) from Section 3.1 one may regard X^M as being a process with trajectories in D_∞. If the discretization of time is taken into account, then the chain X^M induces a probability measure on $\mathcal{B}(D_\infty)$ for every initial distribution over $\mathrm{S}_M \subset D_0$.

Weak convergence of the stationary distributions or, equivalently, convergence of the π_M with respect to the Prohorov metric induced by the Skorokhod topology on D_0 will be established in Section 4.2. Weak convergence of the distributions on $\mathcal{B}(D_\infty)$ is the object of Section 4.3.

Finally, in Section 4.4, we return to the question of identity between the residence time distribution for the limit process and the one we obtained above as the limit of discrete distributions.

4.1 Embedding of the discrete-time chains

First we interpret the finite enlarged state space S_M as a subset of D_0. After that, we change philosophy and regard a chain X^M as being equivalent to a $\{-1,1\}$-valued càdlàg process.

The embedding of S_M, the state space of the Markov chain X^M, into D_0 is in a sense the reverse of what one does when approximating solutions to stochastic delay differential equations by Markov chains in discrete time.[7] Approximation results of this kind were obtained for the multi-dimensional version of Equation (10) by Scheutzow [21, 22]. The method is more powerful, though, as Lorenz [26] shows, where weak convergence of the approximating processes to solutions of multi-dimensional SDDEs is related to a martingale problem that can be associated with the coefficients of the target equation.

Of course, D_0 is a toy space compared to $\mathbf{C}([-r,0],\mathbb{R}^d)$. Notice, however, that linear interpolation as in the case of $\mathbf{C}([-r,0],\mathbb{R}^d)$ is excluded, because the only continuous functions in D_0 are the two constant functions -1 and 1.

Let $M \in \mathbb{N}$, $Z \in \mathrm{S}_M$, and associate with $Z = (Z^{(-M)},\dots,Z^{(0)})$ a function $f_Z : [-r,0] \to \{-1,1\}$ defined by

$$f_Z(t) := Z^{(0)} \cdot \mathbf{1}_{\{0\}}(t) + \sum_{i=-M}^{-1} Z^{(i)} \cdot \mathbf{1}_{[i\frac{r}{M},(i+1)\frac{r}{M})}(t), \quad t \in [-r,0].$$

Clearly, $f_Z \in D_0$. Hence, $\tilde{\iota}_M :\ Z \mapsto f_Z$ defines a natural injection $\mathrm{S}_M \hookrightarrow D_0$, which induces the following embedding of probability measures on $\wp(\mathrm{S}_M)$ into the set of probability measures on $\mathcal{B}(D_0)$.

$$\mathcal{M}^1_+(\mathrm{S}_M) \ni \mu \mapsto \tilde{\mu} := \sum_{Z \in \mathrm{S}_M} \mu(Z) \cdot \delta_{f_Z} \in \mathcal{M}^1_+(D_0),$$

where δ_f is the Dirac or point measure concentrated on $f \in D_0$.

Denote by $\tilde{\pi}_M$ the probability measure on $\mathcal{B}(D_0)$ associated with the stationary distribution π_M for the chain X^M, and write $\tilde{X}^M$ for the corresponding D_0-valued Markov chain. Since all we have done so far is a reinterpretation of the state space the results obtained in Section 3 regarding X^M are also valid for $\tilde{X}^M$.

Although the embedding $\tilde{\iota}_M$ given above is natural in view of how the delay interval $[-r,0]$ should be partitioned according to Section 3.1, it is not the only one possible. Indeed, one could select different interpolation points in the definition of f_Z. As the degree of discretization M increases the complete Skorokhod distance between the different functions f_Z, $Z \in \mathrm{S}_M$ being fixed, tends to zero, and the convergence results stated in Sections 4.2 and 4.3 still hold true.

Following the notation of Appendix A.3, for $Z \in \mathrm{S}_M$ we write

$$J(Z) := J(f_Z), \qquad \dot{J}(Z) := \dot{J}(f_Z), \qquad \zeta_Z := \zeta_{f_Z},$$

thereby denoting the sets of (inner) discontinuities or jumps of Z, and the minimal distance between two discontinuities. Notice that our new definition of $J(Z)$ agrees with the number of jumps $\mathcal{J}(Z)$ defined in Section 3.1 in the sense that $\#J(Z) = \mathcal{J}(Z)$.

Recall the notation of Section 3.1. Let ν be a distribution on $\wp(\mathrm{S}_M)$ and denote by P^M_ν the probability measure on $\mathcal{F}_M$ such that X^M is a Markov chain with transition matrix $\mathbf{P}_M$ and initial distribution $X^M_0 \overset{d}{\sim} \nu$. For a "point distribution" on $Z \in \mathrm{S}_M$ write P^M_Z.

[7] Under suitable conditions the approximating time series converge in distribution to the (weakly unique) solution of the SDDE.

For $f \in D_0$ let $Z(f)$ be the element of S_M such that $Z^{(i)} = f(\frac{r}{M} \cdot i)$ for all $i \in \{-M, \ldots, 0\}$. Let $(Y_n^M)_{n \in \{-M, -M+1, \ldots\}}$ be the sequence of current states of X^M as defined at the beginning of Section 3.2. Write

$$\tilde{Y}^M(t) := Y^M_{\lfloor \frac{t}{r} M \rfloor}, \quad t \geq -r.$$

For $A \in \mathcal{B}(D_\infty)$ set

$$\tilde{\mathrm{P}}_f^M(A) := \mathrm{P}^M_{Z(f)}\big(\tilde{Y}^M \in A\big), \qquad \tilde{\mathrm{P}}_M(A) := \mathrm{P}^M_{\pi_M}\big(\tilde{Y}^M \in A\big),$$

thereby defining probability measures on $\mathcal{B}(D_\infty)$. Note that $\tilde{\mathrm{P}}_f^M$, $\tilde{\mathrm{P}}_M$ are well defined and correspond to the distribution of X^M with $X_0^M = Z(f)$ P_M-almost surely and $X_0^M \overset{d}{\sim} \pi_M$, respectively.

4.2 Convergence of the stationary distributions on D_0

The aim of this section is to prove that the sequence $(\tilde{\pi}_M)_{M \in \mathbb{N}}$ of probability measures on $\mathcal{B}(D_0)$ induced by the sequence (π_M) of stationary distributions converges weakly to a probability measure $\tilde{\pi}$. Since (D_0, d_S°) is separable, Theorem 3 in Appendix A.1 says that weak convergence of $(\tilde{\pi}_M)$ to $\tilde{\pi}$ is equivalent to convergence under the Prohorov metric induced by d_S°.

The proof follows the usual strategy for this kind of convergence. First, we check that the closure of $\{\tilde{\pi}_M \mid M \in \mathbb{N}\}$ is compact in $\mathcal{M}_+^1(D_0)$ with respect to the Prohorov topology. Now, (D_0, d_S°) is also complete. According to the Prohorov compactness criterion, cited as Theorem 4 in Appendix A.1, it is therefore sufficient to show that the set $\{\tilde{\pi}_M \mid M \in \mathbb{N}\}$ is tight.

For the second step, choose a limit point $\tilde{\pi} \in \mathcal{M}_+^1(D_0)$ of $\{\tilde{\pi}_M \mid M \in \mathbb{N}\}$, which exists according to the first step. It remains to show that any convergent subsequence of $(\tilde{\pi}_M)$ actually converges to $\tilde{\pi}$.

Before embarking on the actual proof of convergence we need some technical preparation, which consists in defining suitable approximation sets and estimating their probability under the measures π_M.

Let $N \in \mathbb{N}$, $Z \in \mathrm{S}_N$, and set for $M \in \mathbb{N} \setminus \{1, \ldots, N-1\}$

$$\mathrm{U}_M^N(Z) := \Big\{\tilde{Z} \in \mathrm{S}_M \;\Big|\; \#\dot{J}(\tilde{Z}) = \#\dot{J}(Z) \;\wedge\; \big(\exists\, \lambda \in \Lambda : \sup_{s \in [-r,0]} |\lambda(s) - s| \leq \tfrac{r}{2N+1} \;\wedge\; f_{\tilde{Z}} \circ \lambda = f_Z\big)\Big\}.$$

For N big enough in comparison to r, $\mathrm{U}_M^N(Z) \subset \mathrm{S}_M$ is the set of elements $\tilde{Z} \in \mathrm{S}_M$ such that $d_S(f_Z, f_{\tilde{Z}}) \leq \frac{r}{2N+1}$. Furthermore, $\#J(f_{\tilde{Z}}) = \#J(f_Z)$ for all $\tilde{Z} \in \mathrm{U}_M^N(Z)$.

Notice that $f_{\tilde{Z}}$ is not necessarily an approximation of f_Z with respect to the complete metric d_S°, because the slope of λ can be of order N for all admissible time transformations.

Recall from Proposition 2 that the probability $\pi_M(Z)$ of an element $Z \in \mathrm{S}_M$ under the stationary distribution π_M depends only on the number of jumps of Z. It will be useful to partition S_M into subsets of elements which have equal number of inner jumps. Set

$$\mathrm{S}_M(i) := \Big\{Z \in \mathrm{S}_M \mid \#\dot{J}(f_Z) = i\Big\}, \qquad i \in \{0, \ldots, M-1\}.$$

Clearly, $\mathrm{S}_M = \mathrm{S}_M(0) \cup \ldots \cup \mathrm{S}_M(M-1)$ is a pairwise disjoint union. For $Z, \tilde{Z} \in \mathrm{S}_M(i)$ we have $|\#J(Z) - \#J(\tilde{Z})| \in \{0, 1\}$. Notice that we prescribed $\#\dot{J}(Z) = \#\dot{J}(\tilde{Z})$ instead of $\#J(Z) = \#J(\tilde{Z})$ in the definition of $\mathrm{S}_M(i)$. Elements $Z \in \mathrm{S}_M$ such that f_Z jumps at position 0 play a special role, as their accumulated probability under π_M tends to zero as M tends to infinity.

Before establishing this point, we need two lemmata, see Appendix B. Lemma 1 estimates the number of elements of $\mathrm{S}_M(i)$ and $\mathrm{U}_M^N(Z)$, respectively. Lemma 2 shows that for $M \in \mathbb{N}$ large most of the probability mass of π_M is concentrated on elements of S_M which have a number of jumps small in comparison to M.

It is even sufficient to restrict attention to elements of $\mathrm{U}_M^N(Z)$, where $Z \in \mathrm{S}_N$ is such that the number of jumps of Z is small in comparison to N which in turn must be small against M. We also see that the probability of a set $\mathrm{U}_M^N(Z)$ under π_M gives a good approximation of the probability which the "generating" element $Z \in \mathrm{S}_N$ receives under π_N.

If we compare probabilities with respect to probability measures π_M for different indices $M \in \mathbb{N}$, we have to assume that an appropriate relation holds between the corresponding transition probabilities α_M, γ_M as M varies. We assume scaling relation (31) as in Section 3.2, where we considered convergence of the residence time distributions.

For $M \in \mathbb{N}$ let $\tilde{\pi}_M \in \mathcal{M}_+^1(D_0)$ be the probability measure which corresponds to the stationary distribution π_M, if we embed S_M into D_0 as was done in Section 4.1.

Proposition 4. *Suppose the sequences of transition probabilities $(\alpha_M)_{M\in\mathbb{N}}$, $(\gamma_M)_{M\in\mathbb{N}}$ satisfy relation* (31) *for some transition rates $\alpha, \gamma > 0$. Then there is a probability measure $\tilde{\pi}$ on $\mathcal{B}(D_0)$ such that $(\tilde{\pi}_M)$ converges weakly to $\tilde{\pi}$ as M tends to infinity.*

Proof. Lemma 2 from Appendix B.1 must be applied several times, see Appendix B.2. □

For some special sets we can calculate their probability with respect to $\tilde{\pi}$.

Proposition 5. *Let $\tilde{\pi}$ be the weak limit of $(\tilde{\pi}_M)_{M\in\mathbb{N}}$ according to Proposition 4. For $i \in \mathbb{N}_0$ set*

$$H_i := \{f \in D_0 \mid \#J(f) = \#\dot{J}(f) = i\}, \qquad \hat{H}_i := \{f \in D_0 \mid \#J(f) = \#\dot{J}(f)+1 = i+1\}.$$

Then for all $i \in \mathbb{N}_0$

$$\tilde{\pi}(H_i) \;=\; \frac{2}{c_\infty \cdot i!} \cdot \alpha^{\lfloor \frac{i+1}{2} \rfloor} \gamma^{\lfloor \frac{i}{2} \rfloor}, \qquad\qquad \tilde{\pi}(\hat{H}_i) \;=\; 0.$$

Proof. Observe that H_i, $\hat{H}_i$, $i \in \mathbb{N}_0$, are disjoint closed subsets of D_0, because convergence with respect to the Skorokhod topology on D_0 preserves the number of inner jumps.[8] Indeed, H_i, $\hat{H}_i$, $i \in \mathbb{N}_0$, are the connected components of D_0, and they are also open sets, because $d_S(f, g) = 2$ for all $f, g \in D_0$ such that $f(0) \neq g(0)$ or $\#\dot{J}(f) \neq \#\dot{J}(g)$.

The assertion now follows from Theorem 3 in Appendix A.1, Proposition 2, Equations (33) and (60) of Proposition 3 and Lemma 1 in Appendix B.1, respectively, under scaling condition (31). □

4.3 Convergence of the chain distributions on D_∞

Let the notation be that of Section 4.1, let us write $D := D_{\{-1,1\}}([0,\infty))$, $D_{\mathbb{R}} := D_{\mathbb{R}}([0,\infty))$ and recall $D_0 = D_{\{-1,1\}}([-r,0])$, $D_{\mathbb{R}}^0 = D_{\mathbb{R}}([-r,0])$, $D_\infty = D_{\{-1,1\}}([-r,\infty))$, $D_{\mathbb{R}}^\infty = D_{\mathbb{R}}([-r,\infty))$. All spaces come with their respective Skorokhod topology, and $D \subset D_{\mathbb{R}}$, $D_0 \subset D_{\mathbb{R}}^0$ are closed subsets.

We sketch a proof for weak convergence of the sequence $(\tilde{\mathrm{P}}_M)$ in $\mathcal{M}_+^1(D_\infty)$ applying results from semimartingale theory as developed in Jacod and Shiryaev [27].

A semimartingale with values in $D_{\mathbb{R}}$ is described in terms of its *characteristics*, a triplet (B, C, ν), where B is a truncated predictable process ("drift"), C the quadratic variation process of the continuous martingale part and ν a random measure, namely the compensator of the jump measure of the semimartingale [27, pp. 75-76].

Any probability measure Q on $\mathcal{B}(D)$ gives the distribution of a $\{-1,1\}$-valued jump process. The characteristics (B, C, ν) of such a process take on a special form. One may choose a continuous truncation

[8] Skorokhod convergence in D_∞ does not necessarily preserve the number of jumps.

function with support contained in $(-2,2)$, thereby eliminating the contribution of B. The quadratic variation process C of the continuous martingale part disappears, because the only continuous functions in D are the two constant functions -1 and 1. The important characteristic is therefore the compensator measure ν. If Q corresponds to a $\{-1,1\}$-valued process in discrete time, then the compensator can be calculated explicitly [27, pp. 93-94].

Let $M \in \mathbb{N}$, $Z \in \mathrm{S}_M$ and let $\tilde{\mathrm{P}}_Z^M$ be the corresponding probability measure on $\mathcal{B}(D_\infty)$ as defined in Section 4.1. Recall that $\tilde{\mathrm{P}}_Z^M$ is the distribution of the $\{-1,1\}$-valued càdlàg process $(\tilde{Y}^M(t))_{t\geq -r}$ induced by the sequence Y^M of current states of X^M when time discretization is taken into account. Denote by $(Y(t))_{t\geq -r}$ the canonical process on D_∞ and by $(\mathcal{F}_t)_{t\geq -r}$ the canonical filtration in $\mathcal{B}(D_\infty)$.

Notice that $(Y(t))_{t\geq -r}$ under $\tilde{\mathrm{P}}_Z^M$ is equivalent to the process $(\tilde{Y}^M(t))_{t\geq -r}$ under P_Z^M and that the jumps of $(Y(t))_{t\geq 0}$ under $\tilde{\mathrm{P}}_Z^M$ are concentrated on $\{\frac{r}{M}k \mid k\in\mathbb{N}\}$. We can now calculate the compensator measure $\tilde{\nu}^{M,Z} : D_\infty \times [0,\infty)\times\mathbb{R} \to [0,\infty]$ of $(Y(t))_{t\geq 0}$ under $\tilde{\mathrm{P}}_Z^M$ in terms of the increment process of Y. Observe that $\tilde{\nu}^{M,Z}$ is determined by the integral processes $(\psi * \tilde{\nu}^{M,Z})_{t\geq 0}$, ψ any bounded Borel function.[9] Set $s(k) := \frac{r}{M}\cdot k$, $k\in\{-M,-M+1,\ldots\}$. According to II.3.11 in Jacod and Shiryaev [27, p. 94] it holds for all functions ψ, all $t\geq 0$, $\tilde{\omega}\in D_\infty$ (all probabilities with respect to $\tilde{\mathrm{P}}_Z^M$)

$$
\begin{aligned}
(\psi * \tilde{\nu}^{M,Z})_t(\tilde{\omega}) \quad &= \sum_{k=1}^{\lfloor \frac{t}{r}M\rfloor} \mathbf{E}\Big(\psi\big(Y(s(k)) - Y(s(k-1))\big)\cdot \mathbf{1}_{Y(s(k))\neq Y(s(k-1))} \;\big|\; \mathcal{F}_{s(k-1)}\Big)(\tilde{\omega}) \\
= \sum_{k=1}^{\lfloor \frac{t}{r}M\rfloor} &\mathbf{E}\Big(\ \ldots\ \big|\ \sigma\big(Y(s(k-M-1)),\ldots,Y(s(k-1))\big)\Big)(\tilde{\omega}) \\
= \mathbf{1}_{(\tilde{\omega}(s(-M)),\ldots,\tilde{\omega}(s(0)))=Z}&(\tilde{\omega})\cdot \sum_{k=1}^{\lfloor \frac{t}{r}M\rfloor} \\
&\psi(2)\cdot \mathrm{P}\big(Y(s(k))=1 \;\big|\; Y(s(k-M-1))=-1, Y(s(k-1))=-1\big)\cdot \mathbf{1}_{\tilde{\omega}(s(k-M-1)=-1,\tilde{\omega}(s(k-1))=-1}(\tilde{\omega}) \\
&+\ \psi(2)\cdot \mathrm{P}\big(Y(s(k))=1 \;\big|\; Y(s(k-M-1))=1, Y(s(k-1))=-1\big)\cdot \mathbf{1}_{\tilde{\omega}(s(k-M-1)=1,\tilde{\omega}(s(k-1))=-1}(\tilde{\omega}) \\
&+\ \psi(-2)\cdot \mathrm{P}\big(Y(s(k))=-1 \;\big|\; Y(s(k-M-1))=1, Y(s(k-1))=1\big)\cdot \mathbf{1}_{\tilde{\omega}(s(k-M-1)=1,\tilde{\omega}(s(k-1))=1}(\tilde{\omega}) \\
&+\ \psi(-2)\cdot \mathrm{P}\big(Y(s(k))=-1 \;\big|\; Y(s(k-M-1))=-1, Y(s(k-1))=1\big)\cdot \mathbf{1}_{\tilde{\omega}(s(k-M-1)=-1,\tilde{\omega}(s(k-1))=1}(\tilde{\omega}) \\
= \mathbf{1}_{\ldots}(\tilde{\omega})\cdot \sum_{k=1}^{\lfloor \frac{t}{r}M\rfloor} &\psi(2)\cdot\big(\alpha_M\cdot \mathbf{1}_{\tilde{\omega}(s(k-M-1)=-1,\tilde{\omega}(s(k-1))=-1}(\tilde{\omega}) \ +\ \gamma_M\cdot \mathbf{1}_{\tilde{\omega}(s(k-M-1)=1,\tilde{\omega}(s(k-1))=-1}(\tilde{\omega})\big) \\
&+\ \psi(-2)\cdot\big(\alpha_M\cdot \mathbf{1}_{\tilde{\omega}(s(k-M-1)=1,\tilde{\omega}(s(k-1))=1}(\tilde{\omega}) \ +\ \gamma_M\cdot \mathbf{1}_{\tilde{\omega}(s(k-M-1)=-1,\tilde{\omega}(s(k-1))=1}(\tilde{\omega})\big),
\end{aligned}
$$

where Bayes' formula has been applied.

Let $f\in D_0$ with $f(0) = f(0-)$, and write $Z(f) = Z(f,M)$ for the element of S_M such that $f_{Z(f)} = f$. The compensator measure $\tilde{\nu}^{M,Z(f)}$ then induces a random measure

$$
\begin{aligned}
&\nu^{M,f}\colon D_{\mathbb{R}}\times[0,\infty)\times\mathbb{R}\to[0,\infty], &&\nu^{M,f}(\omega) := \ \tilde{\nu}^{M,Z(f)}(\theta_f(\omega)), \quad \text{where} \\
&\theta_f\colon D_{\mathbb{R}}\to D_{\mathbb{R}}^\infty &&\theta_f(\omega)(t) := \ f(t)\cdot\mathbf{1}_{[-r,0)}(t) \ +\ \omega(t)\cdot\mathbf{1}_{[0,\infty)}(t).
\end{aligned}
$$

Assume that scaling relation (31) is satisfied for some positive transition rates α, γ. Then for all functions

[9] See Jacod and Shiryaev [27, p. 66] for a definition of the integral process w. r. t. a random measure.

$\psi\colon \mathbb{R}\to\mathbb{R}$, $\omega\in D_{\mathbb{R}}$, $t\ge 0$ it holds that

$$\begin{aligned}\lim_{M\to\infty}\ \psi * \nu_t^{M,f}(\omega) \;&=\; \int_0^t \psi(2)\cdot\big(\alpha\cdot \mathbf{1}_{\tilde\omega(s-r)=-1,\tilde\omega(s)=-1}(\theta_f(\omega)) \;+\; \gamma\cdot \mathbf{1}_{\tilde\omega(s-r)=1,\tilde\omega(s)=-1}(\theta_f(\omega))\big)ds\\ &\quad+\int_0^t \psi(-2)\cdot\big(\alpha\cdot \mathbf{1}_{\tilde\omega(s-r)=1,\tilde\omega(s)=1}(\theta_f(\omega)) \;+\; \gamma\cdot \mathbf{1}_{\tilde\omega(s-r)=-1,\tilde\omega(s)=1}(\theta_f(\omega))\big)ds,\end{aligned}$$

which defines a random measure $\nu^f : D_{\mathbb{R}}\times[0,\infty)\times\mathbb{R}\to[0,\infty]$.

Let $\mu\colon D_{\mathbb{R}}\times[0,\infty)\times\mathbb{R}\to[0,\infty]$ be the jump measure associated to the D-valued process $(Y(t))_{t\ge0}$, cf. Jacod and Shiryaev [27, pp. 68-69]. We have for all functions $\psi\colon\mathbb{R}\to\mathbb{R}$, $\omega\in D_{\mathbb{R}}$, $t\ge 0$

$$(\psi*\mu)_t(\omega) \;=\; \sum_{0<s\le t} \mathbf{1}_{\omega(s)\neq\omega(s-)}(\omega)\cdot\big(\psi(2)\cdot\mathbf{1}_{\omega(s)=1,\omega(s-)=-1}(\omega) \;+\; \psi(-2)\cdot\mathbf{1}_{\omega(s)=-1,\omega(s-)=1}(\omega)\big).$$

Theorem IX.2.31 in Jacod and Shiryaev [27, p. 495] guarantees the existence of a probability measure Q^f on $\mathcal{B}(D_{\mathbb{R}})$ such that

- $Q^f\big(\{\omega\in D_{\mathbb{R}}\mid \omega(0)=f(0)\}\big)=1$,
- the canonical process is a semimartingale under Q^f with characteristics $(0,0,\nu^f)$.

We notice that $Q^f(D)=1$. Let us interpret Q^f as a probability measure on $\mathcal{B}(D)$. According to Theorem II.2.21 in Jacod and Shiryaev [27, p. 80] the second property implies that

- $(\psi*\mu-\psi*\nu^f)_{t\ge0}$ is a local martingale under Q^f for every function $\psi\colon\mathbb{R}\to\mathbb{R}$.

Observe that Theorem IX.2.31 does not guarantee uniqueness of the probability measure Q^f. Here, however, uniqueness can be established by considering sequences of stopping times $\tau_1,\tau_2,\dots$ which exhaust the jump positions. The above local martingale property must then be applied to show that any two solution measures to the semimartingale problem coincide on the sets $\{\tau_n\le t\}$ for all $t\ge0$, $n\in\mathbb{N}$. Recall that every element $\omega\in D$ is determined by its value $\omega(0)$ and the positions of its discontinuities. By the uniqueness theorem of measure theory we see that Q^f is uniquely determined.

Let $p\colon D_\infty\to D$ be the natural projection. Then Theorem IX.3.21 in Jacod and Shiryaev [27, p. 505] implies that $\tilde{\mathrm{P}}^M_{Z(f,M)}\circ p^{-1}\xrightarrow{w} Q^f$ in $\mathcal{M}^1_+(D)$. Define a probability measure $\tilde{\mathrm{P}}^f\in\mathcal{M}^1_+(D_\infty)$ by $\tilde{\mathrm{P}}^f := Q^f\circ\theta_f^{-1}$. We have $\delta_{Z(f,M)}\xrightarrow{w}\delta_f$ in $\mathcal{M}^1_+(D_0)$. In view of $Q^f(\{\omega\in D_{\mathbb{R}}\mid\omega(0)=f(0)\})=1$ we conclude that $\tilde{\mathrm{P}}^M_{Z(f,M)}\xrightarrow{w}\tilde{\mathrm{P}}^f$.

The last step is to show that $(\tilde{\mathrm{P}}_M)$ converges weakly, that is in place of a deterministic initial condition $f\in D_0$ we have $\tilde\pi_M\in\mathcal{M}^1_+(D_0)$ as initial distribution. Let $\tilde\pi$ be the weak limit of $(\tilde\pi_M)$ according to Proposition 4. As a consequence of Proposition 5 we have $\tilde\pi(\{f\in D_0\mid f(0)=f(0-)\})=1$. Define $\tilde{\mathrm{P}}\in\mathcal{M}^1_+(D_\infty)$ by

$$\tilde{\mathrm{P}}(A) := \int_{D_0}\tilde{\mathrm{P}}^f(A)\,d\tilde\pi(f),\qquad A\in\mathcal{B}(D_\infty).$$

If $f_1,\dots,f_n\in D_0$ with $f_i(0)=f_i(0-)$, $i\in\{1,\dots,n\}$, then any convex combination of the sequences $(\tilde{\mathrm{P}}^M_{Z(f_1,M)}),\dots,(\tilde{\mathrm{P}}^M_{Z(f_n,M)})$ converges weakly to the corresponding convex combination of the measures $\tilde{\mathrm{P}}^{f_1},\dots,\tilde{\mathrm{P}}^{f_n}$. An approximation argument analogous to that in the proof of Proposition 4 leads to

Proposition 6. *Suppose scaling relation* (31) *holds. Let $\tilde{\mathrm{P}}_M$, $M\in\mathbb{N}$, be defined as in Section 4.1, and let $\tilde\pi$ be the weak limit of $(\pi_M)_{M\in\mathbb{N}}$ according to Proposition 4. Then there is a probability measure $\tilde{\mathrm{P}}\in\mathcal{M}^1_+(D_\infty)$ such that $\tilde{\mathrm{P}}_M\xrightarrow{w}\tilde{\mathrm{P}}$.*

4.4 Residence times revisited

In Section 3.2 we calculated the residence time distribution for the two state model of discretization degree M for each $M \in \mathbb{N}$. We then let M tend to infinity in order to obtain the residence time distribution and its density function in the "continuous time" limit.

At that stage, however, we had not yet established the existence of a corresponding limit process. This was done in Section 4.3, where we saw that $(\tilde{\mathrm{P}}_M)$, the sequence of distributions induced by the two state chains in discrete time, converges weakly to a probability measure $\tilde{\mathrm{P}}$ on $\mathcal{B}(D_\infty)$. We are now in a position to show that any process with distribution $\tilde{\mathrm{P}}$ has the same residence time distribution as the one obtained in Section 3.2.

On the probability space $(D_\infty, \mathcal{B}(D_\infty), \tilde{\mathrm{P}})$ a process with distribution $\tilde{\mathrm{P}}$ is, of course, given by the canonical process of coordinate projections $p_t : D_\infty \to \{-1, 1\}$, because p_t is Borel measurable for all $t \geq -r$, cf. Appendix A.4. We continue to work directly on the canonical space. Define a mapping

$$\tau \colon D_\infty \to [0,\infty], \qquad \tau(f) := \inf\{t \geq 0 \mid f(t) = -1\}. \tag{42}$$

Then τ is Borel measurable as we have

$$\tau^{-1}[0,t] \;=\; p_0^{-1}\{-1\} \cup \bigcap_{n=0}^{\infty} \bigcup_{k=1}^{n} p_0^{-1}\{1\} \cap \ldots \cap p_{\frac{k-1}{n}t}^{-1}\{1\} \cap p_{\frac{k}{n}t}^{-1}\{-1\} \quad \text{for all } t \geq 0,$$

where the càdlàg property of the elements of D_∞ has been exploited. Because of this property the infimum in (42) is really a minimum provided $\tau < \infty$. We notice that τ is a stopping time with respect to the natural filtration in $\mathcal{B}(D_\infty)$ and that τ is finite $\tilde{\mathrm{P}}$-almost surely.

For each $\delta \in (0,1)$ denote by $\tilde{A}_\delta$ the event that in the time interval $[-\delta r, 0]$ there is exactly one jump, that jump going from -1 to 1. This means we set

$$\tilde{A}_\delta := \big\{ f \in D_\infty \;\big|\; \exists\, \tilde{\delta} \in [0,\delta) : \; f(t) = -1 \;\forall\, t \in [-\delta r, -\tilde{\delta} r) \;\wedge\; f(t) = 1 \;\forall\, t \in [-\tilde{\delta} r, 0] \big\}. \tag{43}$$

Observe that $\tilde{A}_\delta \in \mathcal{B}(D_\infty)$ and $\tilde{\mathrm{P}}(\tilde{A}_\delta) > 0$ for all $\delta \in (0,1)$. The distribution function of τ conditional on the event of exactly one jump from -1 to 1 "just before" time zero can be approximated by functions of the form

$$F_\delta(t) := \tilde{\mathrm{P}}\big(\tau \leq t \;\big|\; \tilde{A}_\delta\big), \quad t \in [0,\infty), \tag{44}$$

where $\delta > 0$ must be small. Since τ is $\tilde{\mathrm{P}}$-almost surely finite and $\tilde{A}_\delta$ has positive probability under $\tilde{\mathrm{P}}$, the function F_δ indeed determines a probability distribution on $[0,\infty)$.

Let $\tilde{f}_L$ be the residence time distribution density in the limit of discretization degree M tending to infinity as given by (36). Set

$$F(t) := \int_0^t \tilde{f}_L(s)ds, \quad t \in [0,\infty). \tag{45}$$

We have to show that $F_\delta(t)$ tends to $F(t)$ as δ goes to zero for each $t \in [0,\infty)$. In (42) and (43), the definitions of τ and $\tilde{A}_\delta$, respectively, instead of time zero we could have chosen any starting time $t_0 \geq 0$, see Appendix B.3, which gives a proof of

Proposition 7. *Suppose scaling relation* (31) *holds. Let the distribution functions F_δ, $\delta \in (0,1)$, and F be defined by* (44) *and* (45)*, respectively. Then*

$$\lim_{\delta \downarrow 0} F_\delta(t) = F(t) \quad \textit{for all}\; t \in [0,\infty).$$

5 Connection between the reduced and the reference model

The aim of this section is to provide a heuristic way of

establishing the missing link between our original model, which is given by Equation (2), and the reduced model developed in Section 3. The situation here is quite similar to the one that was studied by Tsimring and Pikovsky [12], and we will closely follow their approach in deriving a relation between the transition rates α, γ and the parameters of the original model.

The main ingredient in finding such a relation is the so-called Kramers rate, which gives an asymptotic approximation of the time a Brownian particle needs in order to escape from a parabolic potential well in the presence of white noise only as the noise intensity tends to zero. By means of the Kramers rate we calculate escape rates from potentials that should mirror the "effective dynamics" of solutions to Equation (2). The resonance characteristics defined in Section 3.3 can then be written down explicitly as functions of the noise parameter σ, which allows us to numerically calculate the resonance point and to compare the optimal noise intensity according to the two state model with the behavior of the original model.

Let $\mathcal{U}$ be a smooth double well potential with the positions of the two local minima at x_{left} and x_{right}, respectively, $x_{left} < x_{right}$, the position of the saddle point at $x_{max} \in (x_{left}, x_{right})$ and such that $\mathcal{U}(x) \to \infty$ as $|x| \to \infty$. An example for $\mathcal{U}$ is the double well potential V from Sections 1 and 2. Consider the SDE

$$dX(t) \;=\; -\mathcal{U}'\big(X(t)\big)dt \;+\; \sigma \cdot dW(t), \qquad t \geq 0, \tag{46}$$

where $W(.)$ is a standard one dimensional Wiener process with respect to a probability measure P and $\sigma > 0$ is a noise parameter. Denote by $X^{x,\sigma}$ a solution of Equation (46) starting in $X^{x,\sigma}(0) = x$, $x \in \mathbb{R}$. With $y \in \mathbb{R}$ let $\tau_y(X^{x,\sigma})$ be the first time $X^{x,\sigma}$ reaches y, that is we set

$$\tau_y(X^{x,\sigma}) := \inf\{t \geq 0 \mid X^{x,\sigma} = y\}.$$

As we are interested in the transition behavior of the diffusion, we need estimates for the distribution of $\tau_y(X^{x,\sigma})$ when x and y belong to different potential wells.

In the limit of small noise the Freidlin-Wentzell theory of large deviations [28] allows to determine the exponential order of $\tau_y(X^{x,\sigma})$ by means of the so-called quasipotential $Q(x, y)$ associated with the double well potential $\mathcal{U}$. One may think of $Q(x, y)$ as measuring the work a Brownian particle has to do in order to get from position x to position y. The following transition law holds.

Theorem 2 (Freidlin-Wentzell). *Let Q be the quasipotential associated with $\mathcal{U}$, let $x \in (-\infty, x_{max})$, $y \in (x_{max}, x_{right}]$. Set $q_l := Q(x_{left}, x_{max})$. Then*

$$\lim_{\sigma \downarrow 0} \sigma^2 \cdot \ln\big(\mathbf{E}_{\mathrm{P}}\big(\tau_y(X^{x,\sigma})\big)\big) \;=\; q_l, \tag{47a}$$

$$\lim_{\sigma \downarrow 0} \mathrm{P}\left(\exp\Big(\frac{q_l - \delta}{\sigma^2}\Big) < \tau_y(X^{x,\sigma}) < \exp\Big(\frac{q_l + \delta}{\sigma^2}\Big)\right) \;=\; 1 \qquad \text{for all} \quad \delta > 0. \tag{47b}$$

Moreover, $Q(x_{left}, x_{max}) = 2\big(\mathcal{U}(x_{max}) - \mathcal{U}(x_{left})\big)$. If $x \in (x_{max}, \infty)$, $y \in [x_{left}, x_{max})$ then q_l has to be replaced with $q_r := Q(x_{right}, x_{max})$.

We notice that in traveling from position x in the left potential well to $y \in (x_{max}, x_{right}]$, a position in the downhill part of the right well, the transition time in the limit of small noise is determined exclusively by the way up from position x_{left} of the left minimum to position x_{max} of the potential barrier.

A typical path of $X^{x,\sigma}$, if $\sigma > 0$ is small, will spend most of its time near the positions of the two minima of the double well potential. Typically, the diffusion will reach the minimum of the potential well where it started, before it can cross the potential barrier at x_{max} and enter the opposite well.

Theorem 2 implies the existence of different time scales for Equation (46). On the one hand, there is the time scale induced by the Wiener process, where one unit of time can be chosen as $\frac{1}{\sigma^2}$, that is the time it takes the quadratic variation process associated with $\sigma W(.)$ to reach 1. On the other hand, there is the mean escape time given by (47a), which is proportional to $\exp\big(\frac{2L}{\sigma^2}\big)$, where $L > 0$ is the height of the potential barrier. Clearly, with $\sigma > 0$ small, the time scale induced by the white noise is negligible in comparison with the escape time scale.

Moreover, if $\mathcal{U}(x_{left}) \neq \mathcal{U}(x_{right})$, then there are two different heights L_l and L_r for the potential barrier depending on where the diffusion starts. Suppose, for instance, that $L_l < L_r$. According to (47b), waiting a time of order $\exp\big(\frac{2L_l+\delta}{\sigma^2}\big)$ with $0 < \delta < 2(L_r - L_l)$ one would witness transitions from the left well to the right well, but no transition in the opposite direction. If the waiting time was of an exponential order less than $\exp\big(\frac{2L_l}{\sigma^2}\big)$, there would be no interwell transitions at all, where "no transitions" means that the probability of a transition occurring tends to zero as $\sigma \to 0$. Thus, by slightly and periodically tilting a symmetric double well potential quasi-periodic transitions can be enforced provided the tilting period is of the right exponential order. This is the mechanism underlying stochastic resonance.

Now, let us suppose that $\tau_y(X^{x,\sigma})$, where $x < x_{max}$ and $y \in (x_{max}, x_{right}]$, is exponentially distributed with rate $\mathrm{r_K} > 0$ such that

$$\mathrm{r_K} \sim \exp\Big(-\frac{2\big(\mathcal{U}(x_{max}) - \mathcal{U}(x_{left})\big)}{\sigma^2}\Big). \tag{48}$$

Equations (47a) and (47b) of Theorem 2 would be fulfilled. In the physics literature it is generally assumed that $\tau_y(X^{x,\sigma})$ obeys an exponential distribution with rate $\mathrm{r_K}$ provided $\sigma > 0$ is sufficiently small. This is known as *Kramers's law*, and $\mathrm{r_K}$ is accordingly called the *Kramers rate* of the respective potential well. It is, moreover, assumed that the proportionality factor missing in (48) can be specified as a function of the curvature of $\mathcal{U}$ at the positions of the minimum and the potential barrier, respectively. The Kramers rate thus reads

$$\mathrm{r_K} = \mathrm{r_K}(\sigma,\mathcal{U}) \;=\; \frac{\sqrt{|\,\mathcal{U}''(x_{left})\mathcal{U}''(x_{max})\,|}}{2\pi}\,\exp\left(-\frac{2|\,\mathcal{U}(x_{left}) - \mathcal{U}(x_{max})\,|}{\sigma^2}\right). \tag{49}$$

Observe that both the assumption of exponentially distributed interwell transition times and formula (49) for the Kramers rate are empirical approximations, where the noise parameter σ is supposed to be sufficiently small.

Well known results for one-dimensional diffusions, extended to the multi-dimensional framework in recent papers by Bovier et al. [29, 30], show that in the limit of small noise the distribution of the interwell transition time indeed approaches an exponential distribution with a noise-dependent rate that asymptotically satisfies relation (48). The order of the approximation error can also be quantified. For our purposes, however, Kramers's law and the Kramers rate as given by Equation (49) will be good enough.

In Section 3.1 we introduced the transition rates α, γ as being switching rates in the two state model conditional on whether or not the current state agrees with the last remembered state. The idea, now, is to find two "effective" potentials $\mathcal{U}_\alpha$, $\mathcal{U}_\gamma$ such that α is proportional to the Kramers rate describing the escape time distribution from potential $\mathcal{U}_\alpha$, while γ is proportional to the Kramers rate for potential $\mathcal{U}_\gamma$, where the Kramers rate is given by formula (49). More precisely, we must have

$$\alpha = \alpha(\sigma) \;=\; r\cdot \mathrm{r_K}(\sigma,\mathcal{U}_\alpha), \qquad \gamma = \gamma(\sigma) \;=\; r\cdot \mathrm{r_K}(\sigma,\mathcal{U}_\gamma). \tag{50}$$

Note that the inclusion of the delay time r as a proportionality factor is necessary, because in the construction of our two state model we took one unit of time as equivalent to the length of the interval $[-r, 0]$.

There is an important point to be made here. In the discussion of Section 3 we assumed that $X(t) \approx 1$ or $X(t) \approx -1$. The error of this approximation is of first order in β, and its contribution to the delay force is proportional to $\beta^2 U''(1) + \mathcal{O}(\beta^3)$, i.e. of the second order in β. As long as we content ourselves with an approximation of first order in β, two states corresponding to the positions of the minima around -1 and 1 should be enough in order to model the effective dynamics of the reference equation. If we wanted to capture the influence of second order terms in the delay force, we would have to build up a model of four states corresponding to the positions $\pm \underline{x}_\alpha$, $\pm \underline{x}_\gamma$ of the minima of the distorted potential V.[10]

The problem disappears, of course, if U' is constant except on a small symmetric interval $(-\delta, \delta)$ around the origin (see Fig. 1 c), for in this case the delay force would not depend on the particular value of $X(t-r)$ provided $|X(t-r)| \geq \delta$.

Let $L := V(0) - V(1)$ be the height of the potential barrier of V. Set $\rho := |V''(0)V''(1)|$, $\eta := \frac{V'''(1)\,U'(1)}{(V''(1))^2}$, $\tilde{\eta} := \frac{U'(1)}{L}$. Neglecting terms of order higher than one, from (49) we obtain

$$\text{(51a)} \qquad \alpha = \alpha(\sigma) \quad \approx \quad r \cdot \frac{\sqrt{\rho\,(1-\eta\,\beta)}}{2\pi} \exp\left(-\frac{2L\,(1-\tilde{\eta}\,\beta)}{\sigma^2}\right),$$

$$\text{(51b)} \qquad \gamma = \gamma(\sigma) \quad \approx \quad r \cdot \frac{\sqrt{\rho\,(1+\eta\,\beta)}}{2\pi} \exp\left(-\frac{2L\,(1+\tilde{\eta}\,\beta)}{\sigma^2}\right).$$

Recall that the Kramers rate is exact only in the small noise limit. Thus, for the formulae (51a) to become the actual rates of escape it is necessary that σ tends to zero. If the rates α, γ as functions of r and σ are to converge to some finite non-zero values, we must have $\sigma \to 0$ and $r \to \infty$ such that $\frac{1}{\sigma^2}$ and $\ln(r)$ are of the same order. There remain errors due to the first order approximations of V, V'' and U, which make sense only if V, U are sufficiently regular and the delay parameter β is of small absolute value.

In Section 3.3 we defined two measures of resonance, namely the jump height υ_M of the residence time distribution density and the probabilities $\hat{\kappa}_M$, κ_M of transitions within the first and second delay interval, respectively.[11] Recall that $M \in \mathbb{N} \cup \{\infty\}$ is the degree of discretization, where $M = \infty$ denotes the limit $M \to \infty$. We restrict attention to the case $M = \infty$, that is to the two state model in continuous time.

Suppose the transition rates α, γ are functions of the reference model parameters as given by (51a) read as equalities. In particular, α, γ are functions of the delay length r and the noise parameter σ. Let us further suppose that the delay parameter β is of small absolute value, $r > 0$ is big enough so that the critical parameter region for σ lies within the scope of formula (51a), and that the remaining parameters are sufficiently nice.

As a consequence of the exponential form the Kramers rate possesses, we notice that

$$\sqrt{\alpha\gamma} \quad = \quad r \cdot \frac{\sqrt{\rho}}{2\pi} \cdot \sqrt[4]{1-\eta^2\beta^2}\, \exp\left(-\tfrac{2L}{\sigma^2}\right) \quad \approx \quad r \cdot \frac{\sqrt{\rho}}{2\pi} \exp\left(-\tfrac{2L}{\sigma^2}\right).$$

In first order of β, the geometric mean $\sqrt{\alpha\gamma}$ of α, γ coincides with the transition rate arising in case $\beta = 0$, that is when there is no delay. Compare this with Proposition 3, which states that the residence time density $\tilde{f}_L$ is distributed on the first delay interval according to a mixed hyperbolic sine - cosine distribution with parameter $\sqrt{\alpha\gamma}$.

[10] Cf. also the numerical results in Curtin et al. [15].

[11] The jump height measure corresponds to a measure of resonance proposed by Masoller [13].

The conditions of Definitions 1 and 2 are satisfied. If $\beta > 0$, then the reduced model exhibits stochastic resonance according to both definitions. According to the jump height measure there is no effect in case $\beta = 0$ and pseudo-resonance in case $\beta < 0$, while the time window measure does not distinguish between $\beta = 0$ and $\beta < 0$, classifying both cases as pseudo-resonance.

Let us specify the potentials V and U according to the model studied by Tsimring and Pikovsky [12], that is V is the standard quartic potential and U a parabola, see Fig. 1. For the constants appearing in formula (51a) we have

$$L = \tfrac{1}{4}, \qquad \rho = 2, \qquad \eta = \frac{3}{2}, \qquad \tilde{\eta} = 4.$$

With $r = 500$, $\beta = 0.1$, for example, we obtain the resonance point $\sigma_\upsilon \approx 0.32$ according to the jump height measure, while the time window measure yields $\sigma_\kappa \approx 0.29$ with probability $\kappa_\infty(\sigma_\kappa) \approx 0.88$ for transitions occurring in the second delay interval.

Assume β is negative. Again, both measures yield an optimal noise level. With $\beta = -0.1$ we have $\sigma_\upsilon \approx 0.30$ as the noise level that maximizes the jump height in $\tilde{f}_L$. According to the time window measure optimal noise level is at $\sigma_\kappa \approx 0.34$, but $\kappa_\infty(\sigma_\kappa) \approx 0.02$, that is sojourns of duration between r and $2r$ are rare.

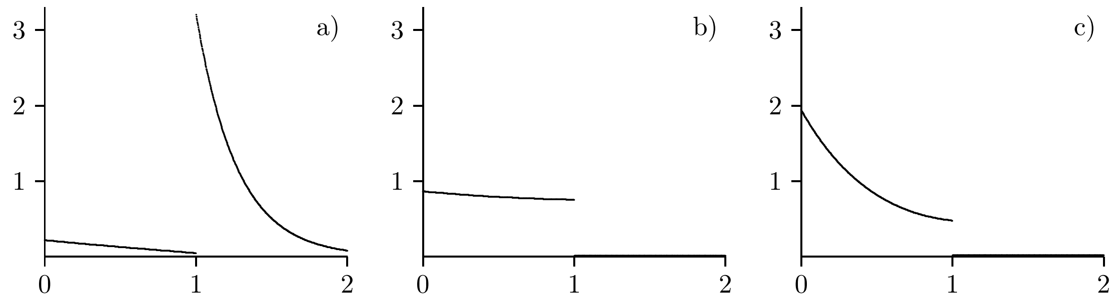

Figure 2: Graphs on $[0, 2]$ of the density f_L of the residence time distribution in normalized time. Parameters of the original model: $r = 500$, a) $\sigma = 0.30$, $\beta = 0.1$, b) $\sigma = 0.30$, $\beta = -0.1$, c) $\sigma = 0.35$, $\beta = -0.1$.

There seems to be a discrepancy, now, between the predicted optimal noise level and the level of "most regular" transition behavior which one would expect from numerical simulation. This is true especially with regard to the jump height measure, the pseudo-resonance point σ_υ being too low.

The problem is that the expected residence time at the level of optimal noise in case $\beta < 0$ is long compared with r. In spite of the fact that long residence times are rare, there is a high probability of finding a solution path remaining in one and the same state for the length of many delay intervals or of witnessing a quasi-periodic transition behavior break down.

For example, let $\sigma = 0.30$, $\beta = 0.1$. The expected residence time is then about $1.16r$, while with $\sigma = 0.30$ and $\beta = -0.1$ the expected residence time is around $4.62r$. More importantly, with β negative the exponential part of the residence time distribution has a "heavy tail" in the sense that long sojourns receive a relatively high probability, cf. Fig. 2.

These properties of the residence time distribution support the distinction made in Definitions 1 and 2 between stochastic resonance and pseudo-resonance.

6 Conclusions and open questions

The main advantage of the two state model which has been our concern for most of this work is that it provides a tool for the analysis of the phenomenon of noise-induced resonance in systems with delay.

The reference model introduced in Section 2 is a more elaborate system exhibiting stochastic resonance. Basic features of this model are the extended Markov property and the existence of an invariant probability measure. Both properties carry over to the two state model.

By first studying the two state model in discrete time we obtained an explicit characterization of its stationary distribution. It was thus possible to calculate the residence time distribution which in turn served as starting point for the definition of two simple measures of resonance. The characterization of the stationary distributions in discrete time together with the passage to the time limit also allows to calculate measures of resonance different from those considered here, for example the entropy of a distribution.

In Section 5 a heuristic link between the reference and the two state model was outlined. The two state model seems to reliably mirror those aspects of the reference model that are responsible for the phenomenon of stochastic resonance. Observe that we did not show whether the dynamics of the original model in the limit of small noise is reducible to the two state model nor whether the resonance measures considered here are indeed robust under model reduction.

There are different ways in which to proceed. The reference model could be modified, for example, by substituting a distributed delay for the point delay. Clearly, the white noise could be replaced with noise of different type, and higher dimensional equations may be considered.

Lastly, the passage to continuous time as addressed in Section 4 should be a special case of more general convergence results for continuous time Markov chains with delay.

A Skorokhod spaces and weak convergence

A.1 Weak convergence in separable metric spaces

The results summarized in this section are taken from Ethier and Kurtz [31, ch. 3 §§ 1-3] and Billingsley [32, § 2]. Let (S, d) be a separable metric space, and denote by $\mathcal{M}^1_+(S)$ the set of probability measures on the Borel σ-algebra $\mathcal{B}(S)$. Define the *Prohorov metric* ρ by

$$\rho(\mathrm{P}, \hat{\mathrm{P}}) := \inf\bigl\{\epsilon > 0 \mid \mathrm{P}(G) \leq \hat{\mathrm{P}}(G^{\epsilon}) + \epsilon \text{ for all closed } G \subseteq S\bigr\}, \qquad \mathrm{P}, \hat{\mathrm{P}} \in \mathcal{M}^1_+(S),$$

where $G^{\epsilon} := \{x \in S \mid \inf_{y \in G} d(x, y) < \epsilon\}$. Then ρ is indeed a metric, and $(\mathcal{M}^1_+(S), \rho)$ is a separable metric space. If, in addition, (S, d) is complete, then $(\mathcal{M}^1_+(S), \rho)$ is complete, too [31, p. 101].

Denote by $\mathbf{C}_b(S)$ the space of all bounded continuous functions on (S, d), topologized with the supremum norm. A sequence $(\mathrm{P}_n)_{n\in\mathbb{N}}$ of probability measures on $\mathcal{B}(S)$ is said to *converge weakly* to a probability measure $\mathrm{P} \in \mathcal{M}^1_+(S)$, in symbols $\mathrm{P}_n \overset{w}{\to} \mathrm{P}$, iff

$$\forall f \in \mathbf{C}_b(S): \qquad \int_S f \, d\mathrm{P}_n \overset{n\to\infty}{\longrightarrow} \int_S f \, d\mathrm{P}\,.$$

The next theorem gives different characterizations of weak convergence and states that weak convergence is equivalent to convergence in the Prohorov metric [31, p. 108]. Recall that we assume (S, d) to be separable. In an arbitrary metric space convergence under ρ would still imply weak convergence and its characterizations, but the converse would not necessarily hold.

Let $\mathrm{P} \in \mathcal{M}_+^1(S)$. A set $A \subseteq S$ is called a P-*continuity set* iff $A \in \mathcal{B}(S)$ and $\mathrm{P}(\partial A) = 0$, i.e. A is Borel measurable and its boundary is a P-null set.

Theorem 3. *Let $(\mathrm{P}_n)_{n\in\mathbb{N}} \subseteq \mathcal{M}_+^1(S)$, $\mathrm{P} \in \mathcal{M}_+^1(S)$. The following conditions are equivalent:*

(i) $\lim_{n\to\infty} \rho(\mathrm{P}_n, \mathrm{P}) = 0$,

(ii) $\mathrm{P}_n \xrightarrow{w} \mathrm{P}$,

(iii) $\int_S f \, d\mathrm{P}_n \overset{n\to\infty}{\longrightarrow} \int_S f \, d\mathrm{P}$ *for all uniformly continuous functions* $f \in \mathbf{C}_b(S)$,

(iv) $\limsup_{n\to\infty} \mathrm{P}_n(A) \leq \mathrm{P}(A)$ *for all closed sets* $A \subseteq S$,

(v) $\liminf_{n\to\infty} \mathrm{P}_n(A) \geq \mathrm{P}(A)$ *for all open sets* $A \subseteq S$,

(vi) $\lim_{n\to\infty} \mathrm{P}_n(A) = \mathrm{P}(A)$ *for all* P-*continuity sets* $A \subseteq S$.

Useful in proving convergence in $\mathcal{M}_+^1(S)$ is the Prohorov criterion for compactness, provided the underlying metric space is complete [31, p. 104].

Theorem 4 (Prohorov). *Let $\Gamma \subseteq \mathcal{M}_+^1(S)$, and suppose that (S, d) is complete. Then the following conditions are equivalent:*

(i) Γ *is tight, i.e.* $\forall \epsilon > 0 \; \exists A \subset S$ *compact*: $\inf_{\mathrm{P}\in\Gamma} \mathrm{P}(A) \geq 1 - \epsilon$,

(ii) $\forall \epsilon > 0 \; \exists A \subset S$ *compact*: $\inf_{\mathrm{P}\in\Gamma} \mathrm{P}(A^\epsilon) \geq 1 - \epsilon$,

(iii) the closure of Γ *is compact in the Prohorov topology.*

The mapping theorem states that under a measurable map weak convergence carries over to the sequence of image measures if the set of discontinuities of the mapping is negligible with respect to the original limit measure [32, p. 21].

Theorem 5. *Let $(\mathrm{P}_n)_{n\in\mathbb{N}} \subset \mathcal{M}_+^1(S)$, $\mathrm{P} \in \mathcal{M}_+^1(S)$. Let (S', d') be a second metric space and $\xi\colon S \to S'$ be a $\mathcal{B}(S')$-$\mathcal{B}(S)$-measurable map. Denote by J_ξ the set of discontinuities of ξ.*

If $\mathrm{P}_n \xrightarrow{w} \mathrm{P}$ and $\mathrm{P}(J_\xi) = 0$ then $\mathrm{P}_n \circ \xi^{-1} \xrightarrow{w} \mathrm{P} \circ \xi^{-1}$.

A.2 The Skorokhod space $D_{\mathbb{R}}^0$

Here, we gather results and definitions from Billingsley [32, §§ 12-13] on the nature of $D_{\mathbb{R}}^0 := D_{\mathbb{R}}([-r, 0])$, the Skorokhod space of all real valued càdlàg functions on the finite interval $[-r, 0]$, i.e. of functions $f : [-r, 0] \to \mathbb{R}$ such that

$$f(t+) := \lim_{s\downarrow t} f(s) \;=\; f(t) \quad \text{for all } t \in [-r, 0), \qquad f(t-) := \lim_{s\uparrow t} f(s) \quad \text{exists for each } t \in (-r, 0].$$

It is possible to define Skorokhod spaces of functions with values in more general spaces than $\mathbb{R}$. In fact, the theory can be developed for $D_E([-r, 0])$ essentially in the same way as for $D_{\mathbb{R}}^0$ as long as E is a Polish space, i.e. a complete and separable metric space.

Let f be any real valued function on $[-r, 0]$. Define the *modulus of càdlàg continuity* as

$$\tilde{w}(f, \delta) := \inf\Big\{ \max_{i\in\{1,\ldots,n\}} w(f, [t_{i-1}, t_i)) \;\Big|\; n \in \mathbb{N},\ -r = t_0 < \ldots < t_n = 0, \min_{i\in\{1,\ldots,n\}} (t_i - t_{i-1}) > \delta \Big\}, \quad \delta \in (0, r), \tag{52}$$

where $w(.,.)$ is the *modulus of uniform continuity* defined as

$$w(f, I) := \sup_{s,t \in I} |f(s) - f(t)|, \quad I \subseteq [-r, 0] \text{ an interval.} \tag{53}$$

A function $f : [-r, 0] \to \mathbb{R}$ is in $D^0_{\mathbb{R}}$ if and only if $\lim_{\delta \downarrow 0} \tilde{w}(f, \delta) = 0$, cf. Billingsley [32, p. 123].

Denote by $\Lambda := \{\lambda : [-r, 0] \to [-r, 0] \mid \lambda \text{ bijective and strictly increasing}\}$ the set of "time transformations" on $[-r, 0]$. For all $\lambda \in \Lambda$ we have $\lambda(-r) = -r$, $\lambda(0) = 0$, and λ is continuous. On Λ define a pseudo-norm

$$\|\lambda\|_\Lambda := \sup_{s,t \in [-r,0],\, s \neq t} \Big|\ln\Big(\frac{\lambda(t) - \lambda(s)}{t - s}\Big)\Big|, \quad \lambda \in \Lambda.$$

Let f, g be elements of $D^0_{\mathbb{R}}$, and define the distances d_S, d°_S as

$$d_S(f, g) := \inf\Big\{\epsilon > 0 \,\Big|\, \exists \lambda \in \Lambda : \sup_{t \in [-r,0]} |\lambda(t) - t| \leq \epsilon \,\wedge\, \sup_{t \in [-r,0]} |f(t) - g(\lambda(t))| \leq \epsilon\Big\}, \tag{54}$$

$$d^\circ_S(f, g) := \inf\Big\{\epsilon > 0 \,\Big|\, \exists \lambda \in \Lambda : \|\lambda\|_\Lambda \leq \epsilon \,\wedge\, \sup_{t \in [-r,0]} |f(t) - g(\lambda(t))| \leq \epsilon\Big\}. \tag{55}$$

Both functionals, d_S and d°_S, measure the distance between f and g in terms of the supremum norm $\|f - g \circ \lambda\|_\infty$. In addition, d_S requires that time transformations λ differ as little as possible from the identity on $[-r, 0]$, while d°_S puts an extra restriction on the slope of the transformations.

The most important difference between d_S and d°_S lies in the fact that they give rise to different sets of Cauchy sequences. Theorem 6 is a summary of Billingsley [32, pp. 125-129].

Theorem 6. *Let d_S, d°_S be defined as above. Then $D^0_{\mathbb{R}}$ is a separable metric space under d_S as well as under d°_S. Both metrics generate the same topology, called the Skorokhod topology.*

Equipped with the Skorokhod topology, $D^0_{\mathbb{R}}$ is a Polish space, and d°_S is a complete metric, while $D^0_{\mathbb{R}}$ is not complete under d_S.

The example below illustrates why d_S does not define a complete metric on $D^0_{\mathbb{R}}$. The sequence to be constructed is a Cauchy sequence with respect to d_S, the only possible limit point of which lies outside the space of càdlàg functions. The same sequence is not Cauchy under d°_S.

Example. Choose $t_0 \in [-r, 0)$. For $n \in \mathbb{N}$ big enough set $f_n := 2 \cdot \mathbf{1}_{I_n} - 1$, where $I_n := [t_0, t_0 + 2^{-n})$. Then $(f_n) \subset D^0_{\mathbb{R}}$ is a Cauchy sequence of $\{-1, 1\}$-valued functions with respect to the metric d_S, and $f_n(t) \overset{n\to\infty}{\to} -1$ for all $t \neq t_0$, but $d_S(f_n, -1) = 2$ for all $n \in \mathbb{N}$, and $f := 2 \cdot \mathbf{1}_{\{t_0\}} - 1$ is no càdlàg function. $\Diamond$

The following criterion, which is Theorem 12.3 in Billingsley [32, p. 130], is an analogue of the Arzelà-Ascoli theorem for compactness in spaces of continuous functions.

Theorem 7. *Let $A \subseteq D^0_{\mathbb{R}}$. Then the closure of A is compact in the Skorokhod topology if and only if the following two conditions hold:*

(i) $\sup_{f \in A} \sup_{t \in [-r,0]} |f(t)| < \infty$,

(ii) $\lim_{\delta \downarrow 0} \sup_{f \in A} \tilde{w}(f, \delta) = 0$.

A.3 The Skorokhod space D_0

There are two equivalent ways of topologizing $D_0 := D_{\{-1,1\}}([-r,0])$, the space of all $\{-1,1\}$-valued càdlàg functions on the finite interval $[-r,0]$. The first is to define metrics d_S, d_S° in analogy to Appendix A.2, where $|.-.|$ should be interpreted as a metric on $\{-1,1\}$. In fact, if (E,d) is a metric space, one can define the Skorokhod space $D_E([-r,0]$ with its accompanying metrics. If, in addition, (E,d) is complete and separable, then an analogue of Theorem 6 holds.[12]

The second option is to restrict the metrics d_S, d_S° and the Skorokhod topology of $D_{\mathbb{R}}^0$ to D_0. This works, because D_0 is a closed subset of $D_{\mathbb{R}}^0$ with respect to the Skorokhod topology. Theorem 6 now implies that D_0 is a separable metric space under d_S, and complete and separable under d_S°, as is the case for $D_{\mathbb{R}}^0$. Define the moduli of continuity w, $\tilde{w}$ by restriction or in analogy to (52) and (53), respectively.

For $f \in D_0$ define $J(f)$, the set of discontinuities or jumps, and ζ_f, the minimal distance between two discontinuities or an inner discontinuity and one of the boundary points of $[-r,0]$, as

$$\begin{aligned} J(f) &:= \left\{t \in (-r,0] \;\middle|\; f(t) \neq f(t-)\right\}, \\ \zeta_f &:= \min\left\{|t-s| \;\middle|\; t,s \in J(f) \cup \{-r,0\}\right\}, \end{aligned}$$

where $f(t-)$ is the left-hand limit of f at t. Set $\dot{J}(f) := J(f) \cap (-r,0)$, the set of inner discontinuities of f. Notice that the only possible discontinuity of f not in $\dot{J}(f)$ is 0, the right boundary of $[-r,0]$.

Proposition 8. *Let $f \in D_0$, $\delta \in (0,r)$, and let $I \subseteq [-r,0]$ be an interval. Then*

$$w(f,I) \in \{0,2\}, \qquad w(f,I) = 0 \Leftrightarrow f \text{ is constant on the interval } I, \tag{56}$$

$$\#J(f) \in \mathbb{N}_0, \tag{57}$$

$$\tilde{w}(f,\delta) \in \{0,2\}, \qquad \tilde{w}(f,\delta) = 0 \Leftrightarrow \zeta_f > \delta. \tag{58}$$

Proof. Obviously, $|f(s)-f(t)| \in \{0,2\}$ for all $s,t \in [-r,0]$, and (56) is a consequence of (53), the definition of w.

If there were an $f \in D_0$ with $\#J(f) = \infty$, one could choose a sequence $(t_n)_{n\in\mathbb{N}} \subset J(f)$ such that $t_n \overset{n\to\infty}{\to} t$ and $t_n \neq t$ for all $n \in \mathbb{N}$. Since f is a càdlàg function, there would be $\delta_l, \delta_r > 0$ such that f is constant on the intervals $(t-\delta_l, t)$, $(t, t+\delta_r)$, except if t were a boundary point of $[-r,0]$, in which case only one of the constants δ_l, δ_r could be chosen appropriately. In any case, $t_n \in (t-\delta_l, t)$ or $t_n \in (t, t+\delta_r)$ for n big enough, a contradiction, because f cannot be constant on an open interval and at the same time have a discontinuity in it.

Clearly, $\tilde{w}(f,\delta) \in \{0,2\}$. Suppose $\tilde{w} = 0$. Then there are $m \in \mathbb{N}$ and a partition $-r = t_0 < \ldots < t_m = 0$ such that $t_i - t_{i-1} > \delta$ and $w(f,[t_{i-1},t_i)) = 0$ for all $i \in \{0,\ldots,m\}$. Hence, f is constant on each interval $[t_{i-1},t_i)$, and the minimal distance between two discontinuities or an inner discontinuity and the boundary of $[-r,0]$ is at least $\min\{(t_i - t_{i-1}) \mid i \in \{0,\ldots,m\}\}$.

Conversely, $\zeta_f > \delta$ implies $\tilde{w}(f,\delta) = 0$, because $-r = t_0 < \ldots < t_m = 0$ forms a suitable partition of $[-r,0]$, if one chooses $m = \#\dot{J}(f)+1$ and takes as $t_1,\ldots,t_{m-1}$ the inner discontinuities of f. ☐

Theorem 7, which states necessary and sufficient conditions for compactness in $D_{\mathbb{R}}^0$, takes on a simple form in the present context.

Proposition 9. *Let $A \subseteq D_0$. Then the closure of A is compact in the Skorokhod topology if and only if $\inf\{\zeta_f \mid f \in A\} > 0$.*

[12] Skorokhod spaces for E-valued functions on the infinite interval $[0,\infty)$ are defined in Ethier and Kurtz [31].

Proof. Condition (i) of Theorem 7 is satisfied for any $A \subseteq D_0$. Hence, we must show that condition (ii) of 7 is equivalent to $\inf_{f\in A} \zeta_f > 0$.

Let $f \in D_0$, then $\tilde{w}(f,\delta) \in \{0,2\}$ for all $\delta \in (0,r)$. Therefore, $\lim_{\delta\downarrow 0} \sup_{f\in A} \tilde{w}(f,\delta) = 0$ if and only if there exists $\delta_0 \in (0,r)$ such that for all $\delta \in (0,\delta_0)$ and all $f \in A$ we have $\tilde{w}(f,\delta) = 0$. According to (58) the latter condition is equivalent to the existence of $\delta_0 \in (0,r)$ such that $\zeta_f \geq \delta_0$ for all $f \in A$, which in turn is just $\inf_{f\in A} \zeta_f > 0$. □

Condition $\inf_{f\in A} \zeta_f > 0$ implies $\sup_{f\in A} \#J(f) < \infty$, but the converse implication does not hold, as can be seen by considering the sequence (f_n) defined in the example of Appendix A.2.

A.4 The space $D^\infty_\mathbb{R}$

Denote by $D^\infty_\mathbb{R} := D_\mathbb{R}([-r,\infty))$ the space of all real valued càdlàg functions on the interval $[-r,\infty)$. Observe that a càdlàg function on $[-r,\infty)$ has at most countably many points of discontinuity [31, p. 116]. It is possible to define Skorokhod metrics on $D^\infty_\mathbb{R}$ in a way similar to that of Appendix A.2, cf. Ethier and Kurtz [31, ch. 3 § 5].

There are two noteworthy differences, though, because the interval the elements of $D^\infty_\mathbb{R}$ live on is no longer bounded to the right. In the definitions which correspond to (54) and (55) one needs a “fading function” or a “fading series” to guarantee finiteness of the metrics. More importantly, the special role which the right boundary plays in (53) and in the definition of the set of time transformations Λ has no counterpart with $D^\infty_\mathbb{R}$.

An alternative is provided by Billingsley [32, § 16], where a connection is established between $D^\infty_\mathbb{R}$ and Skorokhod spaces over finite intervals. Observe that all definitions and properties of $D^0_\mathbb{R} = D_\mathbb{R}([-r,0])$ carry over to the space $D_\mathbb{R}([-r,t])$ for any $t \in (-r,\infty)$. Set

$$\theta_t : \; D^\infty_\mathbb{R} \ni f \mapsto f_{|[-r,t\cdot r]} \in D_\mathbb{R}([-r,t\cdot r]), \quad t > -1.$$

Let $m \in \mathbb{N}_0$. Set $D^m_\mathbb{R} := D_\mathbb{R}([-r, m\cdot r])$. Write d_m, d°_m for the corresponding Skorokhod metrics, and define a function h_m and a “continuous restriction” ψ_m by

$$h_m(t) := \begin{cases} 1 & \text{if} \quad t \in [-r,(m-1)r), \\ m - \frac{t}{r} & \text{if} \quad t \in [(m-1)r, m\cdot r), \\ 0 & \text{if} \quad t \geq m\cdot r, \end{cases}$$

$$\psi_m : \; D^\infty_\mathbb{R} \ni f \mapsto \theta_m(f\cdot h_m) \in D^m_\mathbb{R}.$$

Define a Skorokhod metric d°_∞ on $D^\infty_\mathbb{R}$ by

$$d^\circ_\infty(f,g) := \sum_{m=0}^{\infty} 2^{-m}\Big(1 \wedge d^\circ_m\big(\psi_m(f),\psi_m(g)\big)\Big). \tag{59}$$

An equivalent (but incomplete) Skorokhod metric d_∞ can be defined as in (59) by replacing the metrics d°_m with d_m [32, p. 168]. Now define a metric on the product of the spaces $D^m_\mathbb{R}$, $m \in \mathbb{N}_0$, that is one sets

$$\Pi_D := \prod_{m=0}^{\infty} D^m_\mathbb{R}, \qquad \tilde{d}^\circ_\infty\big((f_m)_{m\in\mathbb{N}_0},(g_m)_{m\in\mathbb{N}_0}\big) := \sum_{m=0}^{\infty} 2^{-m}\big(1 \wedge d^\circ_m(f_m,g_m)\big).$$

We have the following embedding theorem [32, p. 170].

Theorem 8. *Define* $\Psi : D_{\mathbb{R}}^{\infty} \to \Pi_D$ *by* $f \mapsto (\theta_m(f \cdot h_m))_{m \in \mathbb{N}_0}$. *Then*

1. Ψ *is an isometry with respect to* d_{∞}° *and* $\tilde{d}_{\infty}^{\circ}$,
2. $\Psi(D_{\mathbb{R}}^{\infty}) \subset \Pi_D$ *is closed,*
3. $(\Pi_D, \tilde{d}_{\infty}^{\circ})$ *is a Polish space, and so is* $(D_{\mathbb{R}}^{\infty}, d_{\infty}^{\circ})$.

The natural projection $p_t : D_{\mathbb{R}}^{\infty} \ni f \mapsto f(t) \in \mathbb{R}$ is Borel measurable for each $t \geq -r$, and p_t is continuous at $f \in D_{\mathbb{R}}^{\infty}$ if and only if f is continuous at t. The projections generate $\mathcal{B}(D_{\mathbb{R}}^{\infty})$ and form a canonical process [32, p. 172]. Notice that the restrictions θ_t, $t > -1$, are measurable, too.

The following proposition characterizes convergence in $D_{\mathbb{R}}^{\infty}$ in terms of convergence of the restricted sequences [32, p. 169].

Proposition 10. *Let* f_n, $n \in \mathbb{N}$, f *be elements of* $D_{\mathbb{R}}^{\infty}$. *Then* $f_n \overset{n\to\infty}{\to} f$ *w. r. t.* d_{∞}° *if and only if* $\theta_t(f_n) \overset{n\to\infty}{\to} \theta_t(f)$ *in* $D_{\mathbb{R}}([-r, t \cdot r])$ *for every continuity point* $t \cdot r$ *of* f.

A.5 The Skorokhod space D_{∞}

First observe that for any $t > -r$ the space $D_{\{-1,1\}}([-r,t])$ of all $\{-1,1\}$-valued càdlàg functions on $[-r,t]$ can be defined in analogy to the space D_0.

Denote by $D_{\infty} := D_{\{-1,1\}}([-r,\infty))$ the set of all $\{-1,1\}$-valued càdlàg functions on $[-r,\infty)$. Clearly, D_{∞} is a subset of $D_{\mathbb{R}}^{\infty}$, but it is also closed with respect to the Skorokhod topology of $D_{\mathbb{R}}^{\infty}$ as can be seen from Proposition 10 and Appendix A.3. We may therefore restrict the topology of $D_{\mathbb{R}}^{\infty}$ and its metrics d_{∞}, d_{∞}°, thus topologizing D_{∞}.

B Convergence to continuous time

B.1 Approximating sets in S_M

Lemma 1.

$$\forall M \in \mathbb{N}\, \forall i \in \{0,\ldots,M-1\} : \quad \#\mathrm{S}_M(i) \;=\; 4 \cdot \binom{M-1}{i}, \tag{60}$$

$$\forall N \in \mathbb{N}\, \forall Z_1, Z_2 \in \mathrm{S}_N\; \forall M \geq N : \quad Z_1 \neq Z_2 \;\Rightarrow\; \mathrm{U}_M^N(Z_1) \cap \mathrm{U}_M^N(Z_2) = \emptyset, \tag{61}$$

$$\begin{gathered} \forall N \in \mathbb{N}\, \forall i \in \{0,\ldots,\lfloor\sqrt{N}\rfloor - 1\}\; \forall M \geq N\; \forall Z \in \mathrm{S}_N(i) : \\ \left(\lfloor \tfrac{2M}{2N+1} \rfloor\right)^i \;\leq\; \#\mathrm{U}_M^N(Z) \;\leq\; \left(\lfloor \tfrac{2M}{2N+1} \rfloor + 1\right)^i. \end{gathered} \tag{62}$$

Proof. Any element $f \in \tilde{\iota}_M(\mathrm{S}_M(i))$ can be described as follows. Choose i out of $M-1$ possible positions for the inner discontinuities, and decide on the binary values of $f(-r)$, $f(0)$. This determines f, and (60) follows.

Let $Z_1 = (Z_1^{(-N)}, \ldots, Z_1^{(0)})$, $Z_2 = (Z_2^{(-N)}, \ldots, Z_2^{(0)})$ be elements of S_N. By definition of $\mathrm{U}_M^N(.)$, for $\mathrm{U}_M^N(Z_1) \cap \mathrm{U}_M^N(Z_2) \neq \emptyset$ we must have $\#\dot{J}(Z_1) = \#\dot{J}(Z_2)$ as well as $Z_1^{(-N)} = Z_2^{(-N)}$ and $Z_1^{(0)} = Z_2^{(0)}$. Suppose that Z_1, Z_2 have the same number of inner discontinuities and agree at $-N$ and 0, but still $Z_1 \neq Z_2$. Then Z_1, Z_2 differ in the position of at least one inner discontinuity, that is to say there is $s_1 \in (-r,0)$ such that $s_1 \in J(Z_1) \setminus J(Z_2)$; by symmetry, there is also $s_2 \in (-r,0)$ such that $s_2 \in J(Z_2) \setminus J(Z_1)$. Select such an s_2, then $|s - s_2| \geq \frac{r}{N}$ for all $s \in J(Z_1)$.

Let $\tilde{Z}_1$ be an element of $\mathrm{U}_M^N(Z_1)$. Then for any inner discontinuity $s \in \dot{J}(Z_1)$ there is exactly one $\tilde{s} \in \dot{J}(\tilde{Z}_1)$ such that $|\tilde{s} - s| \leq \frac{r}{2N+1}$, and vice versa. The same holds true for any element $\tilde{Z}_2$ of $\mathrm{U}_M^N(Z_2)$ with respect to Z_2. In particular, there is $\tilde{s}_2 \in \dot{J}(\tilde{Z}_2)$ such that $|\tilde{s}_2 - s_2| \leq \frac{r}{2N+1}$. But $\tilde{s}_2 \notin J(\tilde{Z}_1)$, because $|\tilde{s}_2 - \tilde{s}| \geq \frac{r}{N} - \frac{2r}{2N+1} > 0$ for all $\tilde{s} \in J(\tilde{Z}_1)$. Since $\tilde{Z}_1$, $\tilde{Z}_2$ were arbitrary, this establishes (61).

An element $\tilde{Z} \in \mathrm{U}_M^N(Z)$ is determined by the positions of its $\#\dot{J}(Z)$ inner discontinuities, where

$$\left\{k \cdot \tfrac{r}{M} - r \mid k \in \{1, \dots, M-1\}\right\}$$

is the set of possible such positions. If $s \in \dot{J}(Z)$, then there is $k \in \{1, \dots, N-1\}$ with $s = k \cdot \frac{r}{N} - r$, and it exists exactly one $\tilde{s} \in \dot{J}(\tilde{Z})$ such that $\tilde{s} \in [s - \frac{r}{2N+1}, s + \frac{r}{2N+1}]$. Equation (62) is now a consequence of

$$\lfloor \tfrac{2M}{2N+1} \rfloor \;\leq\; \#\Big(\left\{k \cdot \tfrac{r}{M} - r \mid k \in \{1, \dots, M-1\}\right\} \cap [s - \tfrac{r}{2N+1}, s + \tfrac{r}{2N+1}]\Big) \;\leq\; \lfloor \tfrac{2M}{2N+1} \rfloor + 1,$$

for all $s \in \{k \cdot \frac{r}{N} - r \mid k \in \{1, \dots, N-1\}\}$, and the fact that $\#\dot{J}(Z) = i$ for $Z \in \mathrm{S}_N(i)$. $\square$

Lemma 2. *Let M, N, N_0 be natural numbers such that $N_0 < N \leq M$, let $\epsilon > 0$, and define the expressions ψ_1, ψ_2 and ψ_3 as*

$$\begin{aligned}
\psi_1 \;&:=\; \pi_M\Big(\bigcup_{i=0}^{N_0} \bigcup_{Z \in \mathrm{S}_N(i)} \mathrm{U}_M^N(Z)\Big) \;\geq\; 1 - \epsilon,\\
\psi_2 \;&:=\; \forall\, i \in \{0, \dots, N_0\}\; \forall\, Z \in \mathrm{S}_N(i): \quad \big|\pi_M\big(\mathrm{U}_M^N(Z)\big) - \pi_N(Z)\big| \;\leq\; \frac{\epsilon}{N^i},\\
\psi_3 \;&:=\; \sum_{i=N_0+1}^{N-1} \pi_N\big(\mathrm{S}_N(i)\big) \;\leq\; \epsilon.
\end{aligned}$$

Suppose that the sequences of transition probabilities $(\alpha_M)_{M \in \mathbb{N}}$, $(\gamma_M)_{M \in \mathbb{N}}$ satisfy relation (31) *for some transition rates $\alpha, \gamma > 0$. Then for all $\epsilon > 0$*

$$\exists\, \tilde{N}_0 \in \mathbb{N}\; \forall\, N_0 \geq \tilde{N}_0\; \exists\, \tilde{N} \in \mathbb{N}\; \forall\, N \geq \tilde{N}\; \exists\, \tilde{M} \in \mathbb{N}\; \forall\, M \geq \tilde{M}: \quad \psi_1, \tag{63}$$

$$\forall\, N_0 \in \mathbb{N}\; \exists\, \tilde{N} \in \mathbb{N}\; \forall\, N \geq \tilde{N}\; \exists\, \tilde{M} \in \mathbb{N}\; \forall\, M \geq \tilde{M}: \quad \psi_2, \tag{64}$$

$$\exists\, \tilde{N}_0 \in \mathbb{N}\; \forall\, N_0 \geq \tilde{N}_0\; \exists\, \tilde{N} \in \mathbb{N}\; \forall\, N \geq \tilde{N}: \quad \psi_3. \tag{65}$$

Finally, it holds that

$$\forall\, \epsilon > 0\; \exists\, N_0, \tilde{N} \in \mathbb{N}\; \forall\, N \geq \tilde{N}\; \exists\, \tilde{M} \in \mathbb{N}\; \forall\, M \geq \tilde{M}: \quad \psi_1 \;\wedge\; \psi_2 \;\wedge\; \psi_3. \tag{66}$$

Proof. Formula (66) follows by "putting together" (63), (64) and (65), where $N_0 = N_0(\epsilon)$ can be chosen as the maximum of $\tilde{N}_0$ according to (63) and $\tilde{N}_0$ according to (65), $\tilde{N} = \tilde{N}(\epsilon, N_0)$ as the maximum of the respective variables $\tilde{N}$, and in the same way for $\tilde{M} = \tilde{M}(\epsilon, N_0, \tilde{N}, N)$.

The remaining formulae will be established one by one. Let $\epsilon > 0$, without loss of generality $\epsilon < 1$. Recall Proposition 2, where the normalizing constant c_M for the probability measure π_M was defined, and Equation (32) of Proposition 3, where we obtained an explicit expression for $c_\infty = \lim_{M \to \infty} c_M$. In analogy to Propositions 2 and 3, respectively, we set

$$\begin{aligned}
c_{M,M_0} &:= 2 \cdot \sum_{k=0}^{M_0} \binom{M}{k} \Big(\frac{\alpha_M}{1 - \gamma_M}\Big)^{k \bmod 2} \Big(\frac{\alpha_M \cdot \gamma_M}{(1 - \alpha_M)(1 - \gamma_M)}\Big)^{\lfloor \frac{k}{2} \rfloor}, \qquad && M \in \mathbb{N},\; M_0 \in \{1, \dots, M\}.\\
c_{\infty,M_0} &:= 2 \cdot \sum_{k=0}^{M_0} \frac{1}{k!}\, \alpha^{k \bmod 2}\, (\alpha\gamma)^{\lfloor \frac{k}{2} \rfloor}, && M_0 \in \mathbb{N},
\end{aligned}$$

Because of relation (31) it holds that $\forall \tilde\epsilon \in (0,1)\, \forall M_0 \in \mathbb{N}\, \exists \tilde M \in \mathbb{N}\, \forall M \geq \tilde M\, \forall \{0,\dots,M_0\}$:

$$(67)\qquad \begin{aligned} (1+\tilde\epsilon)\cdot(\alpha\gamma)^k \ \geq\ & M^{2k}\cdot\big(\tfrac{\alpha_M\cdot\gamma_M}{(1-\alpha_M)(1-\gamma_M)}\big)^k \ \geq\ (1-\tilde\epsilon)\cdot(\alpha\gamma)^k \\ \wedge\ \ (1+\tilde\epsilon)\alpha \ \geq\ & M\cdot\tfrac{\alpha_M}{1-\gamma_M} \ \geq\ (1-\tilde\epsilon)\alpha \\ \wedge\ \ 1 \ \geq\ & \tfrac{M\cdot\ldots\cdot(M-2k+1)}{M^{2k}} \ \geq\ \tfrac{M\cdot\ldots\cdot(M-2k)}{M^{2k+1}} \ \geq\ 1-\tilde\epsilon. \end{aligned}$$

In view of the above we have

$$(68)\qquad \begin{aligned} &\forall \tilde\epsilon\in(0,1)\ \exists \tilde M_0\in\mathbb{N}\ \forall M_0\geq \tilde M_0\ \exists \tilde M\in\mathbb{N}\ \forall M\geq \tilde M:\\ &1\ \geq\ \frac{c_{\infty,M_0}}{c_\infty}\ \geq\ 1-\tilde\epsilon \quad\wedge\quad M\ \geq\ M_0 \quad\wedge\quad \big|\frac{c_{M,M_0}}{c_\infty}-1\big|+\big|\frac{c_\infty}{c_{M,M_0}}-1\big|\ \leq\ \tilde\epsilon. \end{aligned}$$

To conclude the preparations, recall that for $Z \in \mathrm{S}_M(i)$, where $M\in\mathbb{N}$, $i\in\{0,\dots,M-1\}$, we have $\#J(Z)\in\{i,i+1\}$, and that exactly half of the elements of $\mathrm{S}_M(i)$ has a discontinuity at 0.

Let $\tilde\epsilon>0$. Choose $\tilde N_0\in\mathbb{N}$ such that $\frac{c_{\infty,N_0}}{c_\infty}\geq 1-\tilde\epsilon$ for all $N_0\geq\tilde N_0$. Let $N_0\in\mathbb{N}$ with $N_0\geq\tilde N_0=\tilde N_0(\tilde\epsilon)$. Choose $\tilde N\in\mathbb{N}$ such that $\forall N\geq\tilde N\ \forall i\in\{0,\dots,N_0\}$:

$$\big(\tfrac{2}{2N+1}\big)^i \ \geq\ \tfrac{1-\tilde\epsilon}{N^i} \quad\text{and}\quad \tfrac{1}{N^i}\tbinom{N-1}{i}\ \geq\ \tfrac{1-\tilde\epsilon}{i!}.$$

Let $N\in\mathbb{N}$ with $N\geq\tilde N=\tilde N(\tilde\epsilon,N_0)$. Choose $\tilde M$ such that $\forall M\geq\tilde M\ \forall i\in\{0,\dots,N_0\}$:

$$\begin{aligned} &\big(\lfloor\tfrac{2M}{2N+1}\rfloor\big)^i \ \geq\ \big(\tfrac{2M-(2N+1)}{2N+1}\big)^i \ \geq\ (1-\tilde\epsilon)\big(\tfrac{2M-(2N+1)}{2N}\big)^i \ \geq\ (1-\tilde\epsilon)^2\big(\tfrac{M}{N}\big)^i\\ \wedge\ \ &\tfrac{c_\infty}{c_M}\ \geq\ 1-\tilde\epsilon \quad\wedge\quad \tbinom{M}{i}\big(\tfrac{\alpha_M}{1-\gamma_M}\big)^{i \bmod 2}\big(\tfrac{\alpha_M\cdot\gamma_M}{(1-\alpha_M)(1-\gamma_M)}\big)^{\lfloor\frac{i}{2}\rfloor}\ \geq\ \tfrac{1-\tilde\epsilon}{i!}\,\alpha^{i \bmod 2}\,(\alpha\gamma)^{\lfloor\frac{i}{2}\rfloor}, \end{aligned}$$

where (67) has been applied. For $N_0\geq\tilde N_0(\tilde\epsilon)$, $N\geq\tilde N(\tilde\epsilon,N_0)$, $M\geq\tilde M(\tilde\epsilon,N_0,N)$ we have

$$\begin{aligned} \pi_M\Big(\bigcup_{i=0}^{N_0}\bigcup_{Z\in\mathrm{S}_N(i)}\mathrm{U}_M^N(Z)\Big) \ &=\ \sum_{i=0}^{N_0}\Big(\sum_{Z\in\mathrm{S}_N(i)\,\wedge\,J(Z)=i}+\sum_{Z\in\mathrm{S}_N(i)\,\wedge\,J(Z)=i+1}\Big)\pi_M\big(\mathrm{U}_M^N(Z)\big)\\ \geq\ \ &\tfrac{1}{2c_M}\sum_{i=0}^{N_0}\big(\#\mathrm{S}_N(i)\big)\,\big(\lfloor\tfrac{2M}{2N+1}\rfloor\big)^i\\ &\Big(\big(\tfrac{\alpha_M}{1-\gamma_M}\big)^{i \bmod 2}\big(\tfrac{\alpha_M\cdot\gamma_M}{(1-\alpha_M)(1-\gamma_M)}\big)^{\lfloor\frac{i}{2}\rfloor}+\big(\tfrac{\alpha_M}{1-\gamma_M}\big)^{(i+1) \bmod 2}\big(\tfrac{\alpha_M\cdot\gamma_M}{(1-\alpha_M)(1-\gamma_M)}\big)^{\lfloor\frac{i+1}{2}\rfloor}\Big) \end{aligned}$$

as a consequence of Proposition 2. According to the choice of N_0, N, M and because of (60) it holds that

$$\#\mathrm{S}_N(i)\ \geq\ 4\cdot(1-\tilde\epsilon)\cdot\tfrac{N^i}{i!},\qquad \big(\lfloor\tfrac{2M}{2N+1}\rfloor\big)^i\ \geq\ (1-\tilde\epsilon)^2\big(\tfrac{M}{N}\big)^i,\qquad \tfrac{c_\infty}{c_M}\ \geq\ 1-\tilde\epsilon.$$

We therefore have

$$\begin{aligned} &\pi_M\Big(\bigcup_{i=0}^{N_0}\bigcup_{Z\in\mathrm{S}_N(i)}\mathrm{U}_M^N(Z)\Big)\\ \geq\ \ &\tfrac{2}{c_\infty}(1-\tilde\epsilon)^4\sum_{i=0}^{N_0}\tbinom{M-1}{i}\Big(\big(\tfrac{\alpha_M}{1-\gamma_M}\big)^{i \bmod 2}\big(\tfrac{\alpha_M\cdot\gamma_M}{(1-\alpha_M)(1-\gamma_M)}\big)^{\lfloor\frac{i}{2}\rfloor}+\big(\tfrac{\alpha_M}{1-\gamma_M}\big)^{(i+1) \bmod 2}\big(\tfrac{\alpha_M\cdot\gamma_M}{(1-\alpha_M)(1-\gamma_M)}\big)^{\lfloor\frac{i+1}{2}\rfloor}\Big)\\ \geq\ \ &\tfrac{2}{c_\infty}(1-\tilde\epsilon)^4\sum_{i=0}^{N_0}\tbinom{M}{i}\big(\tfrac{\alpha_M}{1-\gamma_M}\big)^{i \bmod 2}\big(\tfrac{\alpha_M\cdot\gamma_M}{(1-\alpha_M)(1-\gamma_M)}\big)^{\lfloor\frac{i}{2}\rfloor}\\ \geq\ \ &\tfrac{2}{c_\infty}(1-\tilde\epsilon)^5\sum_{i=0}^{N_0}\tfrac{1}{i!}\,\alpha^{i \bmod 2}\,(\alpha\gamma)^{\lfloor\frac{i}{2}\rfloor}\ \ =\ \ (1-\tilde\epsilon)^5\,\tfrac{c_{\infty,N_0}}{c_\infty}\ \ \geq\ \ (1-\tilde\epsilon)^6\ \ \geq\ \ 1-6\tilde\epsilon. \end{aligned}$$

Since $\tilde\epsilon \in (0,1)$ was arbitrary, we may set $\tilde\epsilon := \frac{\epsilon}{6}$, thereby establishing (63).

Let $\tilde\epsilon > 0$, $N_0 \in \mathbb{N}$. Choose $\tilde N \in \mathbb{N}$ such that $\forall N \geq \tilde N\ \forall i \in \{0,\dots,N_0{+}1\}$:

$$\lfloor\sqrt{N}\rfloor \ \geq\ 2N_0+3 \quad\wedge\quad \tfrac{1}{N} \ \leq\ \tilde\epsilon \quad\wedge\quad \big(\tfrac{2N}{2N+1}\big)^i \ \geq\ 1-\tilde\epsilon \quad\wedge\quad 1+\tilde\epsilon \ \geq\ \tfrac{c_\infty}{c_N} \ \geq\ 1-\tilde\epsilon$$
$$\wedge\quad 1+\tilde\epsilon \ \geq\ \tfrac{N}{\alpha}\cdot\tfrac{\alpha_N}{1-\gamma_N} \ \geq\ 1-\tilde\epsilon \quad\wedge\quad 1+\tilde\epsilon \ \geq\ \big(\tfrac{N^2}{\alpha\gamma}\cdot\tfrac{\alpha_N\cdot\gamma_N}{(1-\alpha_N)(1-\gamma_N)}\big)^{\lfloor\frac{i}{2}\rfloor} \ \geq\ 1-\tilde\epsilon,$$

which is possible because of (67). Let $N \in \mathbb{N}$ with $N \geq \tilde N = \tilde N(\tilde\epsilon, N_0)$. Choose $\tilde M$ such that $\forall M \geq \tilde M\ \forall i \in \{0,\dots,N_0\}$:

$$M \ \geq\ N \quad\wedge\quad \big(\lfloor\tfrac{2M}{2N+1}\rfloor\big)^i \ \geq\ (1-\tilde\epsilon)\big(\tfrac{2M-(N+1)}{2N}\big)^i \ \geq\ (1-\tilde\epsilon)^2\big(\tfrac{M}{N}\big)^i \quad\wedge\quad \big(\lfloor\tfrac{2M}{2N+1}+1\rfloor\big)^i \ \leq\ (1+\tilde\epsilon)\big(\tfrac{M}{N}\big)^i.$$

Let $N \geq \tilde N(\tilde\epsilon, N_0)$, $M \geq \tilde M(\tilde\epsilon, N_0, N)$, $i \in \{0,\dots,N_0\}$, $Z \in \mathrm{S}_N(i)$. We have to distinguish two cases. In each case the first step will be an application of Proposition 2.

Case 1. $\#J(Z) = i$, that is f_Z has no discontinuity at 0. Then

$$\begin{aligned}
&\pi_M\big(\mathrm{U}_M^N(Z)\big) - \pi_N(Z)\\
=\ & \tfrac{1}{c_M}\big(\#\mathrm{U}_M^N(Z)\big)\big(\tfrac{\alpha_M}{1-\gamma_M}\big)^{i \bmod 2}\big(\tfrac{\alpha_M\cdot\gamma_M}{(1-\alpha_M)(1-\gamma_M)}\big)^{\lfloor\frac{i}{2}\rfloor} - \tfrac{1}{c_N}\big(\tfrac{\alpha_N}{1-\gamma_N}\big)^{i \bmod 2}\big(\tfrac{\alpha_N\cdot\gamma_N}{(1-\alpha_N)(1-\gamma_N)}\big)^{\lfloor\frac{i}{2}\rfloor}\\
\geq\ & \tfrac{1}{c_M}\big(\lfloor\tfrac{2M}{2N+1}\rfloor\big)^i(1-\tilde\epsilon)^2\big(\tfrac{\alpha}{M}\big)^{i \bmod 2}\big(\tfrac{\alpha\gamma}{M^2}\big)^{\lfloor\frac{i}{2}\rfloor} - \tfrac{1}{c_N}(1+\tilde\epsilon)^2\big(\tfrac{\alpha}{N}\big)^{i \bmod 2}\big(\tfrac{\alpha\gamma}{N^2}\big)^{\lfloor\frac{i}{2}\rfloor}\\
\geq\ & \tfrac{1}{c_\infty}\,\alpha^{i \bmod 2}\,(\alpha\gamma)^{\lfloor\frac{i}{2}\rfloor}\Big((1-\tilde\epsilon)^5\big(\tfrac{M}{N}\big)^i M^{-(i \bmod 2+2\lfloor\frac{i}{2}\rfloor)} - (1+\tilde\epsilon)^3 N^{-(i \bmod 2+2\lfloor\frac{i}{2}\rfloor)}\Big)\\
=\ & \tfrac{1}{c_\infty}\,\tfrac{1}{N^i}\,\alpha^{i \bmod 2}\,(\alpha\gamma)^{\lfloor\frac{i}{2}\rfloor}\big((1-\tilde\epsilon)^5 - (1+\tilde\epsilon)^3\big)\\
\geq\ & \tfrac{1}{c_\infty}\,\tfrac{1}{N^i}\,\eta(\alpha,\gamma,N_0)\,(1-5\tilde\epsilon-1-7\tilde\epsilon) \quad=\quad -\tfrac{12}{c_\infty}\,\tfrac{\tilde\epsilon}{N^i}\,\eta(\alpha,\gamma,N_0) \quad\geq\quad -\tfrac{12}{c_\infty}\,\tfrac{\tilde\epsilon}{N^i}\,\eta(\alpha,\gamma,N_0{+}1),
\end{aligned}$$

where $\eta(\alpha,\gamma,n) := \max\{\alpha^{k \bmod 2}\,(\alpha\gamma)^{\lfloor\frac{k}{2}\rfloor} \mid k \in \{0,\dots,n\}\}$. On the other hand,

$$\begin{aligned}
&\pi_M\big(\mathrm{U}_M^N(Z)\big) - \pi_N(Z)\\
\leq\ & \tfrac{1}{c_M}\big(\lfloor\tfrac{2M}{2N+1}\rfloor+1\big)^i(1+\tilde\epsilon)^2\tfrac{1}{M^i}\,\alpha^{i \bmod 2}\,(\alpha\gamma)^{\lfloor\frac{i}{2}\rfloor} - \tfrac{1}{c_N}(1-\tilde\epsilon)^2\tfrac{1}{N^i}\,\alpha^{i \bmod 2}\,(\alpha\gamma)^{\lfloor\frac{i}{2}\rfloor}\\
\leq\ & \tfrac{1}{c_\infty}\,\tfrac{1}{N^i}\,\alpha^{i \bmod 2}\,(\alpha\gamma)^{\lfloor\frac{i}{2}\rfloor}\big((1+\tilde\epsilon)^4 - (1-\tilde\epsilon)^3\big) \quad\leq\quad \tfrac{18}{c_\infty}\,\tfrac{\tilde\epsilon}{N^i}\,\eta(\alpha,\gamma,N_0{+}1).
\end{aligned}$$

Case 2. $\#J(Z) = i+1$, that is f_Z jumps at 0. Then

$$\begin{aligned}
&\pi_M\big(\mathrm{U}_M^N(Z)\big) - \pi_N(Z)\\
=\ & \tfrac{1}{c_M}\big(\#\mathrm{U}_M^N(Z)\big)\big(\tfrac{\alpha_M}{1-\gamma_M}\big)^{(i+1) \bmod 2}\big(\tfrac{\alpha_M\cdot\gamma_M}{(1-\alpha_M)(1-\gamma_M)}\big)^{\lfloor\frac{i+1}{2}\rfloor} - \tfrac{1}{c_N}\big(\tfrac{\alpha_N}{1-\gamma_N}\big)^{(i+1) \bmod 2}\big(\tfrac{\alpha_N\cdot\gamma_N}{(1-\alpha_N)(1-\gamma_N)}\big)^{\lfloor\frac{i+1}{2}\rfloor}\\
\geq\ & \tfrac{1}{c_\infty}\,\tfrac{1}{N^i}\,\alpha^{(i+1) \bmod 2}\,(\alpha\gamma)^{\lfloor\frac{i+1}{2}\rfloor}\big(\tfrac{(1-\tilde\epsilon)^5}{M} - \tfrac{(1+\tilde\epsilon)^3}{N}\big)\\
\geq\ & \tfrac{1}{c_\infty}\,\tfrac{1}{N^i}\,\eta(\alpha,\gamma,N_0+1)\,\tfrac{1+7\tilde\epsilon}{N} \quad\geq\quad \tfrac{8}{c_\infty}\,\tfrac{\tilde\epsilon}{N^i}\,\eta(\alpha,\gamma,N_0{+}1).
\end{aligned}$$

In the same way one obtains

$$\pi_M\big(\mathrm{U}_M^N(Z)\big) - \pi_N(Z) \ \leq\ \tfrac{16}{c_\infty}\,\tfrac{\tilde\epsilon}{N^i}\,\eta(\alpha,\gamma,N_0{+}1).$$

Set $\tilde{\epsilon} := \min\{\epsilon, \frac{c_\infty \epsilon}{18\eta(\alpha,\gamma,N_0+1)}\}$, and the proof of (64) is finished.

Let $\tilde{\epsilon} > 0$. Choose $\tilde{N}_0 \in \mathbb{N}$ according to (68) such that

$$\forall N_0 \geq \tilde{N}_0 \; \exists \tilde{N} \in \mathbb{N} \; \forall N \geq \tilde{N}: \quad 1 + \tilde{\epsilon} \;\geq\; \tfrac{c_N}{c_\infty} \;\geq\; \tfrac{c_{N,N_0}}{c_\infty} \;\geq\; 1 - \tilde{\epsilon}.$$

Making again use of Proposition 2 we have for $N_0 \geq \tilde{N}_0$, $N \geq \hat{N} = \hat{N}(\tilde{\epsilon}, N_0)$

$$\begin{aligned} &\sum_{i=N_0+1}^{N-1} \pi_N\big(\mathrm{S}_N(i)\big) \\ = \;& \tfrac{2}{c_N} \sum_{i=N_0+1}^{N-1} \tbinom{N-1}{i} \Big(\big(\tfrac{\alpha_N}{1-\gamma_N}\big)^{i \bmod 2} \big(\tfrac{\alpha_N \cdot \gamma_N}{(1-\alpha_N)(1-\gamma_N)}\big)^{\lfloor \frac{i}{2} \rfloor} + \big(\tfrac{\alpha_N}{1-\gamma_N}\big)^{(i+1) \bmod 2} \big(\tfrac{\alpha_N \cdot \gamma_N}{(1-\alpha_N)(1-\gamma_N)}\big)^{\lfloor \frac{i+1}{2} \rfloor} \Big) \\ \leq \;& \tfrac{1}{c_N}(c_N - c_{N,N_0}) \quad = \quad 1 - \tfrac{c_\infty}{c_N} \tfrac{c_{N,N_0}}{c_\infty} \quad \leq \quad \tfrac{\tilde{\epsilon}}{2}. \end{aligned}$$

This establishes (65). $\square$

B.2 Proof of Proposition 4

The first step is to show that the closure of $\{\tilde{\pi}_M \mid M \in \mathbb{N}\}$ is compact in the Prohorov topology of $\mathcal{M}^1_+(D_0)$. According to Theorem 4 it is sufficient to prove tightness of $\{\tilde{\pi}_M \mid M \in \mathbb{N}\}$, that is

$$\forall \epsilon > 0 \; \exists \tilde{K} \subset D_0 \text{ compact} : \inf\{\tilde{\pi}_M(\tilde{K}) \mid M \in \mathbb{N}\} \geq 1 - \epsilon,$$

where compactness means compactness with respect to the Skorokhod topology of D_0. Recall from Section 4.1 the definition of $\tilde{\iota}_M$. For all natural numbers $N_0 < N \leq M$ we have

$$\tilde{\pi}_M\Big(\tilde{\iota}_M\big(\mathrm{U}^N_M\big)\Big) \;\geq\; \pi_M\Big(\bigcup_{i=0}^{N_0} \bigcup_{Z \in \mathrm{S}_N(i)} \mathrm{U}^N_M(Z)\Big), \qquad \text{where} \quad \mathrm{U}^N_M := \bigcup_{Z \in \mathrm{S}_N} \mathrm{U}^N_M(Z).$$

Let $\epsilon > 0$. According to (63) we can find natural numbers $N_0 < N \leq \tilde{M}$ such that for all $M \geq \tilde{M}$:

$$\pi_M\Big(\bigcup_{i=0}^{N_0} \bigcup_{Z \in \mathrm{S}_N(i)} \mathrm{U}^N_M(Z)\Big) \;\geq\; 1 - \epsilon.$$

Fix N, $\tilde{M}$. In analogy to the definition of U^N_M we set

$$\tilde{A} := \bigcup_{Z \in \mathrm{S}_N} \Big\{ f \in D_0 \;\Big|\; \#\dot{J}(f) = \#\dot{J}(Z) \;\wedge\; \big(\exists \lambda \in \Lambda : \sup_{s \in [-r,0]} |\lambda(s) - s| \leq \tfrac{r}{2N+1} \;\wedge\; f \circ \lambda = f_Z\big) \Big\}.$$

Then $\#J(f) \leq N$ and $\zeta_f \geq \frac{2r}{N(2N+1)}$ for all $f \in \tilde{A}$, and by Lemma 9 we see that $\mathrm{cl}(\tilde{A})$, the closure of $\tilde{A}$, is compact with respect to the Skorokhod topology. By definition we have $\mathrm{U}^N_M \subset \tilde{A}$ for all $M \geq \tilde{M} \geq N$. Define

$$\tilde{K} := \bigcup_{M=1}^{\tilde{M}-1} \tilde{\iota}_M(\mathrm{S}_M) \cup \mathrm{cl}(\tilde{A}).$$

Then $\tilde{K}$ is compact in the Skorokhod topology, and with $M \in \mathbb{N}$ it holds that

$$\tilde{\pi}_M(\tilde{K}) \geq \begin{cases} \tilde{\pi}_M\big(\tilde{\iota}_M(\mathrm{S}_M)\big) = 1 & \text{if} \quad M \in \{1, \ldots, \tilde{M}-1\}, \\ \tilde{\pi}_M\big(\tilde{\iota}_M(\mathrm{U}^N_M)\big) \geq 1 - \epsilon & \text{if} \quad M \geq \tilde{M}. \end{cases}$$

Hence, $\inf\{\tilde{\pi}_M(\tilde{K}) \mid M \in \mathbb{N}\} \geq 1-\epsilon$. Since $\epsilon > 0$ was arbitrary, we now know that $\{\tilde{\pi}_M \mid M \in \mathbb{N}\}$ is relatively compact.

Let $(\tilde{\pi}_{M(j)})_{j\in\mathbb{N}}$ be a weakly convergent subsequence of $(\tilde{\pi}_M)_{M\in\mathbb{N}}$. Denote by $\tilde{\pi}$ the limit of $(\tilde{\pi}_{M(j)})$ in the Prohorov topology. We have to check that $\tilde{\pi}_M \xrightarrow{w} \tilde{\pi}$ as $M \to \infty$. Because of Theorem 3 it is sufficient to show that

$$\int_{D_0} \phi \; d\tilde{\pi}_M \overset{M\to\infty}{\longrightarrow} \int_{D_0} \phi \; d\tilde{\pi} \qquad \forall \, \phi \in \mathbf{C}_b(D_0) \text{ uniformly continuous.}$$

Let ϕ be a bounded and uniformly continuous real function on D_0 and set $K_\phi := \sup\{\, |\phi(f)| \mid f \in D_0\}$. With $M \in \mathbb{N}$ it holds that

$$\left|\int \phi \; d\tilde{\pi}_M - \int \phi \; d\tilde{\pi}\right| \;\leq\; \left|\int \phi \; d\tilde{\pi}_M - \int \phi \; d\tilde{\pi}_{M(j)}\right| + \left|\int \phi \; d\tilde{\pi}_{M(j)} - \int \phi \; d\tilde{\pi}\right| \quad \text{for all } j \in \mathbb{N}.$$

The convergence $\tilde{\pi}_{M(j)} \xrightarrow{w} \tilde{\pi}$ implies $|\int \phi \; d\tilde{\pi}_{M(j)} - \int \phi \; d\tilde{\pi}| \to 0$ as $j \to \infty$. We therefore have to show that

$$\forall \, \epsilon > 0 \; \forall \, j_0 \in \mathbb{N} \; \exists \, j \geq j_0 \; \exists \, \tilde{M} \in \mathbb{N} \; \forall \, M \geq \tilde{M} : \quad \left|\int \phi \; d\tilde{\pi}_M - \int \phi \; d\tilde{\pi}_{M(j)}\right| \;\leq\; \epsilon.$$

Let $\epsilon > 0$, $j_0 \in \mathbb{N}$. Choose natural numbers $N_0 = N_0(\epsilon)$, $\tilde{N} = \tilde{N}(\epsilon)$ according to (66). Choose $\delta = \delta(\epsilon, \phi) > 0$ such that $|\phi(f) - \phi(g)| \leq \epsilon$ for all $f, g \in D_0$ with $d_S(f,g) \leq \delta$. Let $j \in \mathbb{N}$ be big enough so that $j \geq j_0$, $M(j) \geq \tilde{N}$ and $\frac{r}{2M(j)+1} \leq \delta$. Set $N := M(j)$.

Recalling the definition of our approximation sets[13] we see that $d_S(f_Z, f_{\tilde{Z}}) \leq \delta$ for all $Z \in \mathrm{S}_N(i)$ and $\tilde{Z} \in \mathrm{U}^N_M(Z)$ if $i \in \{0, \ldots, N_0\}$ and $M \geq N$. By the choice of δ this implies that $|\phi(f_Z) - \phi(f_{\tilde{Z}})| \leq \epsilon$ for all such Z, $\tilde{Z}$.

Finally, choose a natural number $\tilde{M} = \tilde{M}(\epsilon, N_0, \tilde{N}, N)$ according to (66). Then for $M \geq \tilde{M}$

$$\begin{aligned}
&\left|\int \phi \; d\tilde{\pi}_M - \int \phi \; d\tilde{\pi}_N\right| \\
\leq\;& 2K_\phi \cdot \epsilon \;+\; \sum_{i=0}^{N_0} \sum_{Z \in \mathrm{S}_N(i)} \left| \phi(f_Z)\, \pi_N(Z) \;-\; \sum_{\tilde{Z} \in \mathrm{U}^N_M(Z)} \phi(f_{\tilde{Z}})\, \pi_M(\tilde{Z}) \right| \\
\leq\;& 2K_\phi \cdot \epsilon \;+\; \sum_{i=0}^{N_0} \sum_{Z \in \mathrm{S}_N(i)} |\phi(f_Z)| \cdot \left|\pi_M\big(\mathrm{U}^N_M(Z)\big) - \pi_N(Z)\right| \;+\; \sum_{\tilde{Z} \in \mathrm{U}^N_M(Z)} \big(\phi(f_{\tilde{Z}}) - \phi(f_Z)\big)\pi_M(\tilde{Z}) \\
\leq\;& 2K_\phi \cdot \epsilon \;+\; \epsilon \cdot \left(\sum_{i=0}^{N_0} \sum_{Z \in \mathrm{S}_N(i)} \pi_M\big(\mathrm{U}^N_M(Z)\big) \right) \;+\; K_\phi \cdot \epsilon \cdot \left(\sum_{i=0}^{N_0} N^{-i} \big(\#\mathrm{S}_N(i)\big) \right) \\
\leq\;& 2K_\phi \cdot \epsilon \;+\; \epsilon \cdot \pi_M\big(\mathrm{S}_M\big) \;+\; 4K_\phi \cdot \epsilon \cdot \left(\sum_{i=0}^{N_0} N^{-i} \tbinom{N-1}{i} \right) \\
\leq\;& 2K_\phi \cdot \epsilon \;+\; \epsilon \;+\; 4K_\phi \cdot \epsilon \cdot \left(\sum_{i=0}^{\infty} \tfrac{1}{i!} \right) \quad = \quad (2K_\phi + 1 + 4K_\phi e) \cdot \epsilon.
\end{aligned}$$

□

B.3 Proof of Proposition 7

Clearly, $F_\delta(0) = 0 = F(0)$ for all $\delta \in (0,1)$. With $M \in \mathbb{N}$ let $\tilde{\mathrm{P}}_M$ be the probability measure on $\mathcal{B}(D_\infty)$ as defined in Section 4.1. Recall that $\tilde{\mathrm{P}}_M$ is the measure induced by the sequence of current states of

[13] The sets $\mathrm{U}^N_M(Z)$ were defined at the beginning of Section 4.2.

the Markov chain X^M under P_{π_M}, i.e. in the stationary regime. For $\delta \in (0,1)$, $M \in \mathbb{N}$ set

$$F_\delta^M(t) := \tilde{\mathrm{P}}_M\big(\tau \le t \mid \tilde{A}_\delta\big), \quad t \in [0,\infty).$$

From Proposition 6 we know that $\tilde{\mathrm{P}}_M \xrightarrow{w} \tilde{\mathrm{P}}$ as M tends to infinity. Check that for $\delta \in (0,1)$, $t \in (0,\infty)$ the events $\tilde{A}_\delta$, $\tilde{A}_\delta \cap \{\tau \le t\}$ are $\tilde{\mathrm{P}}$-continuity sets of $\mathcal{B}(D_\infty)$. An application of Theorem 3 yields

$$(69) \quad \tilde{\mathrm{P}}_M\big(\tau \le t \mid \tilde{A}_\delta\big) \overset{M\to\infty}{\longrightarrow} \tilde{\mathrm{P}}\big(\tau \le t \mid \tilde{A}_\delta\big), \quad \text{i.e.} \quad F_\delta^M(t) \overset{M\to\infty}{\longrightarrow} F_\delta(t) \quad \text{for all } \delta \in (0,1),\ t > 0.$$

For all $t > 0$, $\delta \in (0,1)$, $M \in \mathbb{N}$ we have

$$\big|F_\delta(t) - F(t)\big| \quad \le \quad \big|F_\delta(t) - F_\delta^M(t)\big| \ + \ \big|F_\delta^M(t) - F(t)\big|.$$

In view of (69) it is sufficient to show that for each $t > 0$ and each $\varepsilon > 0$ there are $\delta_0 \in (0,1)$, $M_0 \in \mathbb{N}$ such that

$$(70) \qquad \big|F_\delta^M(t) - F(t)\big| \ \le \ \varepsilon \quad \text{for all } \delta \in (0,\delta_0),\ M \ge M_0.$$

As in Section 3.2, let $(Y_n^M)_{n\in\{-M,-M+1,\dots\}}$ be the random sequence of current states on $(\Omega,\mathcal{F})$ at discretization degree $M \in \mathbb{N}$. Let $\delta \in (0,1)$ and let $M \in \mathbb{N}$ be such that $\delta \cdot M \ge 1$. Set

$$A_{\delta,M}^j := \big\{Y_{-\lfloor \delta M\rfloor}^M = -1, \dots, Y_{-j-1}^M = -1, Y_{-j}^M = 1, \dots, Y_0^M = 1\big\}, \qquad j \in \{0,\dots,\lfloor \delta M\rfloor - 1\}.$$

Notice that $A_{\delta,M}^j$ is an event in $\mathcal{F}$. The corresponding event in $\mathcal{B}(D_\infty)$ is given by

$$\tilde{A}_{\delta,M}^j := \big\{f \in D_\infty \mid \forall l \in \{j+1,\dots,\lfloor \delta M\rfloor\}\colon\ f\big(-\tfrac{l}{M}r\big) = -1 \ \wedge\ \forall l \in \{0,\dots,j\}\colon\ f\big(-\tfrac{l}{M}r\big) = 1\big\}.$$

For all $\delta \in (0,1)$ and all $M \in \mathbb{N}$ such that $\delta \cdot M \ge 1$ it holds that

$$\mathrm{P}_{\pi_M}\big(A_{\delta,M}^j\big) \ = \ \tilde{\mathrm{P}}_M\big(\tilde{A}_{\delta,M}^j\big) \qquad \text{for all } j \in \{0,\dots,\lfloor \delta M\rfloor - 1\},$$

$$\tilde{A}_\delta \ = \ \tilde{A}_{\delta,M}^0 \cup \ldots \cup \tilde{A}_{\delta,M}^{\lfloor \delta M\rfloor - 1} \qquad \tilde{\mathrm{P}}_M\text{-almost surely.}$$

In analogy to (26), the definition of the residence time distribution $L_M(.)$ of discretization degree M, we set

$$L_{\delta,M}^j(k) := \mathrm{P}_{\pi_M}\big(Y_0^M = 1, \dots, Y_{k-1}^M = 1, Y_k^M = -1 \mid A_{\delta,M}^j\big), \quad k \in \mathbb{N}.$$

Then, by construction of F_δ^M, for all $\delta \in (0,1)$ and all $M \in \mathbb{N}$ such that $\delta \cdot M \ge 1$ we have

$$F_\delta^M(t) \ = \ \sum_{k=1}^{\lfloor \frac{t}{r} M\rfloor} \ \sum_{j=0}^{\lfloor \delta M\rfloor - 1} \tilde{\mathrm{P}}_M\big(\tilde{A}_{\delta,M}^j \mid \tilde{A}_\delta\big) \cdot L_{\delta,M}^j(k), \quad t > 0.$$

It is not necessary to calculate the probabilities $\tilde{\mathrm{P}}_M(\tilde{A}_{\delta,M}^j \mid \tilde{A}_\delta)$. Instead, proceeding in a way very much as in Section 3.2, we will estimate limes inferior and limes superior of $M \cdot L_{\delta,M}^{j_M}(\lfloor qM\rfloor)$ as M tends to infinity, where $q > 0$ and $(j_M) \subset \mathbb{N}_0$ is any sequence such that $j_M \in \{0,\dots,\lfloor \delta M\rfloor - 1\}$ for all $M \in \mathbb{N}$. The estimates will be uniform in $\delta \in (0,\delta_0]$ for any small $\delta_0 > 0$.

In analogy to (27), the definition of the tail constant K_M, we set for $\delta \in (0,1)$ and M big enough

$$K_{\delta,M}^j := \mathrm{P}_{\pi_M}\big(Y_{-\lfloor \delta M\rfloor}^M = 1, \dots, Y_{-j-1}^M = 1, Y_{-j}^M = 1, \dots, Y_{M-\lfloor \delta M\rfloor}^M = 1 \mid A_{\delta,M}^j\big).$$

Because of the shift invariance of Y^M under P_{π_M}, the above definition of $K_{\delta,M}^j$ is really analogue to that of K_M. Exploiting the stationarity of X^M under P_{π_M}, we obtain

$$K_{\delta,M}^j \quad = \quad \frac{\pi_M\big((\overbrace{-1,\dots,-1}^{\lfloor \delta M\rfloor - j}, \overbrace{1,\dots,1}^{M-\lfloor \delta M\rfloor + j+1})\big)}{\pi_M\big(\{(\underbrace{*,\dots,*}_{M-\lfloor \delta M\rfloor}, \underbrace{-1,\dots,-1}_{\lfloor \delta M\rfloor - j}, \underbrace{1,\dots,1}_{j+1})\}\big)}.$$

As a consequence of Proposition 2, the formula for the stationary distributions π_M, we see that $K^j_{\delta,M}$ is the same for all $j \in \{0,\dots,\lfloor \delta M\rfloor -1\}$. Proceeding as in the derivation of (29) we find that

$$(71)\qquad K^j_{\delta,M} \;=\; \frac{2}{\Big(1+\sqrt{\frac{\tilde\gamma_M}{\tilde\alpha_M}}\Big)\big(1+\sqrt{\tilde\eta_M}\big)^{M-\lfloor\delta M\rfloor} + \Big(1-\sqrt{\frac{\tilde\gamma_M}{\tilde\alpha_M}}\Big)\big(1-\sqrt{\tilde\eta_M}\big)^{M-\lfloor\delta M\rfloor}} \;=:\; K_{\delta,M}.$$

In order to calculate $L^j_{\delta,M}$ we apply Proposition 2 again in a way similar to that of Section 3.2. Let $\delta\in(0,1)$, let $M\in\mathbb{M}$ be such that $\lfloor\delta M\rfloor\ge 1$, and $j\in\{0,\dots,\lfloor\delta M\rfloor-1\}$. Then for $k\in\{1,\dots,M-\lfloor\delta M\rfloor\}$

$$(72a)\qquad \begin{aligned} L^j_{\delta,M}(k) \;=\; \frac{\sqrt{\tilde\gamma_M}}{2}\cdot K_{\delta,M}\cdot\Big(&\sqrt{\tilde\gamma_M}\,\big((1+\sqrt{\tilde\eta_M})^{M-\lfloor\delta M\rfloor-k} + (1-\sqrt{\tilde\eta_M})^{M-\lfloor\delta M\rfloor-k}\big)\\ &+\sqrt{\tilde\alpha_M}\,\big((1+\sqrt{\tilde\eta_M})^{M-\lfloor\delta M\rfloor-k} - (1-\sqrt{\tilde\eta_M})^{M-\lfloor\delta M\rfloor-k}\big)\Big).\end{aligned}$$

While $L^j_{\delta,M}(k)$ in (72a) does not vary with j as long as $k\le M-\lfloor\delta M\rfloor$, for $k\in\{M-\lfloor\delta M\rfloor+1,\dots,M-j\}$ it holds that

$$(72b)\qquad L^j_{\delta,M}(k) \;=\; K_{\delta,M}\cdot\gamma_M\cdot(1-\gamma_M)^{k-M+\lfloor\delta M\rfloor-1},$$

and for $k\ge M-j+1$ we have

$$(72c)\qquad L^j_{\delta,M}(k) \;=\; K_{\delta,M}\cdot\alpha_M\cdot(1-\gamma_M)^{\lfloor\delta M\rfloor-j}\cdot(1-\alpha_M)^{k-M+j-1}.$$

Now, let the discretization degree M tend to infinity, where we assume that scaling relation (31) holds for some rates α, γ. From (71) we see that

$$(73)\qquad \begin{aligned} K_{\infty,\delta} := \lim_{M\to\infty} K_{M,\delta} \;&=\; \frac{2}{\big(1+\sqrt{\frac{\gamma}{\alpha}}\big)e^{(1-\delta)\sqrt{\alpha\gamma}} + \big(1-\sqrt{\frac{\gamma}{\alpha}}\big)e^{-(1-\delta)\sqrt{\alpha\gamma}}}\\ &=\; \frac{\sqrt{\alpha}}{\sqrt{\alpha}\cosh\big((1-\delta)\sqrt{\alpha\gamma}\big)+\sqrt{\gamma}\sinh\big((1-\delta)\sqrt{\alpha\gamma}\big)}.\end{aligned}$$

Let $q>0$ and let $(j_M)\subset\mathbb{N}_0$ be any sequence such that $j_M\in\{0,\dots,\lfloor\delta M\rfloor-1\}$ for all $M\in\mathbb{N}$. If $q\in(0,1-\delta]$, then from (72a) we find that

$$(74a)\qquad \lim_{M\to\infty} M\cdot L^{j_M}_{\delta,M}\big(\lfloor qM\rfloor\big) \;=\; \sqrt{\gamma}\cdot K_{\delta,\infty}\cdot\Big(\sqrt{\gamma}\,\cosh\big(\sqrt{\alpha\gamma}(1-\delta-q)\big)+\sqrt{\alpha}\,\sinh\big(\sqrt{\alpha\gamma}(1-\delta-q)\big)\Big).$$

If $q\in(1-\delta,1)$, then a rough estimate of (72b) and (72c), respectively, yields

$$(74b)\qquad \begin{aligned} \limsup_{M\to\infty} M\cdot L^{j_M}_{\delta,M}\big(\lfloor qM\rfloor\big) \;&\le\; \max\{\alpha,\gamma\}\cdot K_{\delta,\infty},\\ \liminf_{M\to\infty} M\cdot L^{j_M}_{\delta,M}\big(\lfloor qM\rfloor\big) \;&\ge\; \min\{\alpha,\gamma\}\cdot K_{\delta,\infty}\cdot e^{-\delta\gamma}\cdot e^{-(q+\delta-1)\alpha}.\end{aligned}$$

On the other hand, if $q\ge1$, then by (72c) we have

$$(74c)\qquad \begin{aligned} \limsup_{M\to\infty} M\cdot L^{j_M}_{\delta,M}\big(\lfloor qM\rfloor\big) \;&\le\; \alpha\cdot K_{\delta,\infty}\cdot e^{-(q-1)\alpha},\\ \liminf_{M\to\infty} M\cdot L^{j_M}_{\delta,M}\big(\lfloor qM\rfloor\big) \;&\ge\; \alpha\cdot K_{\delta,\infty}\cdot e^{-\delta\gamma}\cdot e^{-(q+\delta-1)\alpha}.\end{aligned}$$

Notice that convergence in (73) as well as in (74) is uniform in $\delta\in(0,\delta_0]$ for arbitrary $\delta_0\in(0,1)$. If we let δ tend to zero, we recover the residence time distribution density f_L of Proposition 3. Taking the time discretization into account, we obtain $\tilde f_L$ as given by (36) instead of f_L.

Given $t>0$, $\varepsilon>0$, uniform convergence of $(L^j_{\delta,M})$ in δ and dominated convergence of the corresponding residence time distribution densities over the interval $(0,t]$ imply that we can find $\delta_0\in(0,1)$ and $M_0\in\mathbb{N}$ such that inequality (70) is fulfilled. The assertion of Proposition 7 then follows.

□

References

[1] R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani. The mechanism of stochastic resonance. *J. Phys. A, Math. Gen.*, 14:453–457, 1981.

[2] R. Benzi, G. Parisi, A. Sutera, and A. Vulpiani. Stochastic resonance in climatic change. *Tellus*, 34:10–16, 1982.

[3] C. Nicolis. Stochastic aspects of climatic transitions – responses to periodic forcing. *Tellus*, 34:1–9, 1982.

[4] Luca Gammaitoni, Peter Hänggi, Peter Jung, and Fabio Marchesoni. Stochastic resonance. *Rev. Mod. Phys.*, 70(1).

[5] V. S. Anishchenko, A. B. Neiman, F. Moss, and L. Schimansky-Geier. Stochastic resonance: noise-enhanced order. *Physics – Uspekhi*, 42(1):7–36, 1999.

[6] A. Krawiecki and J.A. Holyst. Stochastic resonance as a model for financial market crashes and bubbles. *Physica A*, 317(3-4):597–608, 2003.

[7] Ilya E. Pavlyukevich. *Stochastic Resonance*. PhD thesis, Humboldt-Universität zu Berlin, 2002.

[8] Peter Imkeller and Ilya Pavlyukevich. Model reduction and stochastic resonance. *Stoch. Dyn.*, 2(4): 463–506, 2002.

[9] B. McNamara and K. Wiesenfeld. Theory of stochastic resonance. *Phys. Rev. A*, 39(9):4854–4869, 1989.

[10] Samuel Herrmann, Peter Imkeller, and Ilya Pavlyukevich. Stochastic resonance: non-robust and robust tuning notions. In K. Haman, B. Jakubiak, and J. Zabczyk, editors, *Probabilistic Problems in Atmospheric and Water Sciences. Proceedings of the workshop held in Bedlewo Mathematical Conference Center, 16-18 December 2002*, pages 61–90, Warszawa, 2003. Wydawnictwa ICM.

[11] Toru Ohira and Yuzuru Sato. Resonance with noise and delay. *Phys. Rev. Lett.*, 82:2811–2815, 1999.

[12] L. S. Tsimring and A. Pikovsky. Noise-induced dynamics in bistable systems with delay. *Phys. Rev. Lett.*, 87(25):250602, 2001.

[13] Cristina Masoller. Distribution of residence times of time-delayed bistable systems driven by noise. *Phys. Rev. Lett.*, 90(2):020601, 2003.

[14] J. Houlihan, D. Goulding, Th. Busch, C. Masoller, and G. Huyet. Experimental investigation of a bistable system in the presence of noise and delay. *Phys. Rev. Lett.*, 92(5):050601, 2004.

[15] D. Curtin, P. Hegarty, S. D. Goulding, J. Houlihan, Th. Busch, C. Masoller, and G. Huyet. Distribution of residence times in bistable noisy systems with time-delayed feedback. *Phys. Rev. E*, 70 (3):031103, 2004.

[16] Salah-Eldin A. Mohammed. *Stochastic functional differential equations*. Pitman Publishing, London, 1984.

[17] Salah-Eldin A. Mohammed. Stochastic differential systems with memory: Theory, examples and applications. In L. Decreusefond, Jon Gjerde, B. Øksendal, and A. S. Üstünel, editors, *Stochastic Analysis and Related Topics VI. Papers from the Sixth Oslo-Silivri Workshop on Stochastic Analysis held in Geilo*, pages 1–77, Boston, 1996. Birkhäuser.

[18] Richard Durrett. *Stochastic Calculus. A Practical Introduction*. Probability and Stochastics Series. CRC Press, Boca Raton (FL), 1996.

[19] Daniel W. Stroock and S. R. Srinivasa Varadhan. *Multidimensional Diffusion Processes*, volume 233 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin, 1979.

[20] Samuel Herrmann, Peter Imkeller, and Dierk Peithmann. Large deviations for diffusions with time

periodic drift and stochastic resonance. Technical report, Humboldt-Universität zu Berlin and Université de Nancy, April 2005.

[21] Michael Scheutzow. *Qualitatives Verhalten der Lösungen von eindimensionalen nichtlinearen stochastischen Differentialgleichungen mit Gedächtnis.* PhD thesis, Fachbereich Mathematik der Universität Kaiserslautern, 1983.

[22] Michael Scheutzow. Qualitative behaviour of stochastic delay equations with a bounded memory. *Stochastics*, 12:41–80, 1984.

[23] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus.* Springer-Verlag, New York, 2nd edition, 1991.

[24] Brian F. Redmonda, Victor G. LeBlanc, and André Longtin. Bifurcation analysis of a class of first-order nonlinear delay-differential equations with reflectional symmetry. *Physica D*, 166:131–146, 2002.

[25] Pierre Brémaud. *Markov chains: Gibbs fields, Monte Carlo simulation, and queues.* Springer-Verlag, New York, 1999.

[26] Robert Lorenz. *Weak Approximation of Stochastic Differential Equations with Bounded Memory by Discrete Time Series.* PhD thesis, Humboldt-Universität zu Berlin, 2005.

[27] Jean Jacod and Albert N. Shiryaev. *Limit Theorems for Stochastic Processes*, volume 288 of *Grundlehren der mathematischen Wissenschaften.* Springer-Verlag, Berlin, 1987.

[28] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems*, volume 260. Springer-Verlag, New York, 2nd edition, 1998.

[29] Anton Bovier, Michael Eckhoff, Véronique Gayrard, and Markus Klein. Metastability in reversible diffusion processes. I. Sharp asymptotics for capacities and exit times. *J. Eur. Math. Soc. (JEMS)*, 6(4):399–424, 2004.

[30] Anton Bovier, Véronique Gayrard, and Markus Klein. Metastability in reversible diffusion processes. II. Precise asymptotics for small eigenvalues. *J. Eur. Math. Soc. (JEMS)*, 7(1):69–99, 2005.

[31] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes: Characterization and Convergence.* Wiley Series in Probability and Statistics. John Wiley & Sons, New York, 1986.

[32] Patrick Billingsley. *Convergence of probability measures.* Wiley series in Probability and Statistics. John Wiley & Sons, New York, 2nd edition, 1999.