A two state model for noise-induced resonance in bistable systems with delay

Markus Fischer

4th March 2005

Abstract

The subject of the present work 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 behaviour 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.

The present work is a slightly modified version of my Diploma thesis, which was submitted to the Department of Mathematics at Humboldt University Berlin in March 2004. The thesis was supervised by Prof. Peter Imkeller and Prof. Salah-Eldin A. Mohammed.

Author's address:

Markus Fischer
Reichenberger Str. 166
10999 Berlin
Germany
E-Mail: markus.fischer@alumni.hu-berlin.de

Zusammenfassung

Gegenstand der vorliegenden Arbeit ist ein vereinfachtes Modell zur Beschreibung des Auftretens eines als stochastische Resonanz bekannten Phänomens in Systemen mit Gedächtnis. Stochastische Resonanz ist ein quasi-deterministisches Verhalten, das sich in zufällig gestörten Systemen bei einer bestimmten niedrigen, aber von null verschiedenen Intensität des Störrauschens zeigt.

Das Ausgangsmodell für ein solches System ist hier durch eine eindimensionale parametrisierte stochastische Differentialgleichung mit Gedächtnis gegeben. Einige wesentliche Eigenschaften der Gleichung werden vorgestellt.

Die Konstruktion des vereinfachten Modells erfolgt in zwei Schritten, zunächst in diskreter Zeit, dann im Grenzübergang zu stetiger Zeit. Dadurch ist es möglich, Größen anzugeben, die das Phänomen der stochastischen Resonanz im vereinfachten Modell quantitativ erfassen. Insbesondere lassen sich die stationäre Verteilung und die Verteilung der Aufenthaltszeiten explizit berechnen. Letztere dient der Definition zweier Resonanzmaße. Schließlich wird eine Heuristik zur Herstellung der Verbindung mit dem Ausgangsmodell entwickelt.

Danksagung

Für den Vorschlag des Themas dieser Arbeit und die über den gesamten Zeitraum ihrer Entstehung hin gewährte Unterstützung danke ich Prof. Peter Imkeller. Für hilfreiche Gespräche und Einblick in ihre Arbeit danke ich Uwe Gross, Robert Lorenz, Ilya Pavlyukevich, Dierk Peithmann und Markus Reiß. Scließlich danke ich Prof. Salah-Eldin A. Mohammed, dass er sich des Themas angenommen und das zweite Gutachten verfasst hat.

Contents

Chapter 1

Introduction

The subject of the present work is a simplified model for a symmetric bistable system with memory, the reference model, which in the presence of noise exhibits a phenomenon similar to what is known as stochastic resonance.

A system with memory or *delay* is one whose evolution depends not only on the current state but also on the past. The reference model is given by a one dimensional parametrized stochastic differential equation with point delay. It will be described in chapter 2.

With a view to capturing the effective dynamics and, in particular, the resonance-like behaviour 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. This is done in chapters 3 and 4. 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, in chapter 5 we present 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.

In sections 1.1 and 1.2 of this chapter we briefly describe what stochastic resonance is and how to quantify it. In section 1.3 we introduce the idea of model reduction, which is fundamental to the approach taken here. Results from the physics literature about noise-induced resonance in systems with delay are summarized in section 1.4.

1.1 The phenomenon of stochastic resonance

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 two macroscopic states in the output signal.

Let us 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 figure 1.1 a). The barrier mentioned 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 given by the stochastic differential equation (SDE)

$$dX(t) \;=\; -\Big(V'\big(X(t)\big) + a\cdot\sin\big(\tfrac{2\pi}{T}t\big)\Big)dt \;+\; \sigma\cdot dW(t), \qquad t\geq 0, \tag{1.1}$$

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 (1.1) 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 a periodic force. It was originally proposed by Benzi et al. (1981, 1982) and Nicolis (1982) as an energy balance model designed to explain the quasi periodicity in climate changes over the last 700000 years, where the solution process describes the average global temperature on Earth and the weak input signal reflects the variation of the solar constant due to an oscillation of the Earth's eccentricity caused by the gravitational influence of Jupiter.[1]

If $a = 0$, i.e. in the absence of a periodic signal, then equation (1.1) reduces to

$$dX(t) \;=\; -V'\big(X(t)\big)dt \;+\; \sigma \cdot dW(t), \qquad t \geq 0. \tag{1.2}$$

The autonomous SDE (1.2) 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, one corresponding to the quadratic variation of the Wiener process, the other proportional to the average time it takes the diffusion to travel from one of the metastable states to the other.

In section 5.1 we cite results which give approximations for the mean intrawell residence time and the residence time distribution in case $a = 0$. 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 behaviour.

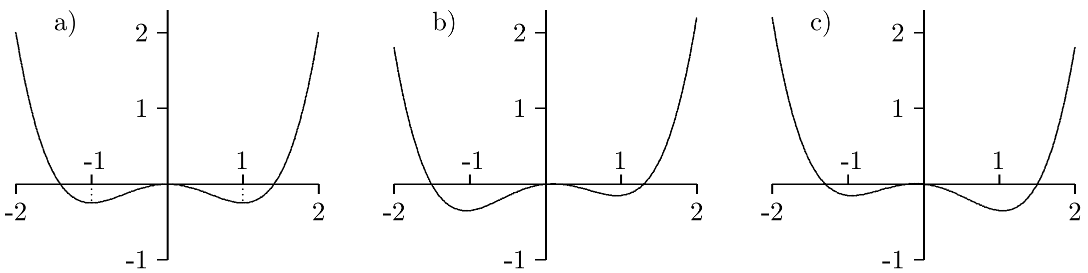

Figure 1.1: Graphs on the interval $[-2, 2]$ of a) symmetric quartic double well potential $V : x \mapsto \frac{1}{4}x^4 - \frac{1}{2}x^2$, b) tilted double well potential $x \mapsto V(x) + ax$, c) tilted double well potential $x \mapsto V(x) - ax$, where $a = 0.1$.

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, see figure 1.1 b) - c), where deflection is maximal. 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 behaviour, thereby inducing an amplification of the input signal.

For a more comprehensive description of stochastic resonance, examples, variants and applications thereof see Gammaitoni et al. or Anishchenko et al. (1999). What models that exhibit stochastic resonance have

[1] Cf. the introduction in Imkeller and Pavlyukevich (2002).

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. We now turn to the question of how to quantify it.

1.2 Measures of resonance

In view of the fact that the system given by equation (1.1) 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 defined as

$$\eta^X(\sigma, T) := \left| \int_0^1 E_\mu\Big(X(Ts) \cdot \exp\big(2\pi\sqrt{-1}s\big)\Big) ds \right|^2,$$

where X is a solution to (1.1) with noise parameter $\sigma > 0$ and input period $T > 0$ and μ is the invariant probability measure on the enlarged state space $\mathbb{R} \times S^1$.[2]

In general, when measuring stochastic resonance, it is assumed that the solution is in a "stationary regime". Since equation (1.1) is time dependent for $a > 0$ we cannot expect $(X(t))_{t \geq 0}$ to be a stationary process. Transforming the non-autonomous SDE (1.1) 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 (2002). In chapter 3, when constructing our reduced model, we will make use of the same idea of appropriately enlarging the state space in order to regain a time homogeneous Markov model.

There are other measures of resonance based on the frequency spectrum, e. g. the signal to noise ratio; for a detailed analysis see Pavlyukevich (2002).

A different starting point for a measure of resonance is the distribution of intrawell residence times. Observe that – because of symmetry – the roles of the two potential wells are interchangeable. We will return to the residence time distribution in the next two sections.

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.

1.3 The idea of model reduction

The fact that with $\sigma > 0$ and $a > 0$ small a typical solution to (1.1) 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.

How can the effective dynamics of equation (1.1) be imitated in the two state model? According to an idea of McNamara and Wiesenfeld (1989) one constructs 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. At times $k \cdot T + \frac{T}{4}$, $k \in \mathbb{N}$, for example, the rate of transition from -1 to 1 is the rate at which a diffusion started in the left well would – in the absence of periodic forcing – leave that well and escape to the right well, see figure 1.1 b).

[2] The circle S^1 corresponds to the phase angle of the periodic input signal. Note that time has been rescaled.

An approximation of the rate of escape from a (parabolic) potential well is given in the limit of small noise by the Kramers formula. In section 5.1 we cite this empirical formula together with mathematical results that partially justify its application.

In the physics literature, a standard ansatz for calculating the two state process given time dependent transition rates $W_{\pm}(.)$ is to solve an associated differential equation for the probabilities $n_{\pm}(t)$ of occupying state ± 1 at time t, the so-called master equation (cf. Gammaitoni et al.):

$$\dot{n}_{\pm}(t) \;=\; -\big(W_{\pm}(t) + W_{\mp}(t)\big)n_{\pm}(t) + W_{\pm}(t), \qquad t \geq 0. \tag{1.3}$$

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 (2002). 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. (2003) show. The simple measures of resonance which we use in chapters 3 and 5 in connection with the delay equation will be based on the distribution of residence times in the reduced model belonging to that equation. Although it is our expectation that those measures are robust under state reduction, we cannot prove this, here.

1.4 Existing results about noise-induced resonance with delay

In equation (1.1) 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 (2.1).

The idea to study such equations with regard to noise-induced resonance seems to originate with Ohira and Sato (1999). 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 (2.1) can be found in Tsimring and Pikovsky (2001). The same model is the object of recent studies by Masoller (2003), Houlihan et al. (2004) and Curtin et al. (2004), and it will be our standard example, too.

While the measure of resonance applied by Tsimring and Pikovsky (2001) 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. Approximative analytical results are obtained via the master equation approach, where in place of equation (1.3) a deterministic delay differential equation is considered and some simplifying assumptions are made.

In chapter 5 we will follow Tsimring and Pikovsky (2001) in establishing the link between the reduced and the reference model. Results by Masoller (2003) 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 will be different, though, in that we do not use any kind of master equation. Instead, we will 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.

Chapter 2

The reference model

In what follows we specify a model that shows resonance-like behaviour for suitable parameter choices. This will be our reference model. It is given by a parameterised one dimensional stochastic delay differential equation (SDDE), our reference equation, and comprises the example that was studied by Tsimring and Pikovsky (2001).

Basic properties of the reference equation are discussed in section 2.2, while section 2.3 heuristically describes "interesting" parameter regions. A comparison between noise-induced resonance with delay and classical stochastic resonance concludes this chapter.

2.1 Motion of a Brownian particle in a symmetric bistable system

A system that exhibits the phenomenon we are interested in can be illustrated as follows: 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 (2001).[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 figure 2.1).

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

$$dX(t) \;=\; -\big(V'\big(X(t)\big) + \beta \cdot U'\big(X(t-r)\big)\big)\,dt \;+\; \sigma \cdot dW(t), \qquad t \geq 0, \tag{2.1}$$

where $W(.)$ is a standard one dimensional Wiener process, $r > 0$ 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

[1] Our notation is slightly different from that of equation (1) in Tsimring and Pikovsky (2001: p. 1). In particular, their parameter ϵ, indicating the "strength of the feedback", corresponds to $-\beta$, here.

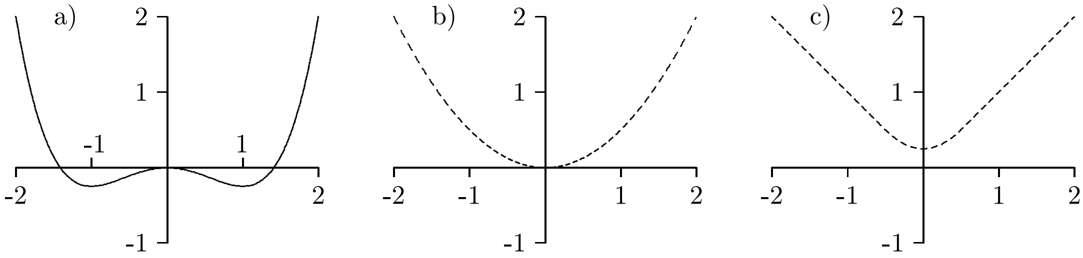

Figure 2.1: Graphs on the interval $[-2,2]$ of a) quartic double well potential V, b) quadratic delay potential $U : x \mapsto \frac{1}{2}x^2$, c) absolute value delay potential $U : x \mapsto |x|, x \in \mathbb{R} \setminus (-\epsilon, \epsilon)$, smoothly continued on $(-\epsilon, \epsilon)$.

and $\sigma \geq 0$ a noise parameter. In the special case $\sigma = 0$ equation (2.1) becomes a deterministic delay differential equation, while in case $\beta = 0$ we have an SDE of the form (1.2).

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

(2.2a)	$V, U \in \mathbf{C}^2(\mathbb{R})$,	
(2.2b)	$V(x) = V(-x)$,	$U(x) = U(-x)$ for all $x \in \mathbb{R}$,
(2.2c)	$V'(x) = 0$ iff $x \in \{-1, 0, 1\}$,	$U'(x) = 0$ iff $x = 0$,
(2.2d)	$V''(-1) = V''(1) > 0$,	
(2.2e)	$\sup\{V'(x) \mid x \in (-\infty, -1) \cup (0,1)\} \leq 0$,	$\sup\{U'(x) \mid x \in (-\infty, 0)\} \leq 0$,
(2.2f)	$\inf\{V'(x) \mid x \in (-1, 0) \cup (1, \infty)\} \geq 0$,	$\inf\{U'(x) \mid x \in (0, \infty)\} \geq 0$.

A further restriction on the geometry of V and U will prove useful in section 2.2, where we derive some properties of equation (2.1). Let us assume that a constant R_{pot} greater than the positive root of V exists such that V and U are linear on $\mathbb{R} \setminus (-R_{pot}, R_{pot})$. In view of the symmetry of V and U it is sufficient to require that

(2.3) there exists $R_{pot} > 1$ such that for all $x \in [R_{pot}, \infty)$:
$$V(x) = V'(R_{pot}) \cdot (x - R_{pot}) + V(R_{pot}) \quad \text{and} \quad U(x) = U'(R_{pot}) \cdot (x - R_{pot}) + U(R_{pot}).$$

Henceforth, whenever the reference model is concerned, we will suppose that conditions (2.2) and (2.3) are satisfied. Certainly, (2.3) is more restrictive than necessary, yet it is not too limiting, since R_{pot} can be chosen arbitrarily big, while the pathwise solutions to (2.1) will stay with high probability in a bounded interval containing the positions of the two minima of the double well potential V, provided the parameters σ and $|\beta|$ are small enough, cf. sections 2.3 and 5.1.

2.2 Properties of the underlying SDDE

By applying results from the literature we check some important properties of (2.1), the reference model equation. Our main sources are Mohammed (1984, 1996) with regard to existence and uniqueness of solutions and the Markov property, and Scheutzow (1983, 1984), where conditions can be found that guarantee the existence of a stationary distribution.

2.2.1 Strong and weak solutions: existence and uniqueness

Let a, b be real-valued Borel measurable functionals defined on $\mathbf{C}([-r, 0])$, where $\mathbf{C}([-r, 0])$ is equipped with the supremum norm $\|.\|_\infty$, which makes it a Banach space. For our purposes it is sufficient to consider autonomous stochastic delay differential equations of the following form:

$$dX(t) \;=\; a(X_t)dt + b(X_t)\cdot dW(t), \qquad t > 0, \tag{2.4}$$

where $W(.)$ is a standard one dimensional Wiener process and $X_t := \big[[-r,0] \ni s \mapsto X(t+s)\big]$ the X-segment of delay length r at time $t \geq 0$.

As in the case of stochastic differential equations without memory there are two notions of a solution and two associated definitions of uniqueness (cf. Karatzas and Shreve, 1991: pp. 285-286, 300-301). The definitions given here in a form adapted to our simplified setting can be found in Mohammed (1996: p. 10), Scheutzow (1983: pp. 12-14) or Larssen (2002).[2]

Definition 2.1. Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space equipped with a filtration $(\mathcal{F}_t)_{t\geq 0}$ satisfying the usual conditions and an $(\mathcal{F}_t)$-adapted standard one dimensional Wiener process $(W(t))_{t\geq 0}$. Let ξ be a $\mathbf{C}([-r,0])$-valued $\mathcal{F}_0$-measurable random variable. Then equation (2.4) has a *strong solution* with initial condition ξ at time 0, if a real-valued process $(X(t))_{t\geq -r}$ exists on $(\Omega, \mathcal{F})$ such that

(i) $(X(t))_{t\geq 0}$ is $(\mathcal{F}_t)$-adapted and $X(s)$ is $\mathcal{F}_0$-measurable for every $s \in [-r, 0]$,

(ii) the trajectories $[-r,\infty) \ni t \mapsto X(t) \in \mathbb{R}$ are continuous P-a.s.,

(iii) integrability: $\mathrm{P}\big(\int_0^t (|a(X_s)| + b^2(X_s))ds < \infty\big) = 1$ for all $t > 0$,

(iv) integral version of (2.4): $X(t) \;=\; X(0) + \int_0^t a(X_s)ds + \int_0^t b(X_s)dW(s)$ for all $t > 0$ P-a. s.,

(v) $X_0 = \xi$ P-a. s.

Definition 2.2. Let $(\Omega, \mathcal{F}, \mathrm{P})$ be any probability space that carries a filtration $(\mathcal{F}_t)_{t\geq 0}$ satisfying the usual conditions and an $(\mathcal{F}_t)$-adapted standard one dimensional Wiener process $(W(t))_{t\geq 0}$.

A strong solution X of (2.4) with respect to $(W, (\mathcal{F}_t))$ and an $\mathcal{F}_0$-measurable initial condition is *unique*, if any other strong solution $\tilde{X}$ of (2.4) with respect to the same Wiener process, filtration and initial condition is indistinguishable from X, i. e. $\mathrm{P}(X(t) = \tilde{X}(t)\ \forall\, t \geq -r) = 1$.

If all strong solutions of (2.4) are unique (for every choice of the filtered probability space and the adapted Wiener process, and for all initial conditions that allow for a strong solution), then one says that *strong uniqueness* holds for equation (2.4) or, equivalently, for the functionals a, b.

Definition 2.3. Let ν be a probability measure on $\mathcal{B}(\mathbf{C}([-r,0]))$. Then equation (2.4) has a *weak solution* with initial distribution ν at time 0, if a triple $((X,W), (\Omega, \mathcal{F}, \mathrm{P}), (\mathcal{F}_t))$ exists such that

(i) $(\Omega, \mathcal{F}, \mathrm{P})$ is a probability space equipped with a filtration $(\mathcal{F}_t)_{t\geq 0}$ satisfying the usual conditions,

(ii) $W = (W(t))_{t\geq 0}$ is an $(\mathcal{F}_t)$-adapted standard one dimensional Wiener process,

(iii) $X = (X(t))_{t\geq -r}$ is a P-a. s. pathwise continuous real-valued process on $(\Omega, \mathcal{F})$ such that $(X(t))_{t\geq 0}$ is $(\mathcal{F}_t)$-adapted and $X(s)$ is $\mathcal{F}_0$-measurable for every $s \in [-r, 0]$,

(iv) integrability: $\mathrm{P}\big(\int_0^t (|a(X_s)| + b^2(X_s))ds < \infty\big) = 1$ for all $t > 0$,

(v) integral version of (2.4): $X(t) \;=\; X(0) + \int_0^t a(X_s)ds + \int_0^t b(X_s)dW(s)$ for all $t > 0$ P-a. s.,

(vi) X_0 has distribution ν with respect to P.

[2] Mohammed (1984, 1996) considers only strong solutions.

Definition 2.4. A weak solution $((X,W),(\Omega,\mathcal{F},\mathrm{P}),(\mathcal{F}_t))$ of (2.4) with initial distribution ν is *unique*, if for any other weak solution $((\tilde{X},\tilde{W}),(\tilde{\Omega},\tilde{\mathcal{F}},\tilde{\mathrm{P}}),(\tilde{\mathcal{F}}_t))$ of (2.4) with the same initial distribution ν it holds that X has the same distribution with respect to P as $\tilde{X}$ with respect to $\tilde{\mathrm{P}}$.

If all weak solutions of (2.4) are unique, then it is said that *weak uniqueness* holds for equation (2.4) or, equivalently, for the functionals a, b.

There is a second notion of uniqueness for weak solutions, analogous to the one of definition 2.2.

Definition 2.5. A weak solution $((X,W),(\Omega,\mathcal{F},\mathrm{P}),(\mathcal{F}_t))$ of (2.4) is *pathwise unique*, if the existence of another weak solution $((\tilde{X},W),(\Omega,\mathcal{F},\mathrm{P}),(\mathcal{F}_t))$ of (2.4) such that $X_0=\tilde{X}_0$ P-a. s. implies that $\mathrm{P}(X(t)=\tilde{X}(t)\ \forall\, t\geq -r)=1$.

If all weak solutions of (2.4) are pathwise unique, then we say that *pathwise uniqueness* holds for equation (2.4) or, equivalently, for the functionals a, b.

The following result is a reduced one dimensional version of theorem I.2 in Mohammed (1996: p. 12).

Theorem 2.1 (Mohammed)**.** *Let $(W(t))_{t\geq 0}$ be an adapted standard one dimensional Wiener process on a probability space $(\Omega,\mathcal{F},\mathrm{P})$ which carries a filtration $(\mathcal{F}_t)_{t\geq 0}$ satisfying the usual conditions. Suppose that the functionals a, b are locally Lipschitz and satisfy a linear growth condition, more precisely:*

(locLip) *For every $n\in\mathbb{N}$ there is a constant $C_n>0$ such that for all $f,g\in\mathbf{C}([-r,0])$ with $\|f\|_\infty,\|g\|_\infty\leq C_n$:* $\quad |a(f)-a(g)|+|b(f)-b(g)|\ \leq\ C_n\cdot\|f-g\|_\infty,$

(Growth) *There is a constant $\tilde{C}>0$ such that for all $f\in\mathbf{C}([-r,0])$:* $|a(f)|+|b(f)|\ \leq\ \tilde{C}\cdot(1+\|f\|_\infty).$

If ξ is a $\mathbf{C}([-r,0])$-valued random variable such that $\mathrm{E_P}(\|\xi\|_\infty^2)<\infty$, then there exists a unique strong solution $(X(t))_{t\geq -r}$ of (2.4) with initial condition ξ. In this case, for all $T>0$ it holds that

$$\mathrm{E_P}\Big(\big(\sup_{t\in[-r,T]}|X(t)|\big)^2\Big)<\infty,$$

$$\mathrm{E_P}\big(\|X_t\|_\infty^2\big)\ \leq\ \hat{C}\cdot\big(1+\mathrm{E_P}\big(\|\xi\|_\infty^2\big)\big)\quad \textit{for all } t\in[0,T] \textit{ and some constant } \hat{C}>0.$$

The existence of a strong solution obviously implies the existence of a weak solution. As for weak uniqueness, this is not an immediate consequence of strong existence and uniqueness. In the case of ordinary stochastic differential equations a result of Yamada and Watanabe states that pathwise uniqueness implies weak uniqueness. A proof thereof is given in Karatzas and Shreve (1991: pp. 308-310). That proof carries over to the more general case of stochastic delay differential equations, see Larssen (2002).

Theorem 2.2 (Yamada, Watanabe et al.)**.** *Pathwise uniqueness of* (2.4) *implies weak uniqueness.*

As any weak solution is also a strong solution with respect to the probability space, filtration and Wiener process it comes with, strong uniqueness of equation (2.4) implies pathwise uniqueness, which in turn implies weak uniqueness.

2.2.2 Markov property

With $r = 0$ equation (2.4) reduces to an ordinary stochastic differential equation. In this case, if weak uniqueness holds for all point measures on $\mathcal{B}(\mathbb{R})$, i. e. all deterministic initial conditions, and if the coefficients a, b are locally bounded and measurable, then the solution process $(X(t))_{t\geq 0}$ (on the canonical space) enjoys the strong Markov property with respect to the Brownian standard filtration (cf. Karatzas and Shreve, 1991: 319-322).

An analogous result in the situation of non-zero delay is the Markov property of the $\mathbf{C}([-r,0])$-valued segment process $(X_t)_{t\geq 0}$, while the solution process $(X(t))_{t\geq -r}$, in general, lacks such property. In the special case, where the diffusion functional b is constant, a proof of the strong Markov property of the segment process under assumptions similar to the situation of zero delay can be found in Scheutzow (1983: pp. 31-32). Here, we cite a reduced one dimensional version of theorem II.1 from Mohammed (1996: p. 17), which is more in bearing with the hypotheses of theorem 2.1.

Theorem 2.3 (Mohammed). *Let $(W(t))_{t\geq 0}$ be a standard one dimensional Wiener process on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$, and let $(\mathcal{F}_t)_{t\geq 0}$ be the accompanying complete Brownian filtration. Suppose that the functionals a, b are globally Lipschitz, that is:*

(globLip) *There is a constant $\hat{C} > 0$ such that for all $f, g \in \mathbf{C}([-r,0])$:*
$$|a(f) - a(g)| + |b(f) - b(g)| \;\leq\; \hat{C} \cdot \|f - g\|_\infty .$$

Let $f \in \mathbf{C}([-r,0])$ and denote by $(X^f(t))_{t\geq -r}$ the strong solution of (2.4) with initial condition f. Define

$$p(f,t,B) \;:=\; \mathrm{P}(X_t^f \in B), \qquad f \in \mathbf{C}([-r,0]), \quad t \geq 0, \quad B \in \mathcal{B}\big(\mathbf{C}([-r,0])\big).$$

Then for every $f \in \mathbf{C}([-r,0])$ the process $(X_t^f)_{t\geq 0}$ is a $\mathbf{C}([-r,0])$-valued Markov process with respect to $(\mathcal{F}_t)_{t\geq 0}$, and p is a time-homogeneous transition function for (X_t^f).

As for ordinary stochastic differential equations, one can associate with the transition function p or, equivalently, the Markov family $\{(X_t^f)_{t\geq 0} \mid f \in \mathbf{C}([-r,0])\}$ a contraction semigroup. To this purpose, let $\mathbf{C}_b := \mathbf{C}_b(\mathbf{C}([-r,0]))$ be the space of all bounded and continuous real-valued functions on $\mathbf{C}([-r,0])$, and equip $\mathbf{C}_b$ with the supremum norm. For $t \geq 0$ define the linear operator $T_t : \mathbf{C}_b \to \mathbf{C}_b$ by setting

$$T_t(\phi)(f) \;:=\; \mathrm{E}_{\mathrm{P}}\big(\phi(X_t^f)\big), \qquad \phi \in \mathbf{C}_b, \quad f \in \mathbf{C}([-r,0]).$$

The first part of theorem II.2 in Mohammed (1996: p. 18) states that $(T_t)_{t\geq 0}$ is, indeed, a one-parameter contraction semigroup on $\mathbf{C}_b$. At this point, there arises a difficulty peculiar to delay equations: If $r > 0$, then the semigroup $(T_t)_{t\geq 0}$ "is never strongly continuous on $\mathbf{C}_b$ under the sup norm" (Mohammed, 1996: 18), that is, for any family of solutions to (2.4), there exists $\phi \in \mathbf{C}_b$ such that $\|T_t(\phi) - \phi\| \not\to 0$ as $t > 0$ tends to zero.

The reason for this lies in the fact that not even the shift semigroup on $\mathbf{C}_b$ is strongly continuous on the whole space (Mohammed, 1996: II.2 (iii)), where the shift or "static" contraction semigroup is the one associated with the family of solutions to the trivial equation $dX^f(t) = 0$, $t \geq 0$, with initial condition $f \in \mathbf{C}([-r,0])$. While $X^f(t) = X^f(0)$ for all $t > 0$, the segment X_t^f, in general, is not equal to $X_0^f = f$. We shall encounter a similar problem in chapter 4, where a delay process with only two possible current states will be constructed.

There is a way out, however. If one considers "weak" convergence on $\mathbf{C}_b$ with respect to the set of all finite regular Borel measures, it turns out that (T_t) is weakly continuous (Mohammed, 1996: II.1 (ii)). The semigroup can then be characterized by means of a weak infinitesimal generator in the sense of Dynkin

(1965: p. 37). Such a characterization was developed in Mohammed (1984), where a formula for the weak infinitesimal generator in terms of the coefficients of equation (2.4) is given.[3]

2.2.3 Stationarity

In this section we will see that under assumptions (2.2) and (2.3) the reference equation (2.1) possesses an invariant probability measure for all choices of positive σ and β greater than some negative constant. This means there is a probability measure ν on $\mathcal{B}(\mathrm{C}([-r,0]))$ such that any weak solution of (2.1) with initial distribution ν is stationary.[4] Here, stationarity of a process is defined as follows (cf. Scheutzow, 1983: 23).

Definition 2.6. Let I be one of the sets $\mathbb{Z}$, $\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{R}$ or $\mathbb{R}^+$. Let $X = (X(t))_{t\in\mathrm{I}}$ be a stochastic process on the probability space $(\Omega, \mathcal{F}, \mathrm{P})$ with values in a measurable space $(\mathrm{S}, \mathcal{S})$. For $h \in \mathrm{I}$ define the shifted process $X^{(h)}$ by setting $X^{(h)}(t) := X(t+h)$, $t \in \mathrm{I}$. Then X is called *stationary* if for every $h \in \mathrm{I}$ the processes X and $X^{(h)}$ have the same distribution.

Stationarity is studied in Scheutzow (1983, 1984) for equations of a special form. Let $W(.)$ be a standard one dimensional Wiener process and F a real-valued Borel measurable and bounded functional on $\mathbf{C}([-1,0])$. Consider the SDDE

$$d\tilde{X}(t) = F(\tilde{X}_t)dt + dW(t), \qquad t \geq 0. \tag{2.5}$$

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 (2.5). Let us specify F as

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

Since the coordinate projections are measurable and V', U' are bounded continuous functions because of (2.2a) and (2.3), F is Borel measurable and bounded.

Let $\tilde{f} \in \mathbf{C}([-1,0])$ 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 of (2.5) with F as just defined and $\tilde{X}_0 = \tilde{f}$ $\tilde{\mathrm{P}}$-almost surely. Set $f(t) := \sigma\sqrt{r} \cdot \tilde{f}(\frac{t}{r})$, $t \in [-r, 0]$. 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 (2.1) 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. Scheutzow, 1983: 31). Therefore, if $\sigma > 0$, then our reference equation can be transformed into an instance of equation (2.5).

Theorem 2.4 cites theorem I.7 from Scheutzow (1983: pp. 17-18) and part of theorem 3 and theorem 5 from Scheutzow (1984: pp. 47-48, 55-56).

Theorem 2.4 (Scheutzow). *Let the functional F be bounded and measurable as above, let ν be a probability measure on $\mathcal{B}(\mathbf{C}([-r,0]))$. Denote by $\underline{f}$ the minimum and by $\overline{f}$ the maximum of $f \in \mathbf{C}([-r,0])$.*

1. *Equation (2.5) has a unique weak solution with initial distribution ν.*
2. *Let $((X,W),(\Omega,\mathcal{F},\mathrm{P}),(\mathcal{F}_t))$ be a weak solution of (2.5) with initial distribution ν. Then there is at most one invariant probability measure π for equation (2.5), and if it exists then*

$$\mathrm{P}_{X_t} \overset{t\to\infty}{\to} \pi \quad \textit{in total variation.}$$

[3] See Mohammed (1984: pp. 70-111) and Mohammed (1996: pp. 20-25).

[4] In this case, a strong solution with initial condition having distribution ν would also be stationary. For our purposes the concept of weak solution is more convenient, because the definitions of resonance we will give in section 3.4 for the reduced model depend only on the distribution of the underlying process.

3. *If* $F(f)\cdot\overline{f}\to-\infty$ *as* $\underline{f}\to\infty$ *and* $F(f)\cdot\underline{f}\to-\infty$ *as* $\overline{f}\to-\infty$ *for all* $f\in\mathbf{C}([-r,0])$*, then* (2.5) *possesses an invariant probability measure.*

To conclude we apply the results we have seen so far to our reference model.

Proposition 2.1. *Suppose that* V*,* U *satisfy conditions* (2.2) *and* (2.3)*. Let* $\sigma\geq 0$*,* $\beta\in\mathbb{R}$ *be given. Then the following holds for equation* (2.1)*, which describes the reference model:*

1. *Strong and weak solutions exist for every probability measure on* $\mathcal{B}(\mathbf{C}([-r,0]))$ *as initial distribution, and the solutions are unique in the respective sense.*
2. *The solution processes enjoy the strong Markov property.*
3. *If* $\sigma>0$ *and* $\beta>-\frac{V'(R_{pot})}{U'(R_{pot})}$*, then a unique invariant probability measure exists.*

Proof. Theorems 2.1 and 2.2 guarantee weak and strong existence and uniqueness of solutions to equation (2.1) for all parameter choices and all deterministic initial conditions. An argument as in Scheutzow (1983: pp. 17-18) shows that in place of point distributions one may take probability measures as initial conditions.

The Markov property is a consequence of theorem 2.3. Since weak uniqueness holds, theorem 1 from Scheutzow (1984: p. 42) implies that solutions to (2.1) have the Markov property even with respect to stopping times.

The existence of an invariant probability measure is an application of the third part of theorem 2.4 and condition (2.3) on the growth of the potentials V, U. □

In the situation of proposition 2.1, if $\sigma>0$ and $\beta<-\frac{V'(R_{pot})}{U'(R_{pot})}$, then there is no invariant probability measure for equation (2.1). To see this, let $((X,W),(\Omega,\mathcal{F},\mathrm{P}),(\mathcal{F}_t))$ be a weak solution of (2.1) with deterministic initial condition $f\in\mathbf{C}([-r,0])$ such that $\overline{f}<-R_{pot}$. Set $\mu:=V'(R_{pot})+\beta\cdot U'(R_{pot})$, then $\mu<0$. Define $A:=\{X(t)<-R_{pot}-\frac{|\mu|}{2}t\ \forall t\geq 0\}$. The probability of the event A with respect to P follows from the distribution of the supremum of a Brownian motion with negative drift, that is

$$\mathrm{P}(A)=1-\mathrm{P}\left(\sup_{t\geq 0} f(0)-\tfrac{|\mu|}{2}t+\sigma\cdot W(t)\geq -R_{pot}\right)=1-\exp\left(-\tfrac{|\mu|(|f(0)|-R_{pot})}{\sigma^2}\right)>0.$$

If equation (2.1) had an invariant probability measure, then the convergence part of theorem 2.4 would imply weak convergence of the distribution of $X(t)$ to some probability measure on $\mathcal{B}(\mathbb{R})$ as $t\to\infty$. But $\mathrm{P}(A)>0$, a contradiction to the compact exhaustability of probability measures on Polish spaces.

2.3 Heuristical description and numerical simulation

Let us have a look at basic parameter settings for equation (2.1). Suppose that conditions (2.2) and (2.3) are satisfied. To illustrate the discussion we made some numerical simulation of our reference equation, where we chose as double well potential V the standard quartic potential and as delay potential U a parabola (see figure 2.1). Solutions to equation (2.1) were obtained by approximation with time series, using appropriately scaled i. i. d. Bernoulli trials as approximation of the noise process. Results by Scheutzow (1983, 1984) and Lorenz (2003) guarantee weak convergence of such schemes.

The simplest and least interesting choice of parameters is $\sigma=0$ and $\beta=0$, i. e. no noise and no delay. In this case, (2.1) 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. (2002) 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 $\beta \geq 1$, $r \in [0, 1]$.[5] This is not the parameter region we are interested in, here. Recall from 1.1 that stochastic resonance is a phenomon 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 time enough 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 (2.1) exhibit periodic behaviour even for $\beta > 0$ comparatively small. This is illustrated in figures 2.2 and 2.3. With $\beta = 0.39$, the solution periodically switches between the two stable equilibria, while with $\beta = 0.38$ it converges to the stable equilibrium near which it was started. The switching behaviour does not depend on the special choice of the delay time $r > 0$. Clearly, the behaviour of a deterministic solution may crucially depend on the choice of the initial condition, while for $\sigma > 0$ all solutions converge towards the stationary regime.[6]

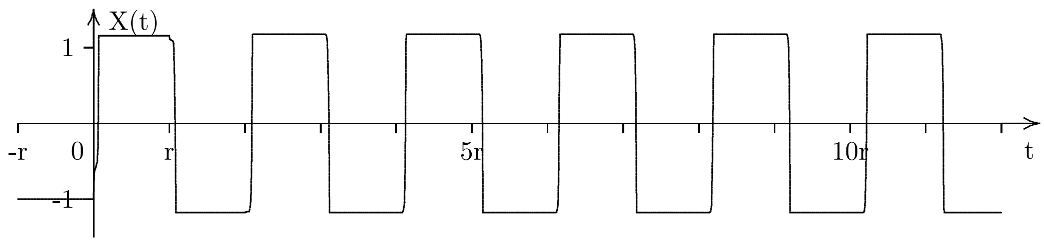

Figure 2.2: Deterministic solution, i. e. $\sigma = 0$, of (2.1) with $r = 500$, $\beta = 0.39$ and initial condition -1 on $[-r, 0]$.

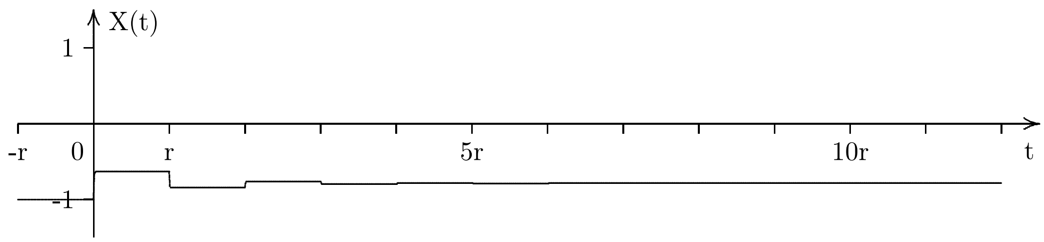

Figure 2.3: Deterministic solution, i. e. $\sigma = 0$, of (2.1) with $r = 500$, $\beta = 0.38$ and initial condition -1 on $[-r, 0]$.

If $\beta = 0$ and $\sigma > 0$, then our SDDE (2.1) reduces to the SDE (1.2), cf. also section 5.1. 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

[5]Our parameter β corresponds to $-\alpha$ in equation (1.3) of Redmonda et al. (2002).

[6]Unfortunately, we do not know of estimates on the rate of convergence towards stochastic equilibrium. In the simulations, we chose a "pre-period" of 100 times the length of the delay ($1000r$ in case $\beta < 0$) in order to get solutions approaching the stationary regime (distance in total variation with respect to the segment process). The rate of convergence depends, of course, on the value of $\sigma > 0$.

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, see figure 2.4.

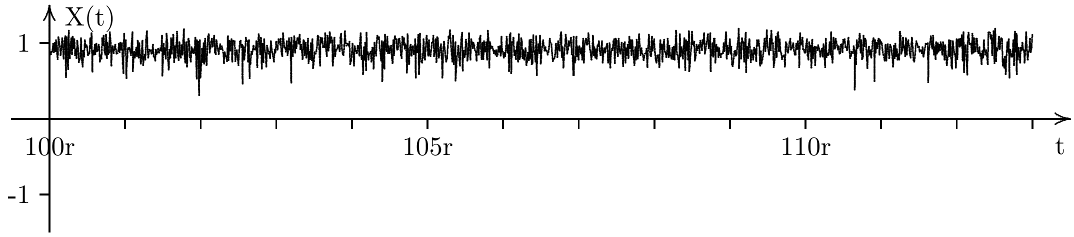

Figure 2.4: Typical path of a solution to (2.1) with $\sigma = 0.2$, $\beta = 0.1$ and $r = 500$ from time $100r$ to $113r$. Noise and delay are too weak to induce frequent interwell transitions.

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, see figure 2.5. 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 behaviour at a non-zero noise level, see figure 2.6. This is what we may call an instance of stochastic resonance.

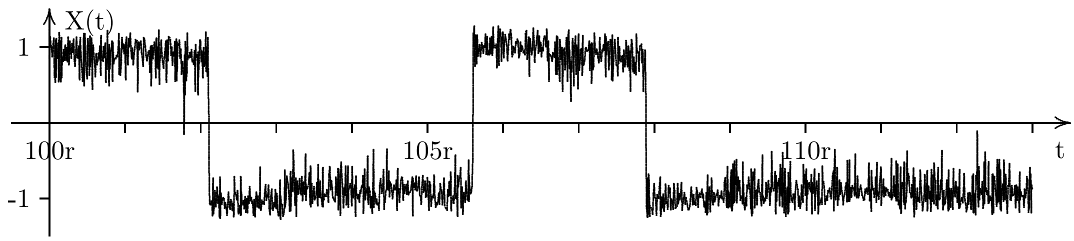

Figure 2.5: Typical path of a solution to (2.1) with $\sigma = 0.25$, $\beta = 0.1$ and $r = 500$ from time $100r$ to $113r$. Interwell transitions become more frequent.

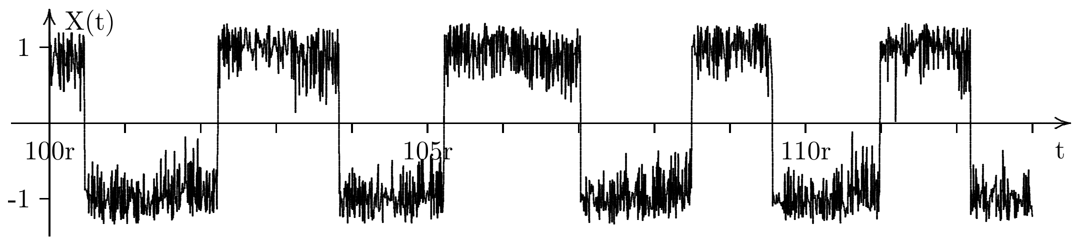

Figure 2.6: Typical path of a solution to (2.1) with $\sigma = 0.30$, $\beta = 0.1$ and $r = 500$ from time $100r$ to $113r$. Interwell transitions at intervals between r and $2r$, quasi-periodic switching behaviour: Stochastic resonance!

Further increasing the noise intensity leads to ever growing intrawell fluctuations which eventually destroy the quasi-periodicity of the interwell transitions, see figures 2.7 and 2.8. 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, see figure 2.9.

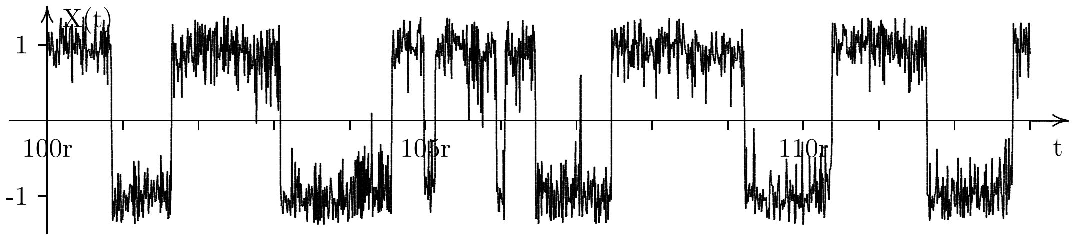

Figure 2.7: Typical path of a solution to (2.1) with $\sigma = 0.35$, $\beta = 0.1$ and $r = 500$ from time $100r$ to $113r$. Large random fluctuations begin to destroy the quasi-periodic switching behaviour.

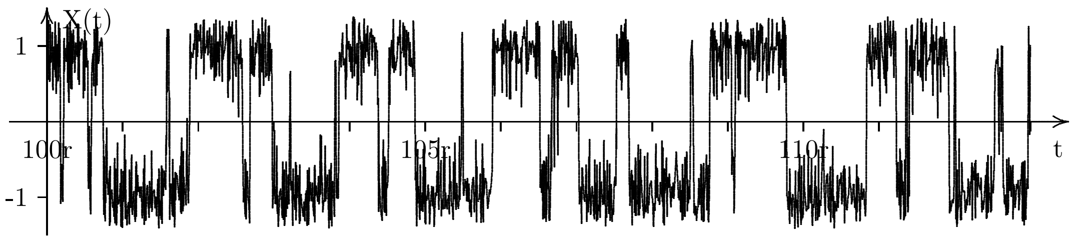

Figure 2.8: Typical path of a solution to (2.1) with $\sigma = 0.40$, $\beta = 0.1$ and $r = 500$ from time $100r$ to $113r$. Large random fluctuations destroy the quasi-periodic switching behaviour.

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 behaviour 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. Typical solution trajectories for $\sigma = 0.35$ and $\sigma = 0.40$ are depicted in figures 2.10 and 2.11. Notice that we have chosen a longer pre-run period than with $\beta > 0$, namely $1000r$ instead of $100r$, in order to let the system come closer to the stationary regime.

In chapter 5 we will state more precisely what regular transition behaviour means in case $\beta < 0$, yet we will not subsume it under the heading of stochastic resonance.

2.4 Classical stochastic resonance versus resonance with delay

Here, we compare some of the main features of equation (1.1), which is a prototypical example of a model that shows stochastic resonance, with those of equation (2.1), which describes our reference model. Let us suppose the symmetric double well potential V is the same in both cases.

If we have $a = 0$ and $\beta = 0$, then both equations reduce to the same SDE, namely equation (1.2). Recall from section 1.1 that in the presence of weak noise there are two distinct time scales, one corresponding to the quadratic variation of the Wiener process, the other proportional to the average time it takes the diffusion to travel from one of the two metastable states to the other.

Now, let $a > 0$, $\beta \neq 0$ be of small absolute value. Then a third time scale enters the scene, namely the period T of the harmonic signal in case of equation (1.1), the length of the delay $r > 0$ in case of equation (2.1). Notice that the delay length r should be equated with the *half* period $\frac{T}{2}$ of the periodic signal.

What is more than a new time scale is the fact that the corresponding periodic or delay force alters

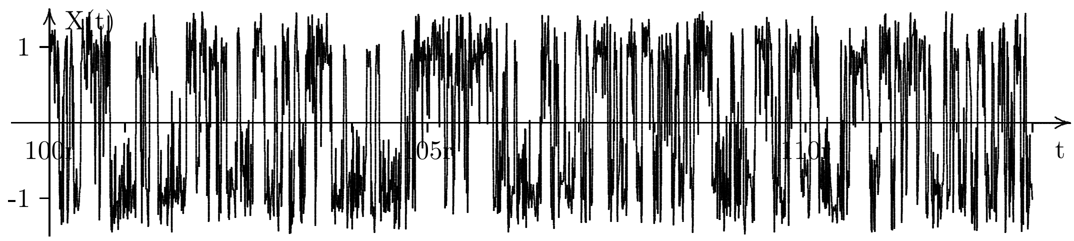

Figure 2.9: Typical path of a solution to (2.1) with $\sigma = 0.50$, $\beta = 0.1$ and $r = 500$ from time $100r$ to $113r$. Large random fluctuations, noise is too strong.

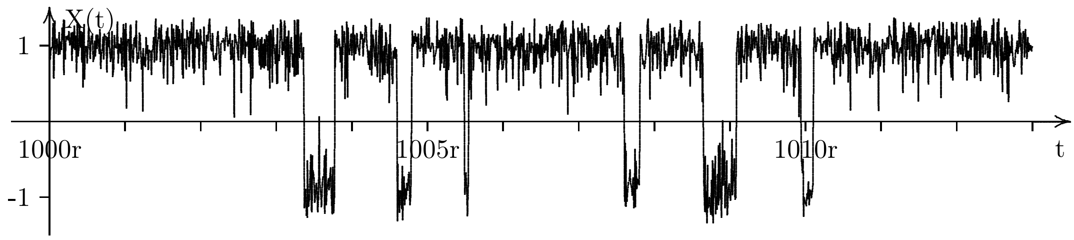

Figure 2.10: Typical path of a solution to (2.1) with $\sigma = 0.35$, $\beta = -0.1$ and $r = 500$ from time $1000r$ to $1013r$. Noise is comparatively strong, delay force prolongs long sojourns and favours short switching transitions.

the symmetric potential landscape creating a shallow and a deep potential well, thereby splitting up the time scale induced by the interwell transitions into two: one corresponding to transitions from the shallow to the deep well, the other corresponding to transitions in the opposite direction. By tuning the noise intensity σ to the periodic or delay force the behaviour of the solution trajectories can now be changed in such a way as to yield an optimal macroscopic response to the modulating force at an optimal non-zero noise level.

What an "optimal macroscopic response" and the corresponding "optimal noise level" is depends, of course, on the measure of resonance. Natural measures in the context of equation (1.1) seem to be those based on the frequency spectrum of the output signal in view of the fact that there is an harmonic input signal. The main difficulty with these measures is that they quantify not only the macroscopic, but also the microscopic behaviour of the solution process and it is often unclear how to tell their respective contributions apart; cf. sections 1.2 and 1.3.

In equation (2.1), there is nothing that could be regarded as an harmonic input signal. Natural measures in this context are those based on the interwell transition time or residence time distribution. In section 3.4 we will define two such measures for the reduced model.

Notwithstanding the similarities sketched above there are important differences between the model for classical stochastic resonance and the delay model. The harmonic signal in equation (1.1) can be regarded as an external force, because it is – as a deterministic function – independent of the filtration generated by the solution process. The delay term in equation (2.1), on the contrary, depends on the solution process itself and is measurable with respect to $(\mathcal{F}^X_{.-r})$. Clearly, the harmonic signal periodically alters the potential landscape, no matter in what way the solution process evolves, while the quasi-periodicity of the delay force must be derived from the solution process itself.

At this point we should mention a difference between the two models that can be overcome. The harmonic signal of equation (1.1) causes the potential landscape to change smoothly. The tilting of the double well potential V due to the delay force, though being continuous, is less regular, because it depends on the point value of a trajectory of usually unbounded variation. In addition, interwell transitions take

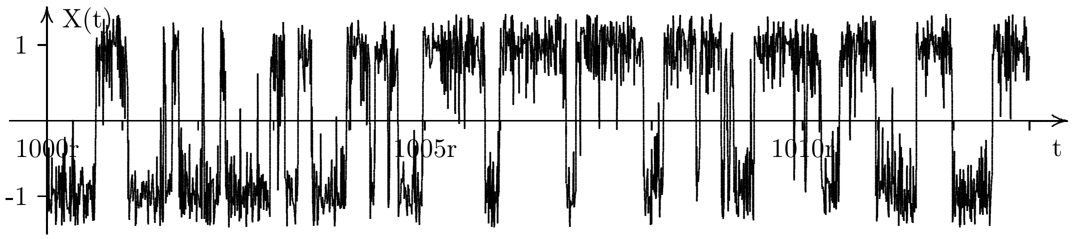

Figure 2.11: Typical path of a solution to (2.1) with $\sigma = 0.40$, $\beta = -0.1$ and $r = 500$ from time $1000r$ to $1013r$. Large random fluctuations, short intrawell sojourns.

place almost instantaneously, triggering almost instantaneous changes of sign in the delay force. This effect could partially be transferred to the classical model by replacing the sinus function in (1.1) with an appropriate periodic step function. The other way round, the two models can be brought into line with one another by introducing a distributed delay in equation (2.1) instead of a point delay.

A more formal distinction regards the Markov property. Equation (1.1) is a non-autonomous SDE, equation (2.1) an autonomous SDDE. Solutions are real-valued stochastic processes indexed by $t \in [0, \infty)$ in the first case and $t \in [-r, \infty)$ in the second. As such they are non-Markovian. In order to recover the Markov property one has to enlarge the state space. In case of equation (1.1) this leads to processes with values in $\mathbb{R} \times S^1$, while solutions of (2.1) are $\mathbf{C}([-r, 0])$-valued Markov processes.

Last but not least, while it makes little difference whether the amplitude a in equation (1.1) is positive or allowed to be negative, there is an important difference in the transition behaviour of solutions to (2.1) depending on the sign of the delay parameter β.

Chapter 3

The two state model in discrete time

Applying the ideas presented in section 1.3, we will develop a reduced model with the aim of capturing the effective dynamics of the reference model from chapter 2. To simplify things further we start with discrete time. As the segment process associated with the unique solution to (2.1), the reference model equation, enjoys the strong Markov property, it is reasonable to approximate the transition behaviour 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.

In section 3.1 the approximating Markov chains are defined, while in section 3.2 an explicit formula for their stationary distributions is obtained. In section 3.3 we make use of this formula in order to calculate, for each $M \in \mathbb{N}$, the residence time distribution in the stationary regime and then derive its density function in the limit of discrete time tending to continuous time. Finally, in section 3.4, 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 (1999), which will be our standard reference.

3.1 A sequence of Markov chains

Let $M \in \mathbb{N}$ be the discretisation 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 often be written as $(M+1)$-tuples having $\{-1, 1\}$-valued entries indexed (from left to right) from $-M$ to 0. The strange choice of the index range serves 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. This is, of course, only a mnemonic, the link to the reference model has to be established in a different way.

To this purpose, let α, γ be positive real numbers. If $X(.)$ is the unique solution to (2.1) in the case of "interesting" noise parameter σ and delay parameter β then 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. In section 5.2 we derive an approximation formula for the transition rates and discuss the assumptions underlying their introduction.

In the discrete model of discretisation 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 discretisation of a Markov process we set

$$R_{sc} : (0,\infty) \times \mathbb{N} \ni (\eta, N) \mapsto \tfrac{\eta}{\alpha+\gamma} \cdot (1 - e^{-\frac{\alpha+\gamma}{N}}) \in (0,1). \tag{3.1}$$

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

$$\forall\, l \in \{-M, \ldots, -1\} : \quad \tilde{Z}^{(l)} = Z^{(l+1)}. \tag{3.2}$$

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 $(\mathbf{1}, \mathbf{-1}, -1)$. ◊

If (3.2) 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 (3.2) we must have

$$\begin{aligned}
Z^{(0)} = Z^{(-M)} \quad &\Rightarrow \quad p^M_{Z\tilde{Z}} = \begin{cases} \alpha_M & \text{if} \quad \tilde{Z}^{(0)} \neq Z^{(0)}, \\ 1-\alpha_M & \text{if} \quad \tilde{Z}^{(0)} = Z^{(0)}, \end{cases} \\
Z^{(0)} \neq Z^{(-M)} \quad &\Rightarrow \quad p^M_{Z\tilde{Z}} = \begin{cases} \gamma_M & \text{if} \quad \tilde{Z}^{(0)} \neq Z^{(0)}, \\ 1-\gamma_M & \text{if} \quad \tilde{Z}^{(0)} = Z^{(0)}. \end{cases}
\end{aligned} \tag{3.3}$$

The fact that – because of (3.1) – we always have $\alpha_M, \gamma_M \in (0,1)$, implies

$$p^M_{Z\tilde{Z}} \neq 0 \quad \Leftrightarrow \quad \text{shift condition (3.2) is satisfied.} \tag{3.4}$$

Now define $\mathbf{P}_M := (p^M_{Z\tilde{Z}})_{Z,\tilde{Z}\in \mathrm{S}_M}$. Clearly, $\mathbf{P}_M$ is a $2^{M+1} \times 2^{M+1}$ transition matrix. Interpreting the elements of S_M as binary numbers, where -1 has to be read as 0, $\mathbf{P}_M$ is of the form (only non-zero entries are indicated)

$$\begin{pmatrix} \mathbf{C}_M & & \\ & \ddots & \\ & & \mathbf{C}_M \\ \tilde{\mathbf{C}}_M & & \\ & \ddots & \\ & & \tilde{\mathbf{C}}_M \end{pmatrix}, \tag{3.5}$$

where $\mathbf{C}_M \equiv \left(\begin{smallmatrix} 1-\alpha_M & \alpha_M & 0 & 0 \\ 0 & 0 & \gamma_M & 1-\gamma_M \end{smallmatrix}\right)$ and $\tilde{\mathbf{C}}_M \equiv \left(\begin{smallmatrix} 1-\gamma_M & \gamma_M & 0 & 0 \\ 0 & 0 & \alpha_M & 1-\alpha_M \end{smallmatrix}\right)$.

Example. The following table illustrates the transition matrix $\mathbf{P}_M$ for $M = 2$.

	$\begin{pmatrix}-1\\-1\\-1\end{pmatrix}$	$\begin{pmatrix}-1\\-1\\1\end{pmatrix}$	$\begin{pmatrix}-1\\1\\-1\end{pmatrix}$	$\begin{pmatrix}-1\\1\\1\end{pmatrix}$	$\begin{pmatrix}1\\-1\\-1\end{pmatrix}$	$\begin{pmatrix}1\\-1\\1\end{pmatrix}$	$\begin{pmatrix}1\\1\\-1\end{pmatrix}$	$\begin{pmatrix}1\\1\\1\end{pmatrix}$
$(-1,-1,-1)^\mathsf{T}$	$1-\alpha_2$	α_2	0	0	0	0	0	0
$(-1,-1,\ 1)^\mathsf{T}$	0	0	γ_2	$1-\gamma_2$	0	0	0	0
$(-1,\ 1,-1)^\mathsf{T}$	0	0	0	0	$1-\alpha_2$	α_2	0	0
$(-1,\ 1,\ 1)^\mathsf{T}$	0	0	0	0	0	0	γ_2	$1-\gamma_2$
$(\ 1,-1,-1)^\mathsf{T}$	$1-\gamma_2$	γ_2	0	0	0	0	0	0
$(\ 1,-1,\ 1)^\mathsf{T}$	0	0	α_2	$1-\alpha_2$	0	0	0	0
$(\ 1,\ 1,-1)^\mathsf{T}$	0	0	0	0	$1-\gamma_2$	γ_2	0	0
$(\ 1,\ 1,\ 1)^\mathsf{T}$	0	0	0	0	0	0	α_2	$1-\alpha_2$

◊

For every $M \in \mathbb{N}$ we may choose an S_M-valued process $(X_n^M)_{n\in\mathbb{N}_0}$ on some measurable space $(\Omega_M, \mathcal{F}_M)$ and probability measures P_Z^M, $Z \in \mathrm{S}_M$, on $\mathcal{F}_M$ such that under P_Z^M the discrete process X^M is a homogeneous Markov chain with transition matrix $\mathbf{P}_M$ and initial condition $\mathrm{P}_Z^M(X_0^M = Z) = 1$.

If ν is a probability measure on the power set $\wp(\mathrm{S}_M)$ then, as usual, P_ν^M will 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_ν^M. Since S_M is finite, we have

$$\mathrm{P}_\nu^M = \sum_{Z\in\mathrm{S}_M} \nu(Z)\, \mathrm{P}_Z^M .$$

When there is no ambiguity about the probability measure P^M, we will write P_ν instead of P_ν^M. For the probability of a singleton $\{Z\}$ under a discrete measure ν let us just write $\nu(Z)$.

3.2 Stationary distributions

From relation (3.4), 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 $(-1,\ldots,-1)$, for example, has positive probability for any finite number of steps.

Now, if the transition matrix of a homogeneous Markov chain is finite-dimensional or, what is the same, the state space 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 (1999: 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

$$\pi_M^\mathsf{T} = \pi_M^\mathsf{T}\mathbf{P}_M, \quad \text{that is}$$

$$\pi_M(\tilde{Z}) = \sum_{Z\in\mathrm{S}_M} \pi_M(Z)\, p_{Z\tilde{Z}}^M \qquad \text{for all } \tilde{Z}\in\mathrm{S}_M. \tag{3.6}$$

Note that X^M is a stationary process on $(\Omega_M, \mathcal{F}_M, \mathrm{P}_{\pi_M})$ in the sense of definition 2.6, where the index set I is chosen to be $\mathbb{N}_0$, because $X_{n+1}^M \overset{d}{\sim} X_n^M$ with respect to P_{π_M} for every $n \in \mathbb{N}_0$ and X_{n+1}^M is independent of $\sigma(X_0^M,\ldots,X_{n-1}^M)$ given $\sigma(X_n^M)$, since X^M has the Markov property.

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

$$\mathcal{J}(Z) := \#\left\{ j \in \{-M+1,\ldots,0\} \;\middle|\; Z^{(j)} \neq Z^{(j-1)} \right\}.$$

At the moment, "number of changes of sign" would be a label more precise for $\mathcal{J}(Z)$ than "number of jumps" in view of the fact that we have not yet defined what a jump for Z should be. This will become clear in section 4.2. Let us here make use of the new notation.

Proposition 3.1 (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 $Z \in \mathrm{S}_M$ it holds that*

$$\pi_M(Z) \;=\; \frac{1}{c_M}\,\tilde{\alpha}_M^{\lfloor\frac{\mathcal{J}(Z)+1}{2}\rfloor}\,\tilde{\gamma}_M^{\lfloor\frac{\mathcal{J}(Z)}{2}\rfloor} \;=\; \frac{1}{c_M}\,\tilde{\alpha}_M^{\mathcal{J}(Z) \bmod 2}\,\tilde{\eta}_M^{\lfloor\frac{\mathcal{J}(Z)}{2}\rfloor}, \tag{3.7}$$

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

Proof. The right-hand part of equation (3.7) 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 $\mathcal{J}(Z) \in \{0,\dots,M\}$ for every $Z \in \mathrm{S}_M$, and with $j \in \{0,\dots,M\}$ there are exactly $2\cdot\binom{M}{j}$ elements in S_M having j jumps.

Let $Z = (Z^{(-M)},\dots,Z^{(0)})$ be an element of S_M. Define elements $\tilde{Z}$, $\hat{Z}$ of S_M as

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

Because of (3.3) and (3.4) the global balance equations (3.6) reduce to

$$\pi_M(Z) \;=\; \begin{cases}(1-\alpha_M)\cdot\pi_M(\tilde{Z}) + (1-\gamma_M)\cdot\pi_M(\hat{Z}) & \text{if } Z^{(-1)} = Z^{(0)},\\ \alpha_M\cdot\pi_M(\tilde{Z}) + \gamma_M\cdot\pi_M(\hat{Z}) & \text{if } Z^{(-1)} = -Z^{(0)}.\end{cases} \tag{3.8}$$

Equations (3.8) 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 (3.8). Let Z, $\tilde{Z}$,$\hat{Z}$ be elements of S_M as above. Then

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

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

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

From proposition 3.1 we see that π_M is symmetric in the sense that for all $(Z^{(-M)},\dots,Z^{(0)}) \in \mathrm{S}_M$

$$\pi\big((Z^{(-M)},\dots,Z^{(0)})\big) \;=\; \pi\big((-Z^{(-M)},\dots,-Z^{(0)})\big) \;=\; \pi\big((Z^{(0)},\dots,Z^{(-M)})\big).$$

For later use, it is convenient to take note of the following lemma.

Lemma 3.1. *Let a, b, η be positive real numbers. For $n \in \mathbb{N}_0$ set*

$$a_n = a_n(a,b,\eta) := \frac{1}{2}\left(a + \frac{b}{\sqrt{\eta}}\right)\cdot(1+\sqrt{\eta})^n + \frac{1}{2}\left(a - \frac{b}{\sqrt{\eta}}\right)\cdot(1-\sqrt{\eta})^n,$$

$$b_n = b_n(a,b,\eta) := \frac{1}{2}(b + a\cdot\sqrt{\eta})\cdot(1+\sqrt{\eta})^n + \frac{1}{2}(b - a\cdot\sqrt{\eta})\cdot(1-\sqrt{\eta})^n.$$

Then

$$a_n = \Big(\sum_{k=0}^{\lfloor\frac{n}{2}\rfloor}\binom{n}{2k}\eta^k\Big)\cdot a + \Big(\sum_{k=0}^{\lfloor\frac{n-1}{2}\rfloor}\binom{n}{2k+1}\eta^k\Big)\cdot b, \tag{3.9a}$$

$$b_n = \Big(\sum_{k=0}^{\lfloor\frac{n}{2}\rfloor}\binom{n}{2k}\eta^k\Big)\cdot b + \Big(\sum_{k=0}^{\lfloor\frac{n-1}{2}\rfloor}\binom{n}{2k+1}\eta^{k+1}\Big)\cdot a. \tag{3.9b}$$

Proof. Splitting up the sum in the binomial formula applied to $(1 \pm \sqrt{\eta})^n$ gives

$$(1 \pm \sqrt{\eta})^n = \sum_{k=0}^{\lfloor\frac{n}{2}\rfloor}\binom{n}{2k}\eta^k \pm \Big(\sum_{k=0}^{\lfloor\frac{n-1}{2}\rfloor}\binom{n}{2k+1}\eta^k\Big)\sqrt{\eta}.$$

The assertion follows by linear combination of $(1+\sqrt{\eta})^n$ and $(1-\sqrt{\eta})^n$. □

The sequences (a_n), (b_n) defined in lemma 3.1 above, arise as the result of a Fibonacci-like iteration if we set $a_0 := a$, $b_0 := b$ and recursively define

$$a_n := a_{n-1} + b_{n-1}, \quad b_n := \eta\cdot a_{n-1} + b_{n-1}, \qquad n \in \mathbb{N}.$$

Let c_M be the normalizing constant from proposition 3.1. Then (3.9a) of lemma 3.1 implies

$$c_M = \Big(1 + \sqrt{\tfrac{\tilde{\alpha}_M}{\tilde{\gamma}_M}}\Big)\big(1+\sqrt{\tilde{\eta}_M}\big)^M + \Big(1 - \sqrt{\tfrac{\tilde{\alpha}_M}{\tilde{\gamma}_M}}\Big)\big(1-\sqrt{\tilde{\eta}_M}\big)^M. \tag{3.10}$$

3.3 Residence time distributions

Let $M \in \mathbb{N}\setminus\{1\}$, let the transition probabilities α_M, γ_M, the transition matrix $\mathbf{P}_M$, the associated S_M-valued Markov chain X^M on $(\Omega_M, \mathcal{F}_M)$, the stationary distribution π_M on $\wp(\mathrm{S}_M)$ and the induced probability measure P_{π_M} on $\mathcal{F}_M$ be defined as in sections 3.1 and 3.2 above.

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

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

Now define

$$L_M(k) := \mathrm{P}_{\pi_M}\big(Y_0^M = 1,\dots,Y_{k-1}^M = 1, Y_k^M = -1 \,\big|\, Y_{-1}^M = -1, Y_0^M = 1\big), \quad k \in \mathbb{N}. \tag{3.11}$$

$L_M(k)$ gives the probability to remain exactly k units of time in the same state conditional on the occurrence of a jump. Note that $L_M(.)$ is well defined, because

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

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

Here, $\{(*,\dots,*,-1,1)\}$ denotes the set $\{Z \in \mathrm{S}_M \mid Z^{(-1)} = -1, Z^{(0)} = 1\}$. Because of the symmetry properties of $\mathbf{P}_M$ and π_M the roles of -1 and 1 in (3.11) 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. For any $n \in \mathbb{N}$, we have

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

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.

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

$$K_M := \mathrm{P}_{\pi_M}\big(Y_0^M = -1, Y_1^M = 1,\dots,Y_M^M = 1 \;\big|\; Y_0^M = -1, Y_1^M = 1\big). \tag{3.12}$$

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

$$\begin{aligned} L_M(k) \;&=\; \mathrm{P}_{\pi_M}\big(\{Y_0^M = -1, Y_1^M = 1,\dots,Y_M^M = 1\} \cap \{Y_{M+1}^M = 1,\dots,Y_k^M = 1, Y_{k+1}^M = -1\} \\ &\qquad\qquad \mid \{Y_0^M = -1, Y_1^M = 1\}\big) \\ &=\; (1-\gamma_M)\cdot K_M \cdot \alpha_M \cdot (1-\alpha_M)^{k-M-1}, \qquad k \ge M+1, \end{aligned} \tag{3.13}$$

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

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

From proposition 3.1 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

$$\begin{aligned} \pi_M\big(\{(*,\dots,*,-1,1)\}\big) \;&=\; \frac{1}{c_M}\cdot\sum_{j=0}^{M-2}\binom{M-2}{j}\big(\tilde\alpha_M^{\,j+1 \bmod 2}\,\tilde\eta_M^{\lfloor\frac{j+1}{2}\rfloor} \;+\; \tilde\alpha_M^{\,j+2 \bmod 2}\,\tilde\eta_M^{\lfloor\frac{j+2}{2}\rfloor}\big) \\ &=\; \frac{\tilde\alpha_M}{c_M}\cdot\Bigg(\Big(\sum_{j=0}^{\lfloor\frac{M-2}{2}\rfloor}\binom{M-2}{2j}\tilde\eta_M^{\,j}\Big) \;+\; \Big(\sum_{j=0}^{\lfloor\frac{M-3}{2}\rfloor}\binom{M-2}{2j+1}\tilde\eta_M^{\,j+1}\Big) \\ &\qquad +\;\Big(\sum_{j=0}^{\lfloor\frac{M-2}{2}\rfloor}\binom{M-2}{2j}\tilde\eta_M^{\,j}\Big)\,\tilde\gamma_M \;+\; \Big(\sum_{j=0}^{\lfloor\frac{M-3}{2}\rfloor}\binom{M-2}{2j+1}\tilde\eta_M^{\,j}\Big)\,\tilde\gamma_M\Bigg) \\ &=\; \frac{\tilde\alpha_M}{2c_M}\cdot\Bigg(\Big(1+\sqrt{\tfrac{\tilde\gamma_M}{\tilde\alpha_M}}\Big)\big(1+\sqrt{\tilde\eta_M}\big)^{M-1} + \Big(1-\sqrt{\tfrac{\tilde\gamma_M}{\tilde\alpha_M}}\Big)\big(1-\sqrt{\tilde\eta_M}\big)^{M-1}\Bigg), \end{aligned}$$

where in the last part of the above equation lemma 3.1 has been applied. We therefore have

$$K_M \quad = \quad \frac{2}{\Big(1+\sqrt{\tfrac{\tilde\gamma_M}{\tilde\alpha_M}}\Big)\big(1+\sqrt{\tilde\eta_M}\big)^{M-1} + \Big(1-\sqrt{\tfrac{\tilde\gamma_M}{\tilde\alpha_M}}\Big)\big(1-\sqrt{\tilde\eta_M}\big)^{M-1}}. \tag{3.14}$$

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

$$L_M(M) \quad = \quad \frac{\mathrm{P}_{\pi_M}\big(Y_0^M = -1, Y_1^M = 1,\dots,Y_M^M = 1, Y_{M+1}^M = -1\big)}{\pi_M\big(\{(*,\dots,*,-1,1)\}\big)} \quad = \quad \gamma_M\cdot K_M, \tag{3.15a}$$

$$L_M(M-1) \quad = \quad \frac{\pi_M\big((-1,1,\dots,1,-1)\big)}{\pi_M\big(\{(*,\dots,*,-1,1)\}\big)} \quad = \quad \tilde\gamma_M\cdot K_M. \tag{3.15b}$$

Now, let $k \in \{1,\dots,M-2\}$. Set $n := M-2-k$ and denote by $\{(-1,1,\dots,1,-1,*,\dots,*)\}$ the set $\{Z \in \mathrm{S}_M \mid Z^{(-M)} = -1, Z^{(-M+1)} = 1,\dots,Z^{(-M+k)} = 1, Z^{(-M+k+1)} = -1\}$. Then

$$\begin{aligned}
L_M(k) &= \frac{\pi_M\big(\{(-1,1,\dots,1,-1,*,\dots,*)\}\big)}{\pi_M\big(\{(*,\dots,*,-1,1)\}\big)} = \frac{c_M}{\tilde\alpha_M}\cdot K_M\cdot \pi_M\big(\{(-1,1,\dots,1,-1,*,\dots,*)\}\big)\\
&= \frac{1}{\tilde\alpha_M}\cdot K_M\cdot \sum_{j=0}^{n}\binom{n}{j}\big(\tilde\alpha_M^{\,j+2 \bmod 2}\,\tilde\eta_M^{\lfloor\frac{j+2}{2}\rfloor} + \tilde\alpha_M^{\,j+3 \bmod 2}\,\tilde\eta_M^{\lfloor\frac{j+3}{2}\rfloor}\big)\\
&= \frac{\tilde\eta_M}{\tilde\alpha_M}\cdot K_M\cdot\Bigg(\Big(\sum_{j=0}^{\lfloor\frac{n}{2}\rfloor}\binom{n}{2j}\tilde\eta_M^{\,j}\Big) + \Big(\sum_{j=0}^{\lfloor\frac{n-1}{2}\rfloor}\binom{n}{2j+1}\tilde\eta_M^{\,j+1}\Big)\\
&\qquad + \Big(\sum_{j=0}^{\lfloor\frac{n}{2}\rfloor}\binom{n}{2j}\tilde\eta_M^{\,j}\Big)\tilde\alpha_M + \Big(\sum_{j=0}^{\lfloor\frac{n-1}{2}\rfloor}\binom{n}{2j+1}\tilde\eta_M^{\,j}\Big)\tilde\alpha_M\Bigg)\\
&= \frac{\tilde\gamma_M}{2}\cdot K_M\cdot\Bigg(\Big(1+\sqrt{\tfrac{\tilde\alpha_M}{\tilde\gamma_M}}\Big)\big(1+\sqrt{\tilde\eta_M}\big)^{n+1} + \Big(1-\sqrt{\tfrac{\tilde\alpha_M}{\tilde\gamma_M}}\Big)\big(1-\sqrt{\tilde\eta_M}\big)^{n+1}\Bigg),
\end{aligned}$$

where we have again made use of lemma (3.1). A last rearrangement yields for $k \in \{1,\dots,M-1\}$

$$\begin{aligned}
L_M(k) &= \frac{\sqrt{\tilde\gamma_M}}{2}\cdot K_M\cdot\Big(\sqrt{\tilde\gamma_M}\,\big((1+\sqrt{\tilde\eta_M})^{M-1-k} + (1-\sqrt{\tilde\eta_M})^{M-1-k}\big)\\
&\qquad + \sqrt{\tilde\alpha_M}\,\big((1+\sqrt{\tilde\eta_M})^{M-1-k} - (1-\sqrt{\tilde\eta_M})^{M-1-k}\big)\Big).
\end{aligned} \tag{3.16}$$

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

Recall (3.1) from section 3.1, where a scaling function R_{sc} was defined 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

$$\alpha_M = \frac{\alpha}{M} + \mathcal{O}(\frac{1}{M^2}), \qquad \gamma_M = \frac{\gamma}{M} + \mathcal{O}(\frac{1}{M^2}). \tag{3.17}$$

Indeed, if condition (3.17) 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 3.2. *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 (3.17) for some positive real numbers α, γ. Then c_M and K_M converge to c_∞ and K_∞, respectively, as $M \to \infty$, where*

$$c_\infty := \lim_{M\to\infty} c_M = \Big(1+\sqrt{\tfrac{\alpha}{\gamma}}\Big)e^{\sqrt{\alpha\gamma}} + \Big(1-\sqrt{\tfrac{\alpha}{\gamma}}\Big)e^{-\sqrt{\alpha\gamma}} = 2\cdot\sum_{k=0}^{\infty}\frac{1}{k!}\,\alpha^{k \bmod 2}\,(\alpha\gamma)^{\lfloor\frac{k}{2}\rfloor}, \tag{3.18}$$

$$K_\infty := \lim_{M\to\infty} K_M = \frac{2}{\big(1+\sqrt{\frac{\gamma}{\alpha}}\big)e^{\sqrt{\alpha\gamma}} + \big(1-\sqrt{\frac{\gamma}{\alpha}}\big)e^{-\sqrt{\alpha\gamma}}} = \frac{\sqrt{\alpha}}{\sqrt{\alpha}\cosh(\sqrt{\alpha\gamma}) + \sqrt{\gamma}\sinh(\sqrt{\alpha\gamma})}. \tag{3.19}$$

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

$$f_L(q) = \begin{cases} \sqrt{\gamma}\cdot K_\infty\cdot\Big(\sqrt{\gamma}\,\cosh\big(\sqrt{\alpha\gamma}(1-q)\big) + \sqrt{\alpha}\,\sinh\big(\sqrt{\alpha\gamma}(1-q)\big)\Big) & \textit{if}\quad q\in(0,1],\\ K_\infty\cdot\alpha\cdot\exp\big(-\alpha(q-1)\big) & \textit{if}\quad q>1. \end{cases} \tag{3.20}$$

Proof. If relation (3.17) holds, then in order to derive (3.18) and (3.19) from (3.10) and (3.14), 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 (3.18) is obtained by series expansion. Similarly, expression (3.20) for f_L follows from equations (3.15a), (3.16) and (3.13). □

Observe that f_L as defined in proposition 3.2 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

$$f_L(1+) - f_L(1-) \quad = \quad K_\infty \cdot (\alpha - \gamma). \tag{3.21}$$

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 section 3.1 we interpreted the discretisation degree M as the number of subintervals of $[-r,0]$, where $r > 0$ is the length of the delay that appears in equation (2.1). Anticipating the discussion of chapter 5, 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 as

$$\tilde{f}_L(t) := \tfrac{1}{r} f_L\big(\tfrac{t}{r}\big), \qquad t \in (0,\infty). \tag{3.22}$$

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. This is the subject of chapter 4. We will return to the issue of residence times in section 4.5.

3.4 Two measures of resonance

Drawing on the residence time distribution of the Markov chain X^M that was studied in section 3.3 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 have the guarantee that an eventual resonance behaviour of the trajectories of X^M is independent of the initial distribution. From section 3.2 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 (Brémaud, 1999: p. 130). In section 2.2.3 we saw an analogous result for the segment process of a solution to (2.1), the equation of the reference model.

Let us 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 (3.1), for example, such that condition (3.17) is satisfied. Under this assumption we let the discretisation 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. This view enables us to define what we mean by stochastic resonance in the two state model.

Recall from section 3.3 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 (3.12). The distribution which L_M induces on $\{1,\ldots,M\}$ is given – up to a renormalizing factor – by equations (3.16) and (3.15a). A natural characteristic seems to be the jump in the density of the residence time distribution f_L that we encountered in section 3.3. In discrete time, i. e. with discretisation degree $M \in \mathbb{N}$, we set

$$\upsilon_M := M\cdot\big(L_M(M+1) - L_M(M)\big). \tag{3.23}$$

Because of (3.13), (3.15a) and (3.21) we have

$$\upsilon_M = M\cdot K_M\cdot\big((1-\gamma_M)\cdot\alpha_M - \gamma_M\big), \tag{3.24}$$

$$\upsilon_\infty := \lim_{M\to\infty}\upsilon_M = K_\infty\cdot(\alpha-\gamma). \tag{3.25}$$

To consider the height of the discontinuity of f_L as a measure of resonance has already been proposed by Masoller (2003), cf. section 1.4. Following this proposal we define what stochastic resonance means according to the jump characteristic.

Definition 3.1. 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 in $\mathbf{C}^2((0,\infty))$,

(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

$$\hat\kappa_M := \sum_{k=1}^{M} L_M(k), \qquad \kappa_M^{(q)} := \sum_{k=M+1}^{\lfloor (q+1)M\rfloor} L_M(k). \tag{3.26}$$

By summation over k we see from (3.13) that

$$\hat\kappa_M = 1-(1-\gamma_M)\cdot K_M, \tag{3.27a}$$

$$\kappa_M^{(q)} = (1-\gamma_M)\cdot K_M\cdot\big(1-(1-\alpha_M)^{\lfloor qM\rfloor}\big), \tag{3.27b}$$

and letting M tend to infinity we get

$$\hat\kappa_\infty := \lim_{M\to\infty}\hat\kappa_M = 1-K_\infty, \tag{3.28a}$$

$$\kappa_\infty^{(q)} := \lim_{M\to\infty}\kappa_M^{(q)} = K_\infty\cdot(1-e^{-q\cdot\alpha}). \tag{3.28b}$$

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.[2]

In (3.26) 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 $\mathbb{N} \setminus \{1,\dots,M\}$, $\kappa_M^{(1)}$ 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 3.2. 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 in $\mathbf{C}^2((0,\infty),[0,1])$,

(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

$$\lim_{q\downarrow 0} \frac{1}{q} \cdot \kappa_\infty^{(q)} \quad = \quad K_\infty \cdot \alpha \quad = \quad f_L(1+). \tag{3.29}$$

Here, f_L is the density of the residence time distribution from proposition 3.2 and $f_L(1+)$ is the right-hand limit appearing in equation (3.21), which gives the height of the discontinuity of f_L.

Of course, definition 3.2 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.

As the resonance characteristics are defined over the distribution of the Markov chain X^M in the stationary regime, it is natural to ask, how the distributions of X^M, $M \in \mathbb{N}$, behave in the limit $M \to \infty$. More precisely, we shall study weak convergence of those distributions on a suitable Skorokhod space. This task is carried out in chapter 4.

[2] Again, the time correspondence is to be understood as a means of illustration only. We are not yet in a position to make those statements precise.

Chapter 4

The two state model in continuous time

Our aim in this chapter is to justify the passage from time discretisation degree M to the limit $M \to \infty$ as undertaken in sections 3.3 and 3.4. 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 3.3. Since the measures of resonance introduced in section 3.4 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 cadlag 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 section 4.1.1, while in section 4.1.2 we present $D_\infty := D_{\{-1,1\}}([-r, \infty))$, the space of all $\{-1, 1\}$-valued cadlag 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 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. This is done in section 4.2.1.

Now, because of the shift condition (3.2) from section 3.1 one may regard X^M as being a process with trajectories in D_∞. If the discretisation 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 $\mathrm{S}_M \subset D_0$. The induced measures are defined in section 4.2.2.

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 4.3. Weak convergence of the distributions on $\mathcal{B}(D_\infty)$ is the object of section 4.4.

Finally, in section 4.5, 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 $\{-1, 1\}$-valued cadlag functions

Here, we specialize results from the literature on the Skorokhod spaces $D^0_{\mathbb{R}} = D_{\mathbb{R}}([-r, 0])$ and $D^\infty_{\mathbb{R}} = D_{\mathbb{R}}([-r, \infty)$, summarized in appendices A.2 and A.3, respectively, to the corresponding Skorokhod spaces of $\{-1, 1\}$-valued cadlag functions, namely $D_0 = D_{\{-1,1\}}([-r, 0])$ and $D_\infty = D_{\{-1,1\}}([-r, \infty))$.

4.1.1 The Skorokhod space D_0

There are two equivalent ways of topologizing D_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 A.4 holds.[1]

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 A.4 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 (A.1) and (A.2), 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 4.1. *Let $f \in D_0$, $\delta \in (0, r)$, and let $I \subseteq [-r, 0]$ be an interval. Then*

$$w(f, I) \in \{0, 2\}, \qquad w(f, I) = 0 \;\Leftrightarrow\; f \text{ is constant on the interval } I, \tag{4.1}$$

$$\#J(f) \in \mathbb{N}_0, \tag{4.2}$$

$$\tilde{w}(f, \delta) \in \{0, 2\}, \qquad \tilde{w}(f, \delta) = 0 \;\Leftrightarrow\; \zeta_f > \delta. \tag{4.3}$$

Proof. Obviously, $|f(s) - f(t)| \in \{0, 2\}$ for all $s, t \in [-r, 0]$, and (4.1) is a consequence of (A.2), 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 cadlag 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 A.5, which states necessary and sufficient conditions for compactness in $D_{\mathbb{R}}^0$, takes on a simple form in the present context.

Proposition 4.2. *Let $A \subseteq D_0$. Then the closure of A is compact in the Skorokhod topology if and only if* $\inf\{\zeta_f \mid f \in A\} > 0$.

[1] Skorokhod spaces for E-valued functions on the infinite interval $[0, \infty)$ are defined in Ethier and Kurtz (1986).

Proof. Condition (i) of theorem A.5 is satisfied for any $A \subseteq D_0$. Hence, we must show that condition (ii) of A.5 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 (4.3) 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.

4.1.2 The Skorokhod space D_∞

First observe that for any $t > -r$ the space $D_{\{-1,1\}}([-r,t])$ of all $\{-1,1\}$-valued cadlag 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 cadlag functions on $[-r,\infty)$. Clearly, D_∞ is a subset of $D^\infty_{\mathbb{R}}$, but it is also closed with respect to the Skorokhod topology of $D^\infty_{\mathbb{R}}$ as can be seen from proposition A.1 and section 4.1.1. We may therefore restrict the topology of $D^\infty_{\mathbb{R}}$ and its metrics d_∞, d°_∞, thus topologizing D_∞.

4.2 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 cadlag process.

4.2.1 Embedding into D_0

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.[2] Approximation results of this kind were obtained for the multi-dimensional version of equation (2.5) by Scheutzow (1983, 1984). The method is more powerful, though, as Lorenz (2003) shows, where weak convergence of the approximating processes to solutions of the multi-dimensional version of (2.4) 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, $\tilde{\imath}_M :\ Z \mapsto f_Z$ 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 $\mathcal{B}(D_0)$.

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

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

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 chapter 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 discretisation 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 4.3 and 4.4 still hold true.

Following the notation of section 4.1.1, 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 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 3.2 in the sense that $\#J(Z) = \mathcal{J}(Z)$.

For the proof of convergence it will be useful to partition S_M into subsets of elements which have equal number of inner jumps. With $i \in \{0, \ldots, M-1\}$ set

$$\mathrm{S}_M(i) := \left\{ Z \in \mathrm{S}_M \mid \#\dot{J}(f_Z) = i \right\}.$$

We then have $\mathrm{S}_M = \bigcup_{i=0}^{M-1} \mathrm{S}_M(i)$ a pairwise disjoint union.

4.2.2 Embedding into D_∞

Let $M \in \mathbb{N}$. 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.

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 $(Y^M_n)_{n \in \{-M, -M+1, \ldots\}}$ be the sequence of current states of X^M as defined at the beginning of section 3.3. 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

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

thereby defining probability measures on $\mathcal{B}(D_\infty)$. Note that $\tilde{\mathrm{P}}^M_f$, $\tilde{\mathrm{P}}_M$ are well defined and correspond to the distribution of X^M with $X^M_0 \overset{d}{\sim} Z(f)$ and $X^M_0 \overset{d}{\sim} \pi_M$, respectively.

4.3 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 the Borel σ-algebra $\mathcal{B}(D_0)$ converges weakly to a probability measure $\tilde{\pi}$. Since (D_0, d°_S) is separable, theorem 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 A.2 in the appendix, 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 $\tilde{\pi}$ is the unique limit point of $(\tilde{\pi}_M)$.

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.

4.3.1 Preliminaries

In order to construct compact sets for the proof of tightness we need special subsets of S_M. Let $N \in \mathbb{N}$, $Z \in \mathrm{S}_N$, and set for $M \in \mathbb{N} \setminus \{1, \ldots, N-1\}$

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

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

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 3.1 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. The sets $\mathrm{S}_M(i)$, $i \in \{0, \ldots, M-1\}$, form a partition of S_M into subsets of elements of equal number of inner discontinuities. 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 in section 4.3.2, we need two more lemmata. Let us start by estimating the number of elements of $\mathrm{S}_M(i)$ and $\mathrm{U}^N_M(Z)$, respectively.

Lemma 4.1.

$$\forall\, M \in \mathbb{N}\; \forall\, i \in \{0, \ldots, M-1\} : \quad \#\mathrm{S}_M(i) \;=\; 4 \cdot \binom{M-1}{i}, \tag{4.4}$$

$$\forall\, N \in \mathbb{N}\; \forall\, Z_1, Z_2 \in \mathrm{S}_N\; \forall\, M \ge N : \quad Z_1 \ne Z_2 \;\Rightarrow\; \mathrm{U}^N_M(Z_1) \cap \mathrm{U}^N_M(Z_2) = \emptyset, \tag{4.5}$$

$$\begin{gathered} \forall\, N \in \mathbb{N}\; \forall\, i \in \{0, \ldots, \lfloor\sqrt{N}\rfloor - 1\}\; \forall\, M \ge N\; \forall\, Z \in \mathrm{S}_N(i) : \\ \Big(\big\lfloor \tfrac{2M}{2N+1} \big\rfloor\Big)^i \;\le\; \#\mathrm{U}^N_M(Z) \;\le\; \Big(\big\lfloor \tfrac{2M}{2N+1} \big\rfloor + 1\Big)^i . \end{gathered} \tag{4.6}$$

Proof. Any element $f \in \tilde{\iota}_M\big(\mathrm{S}_M(i)\big)$ 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 (4.4) 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 $\mathrm{U}^N_M(.)$, for $\mathrm{U}^N_M(Z_1) \cap \mathrm{U}^N_M(Z_2) \ne \emptyset$ 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 $Z_1 \ne Z_2$. 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 $|s - s_2| \ge \frac{r}{N}$ 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 (4.5).

An element $\tilde{Z} \in \mathrm{U}_M^N(Z)$ is determined by the positions of its $\#\dot{J}(Z)$ inner discontinuities, where $\{k \cdot \frac{r}{M} - r \mid k \in \{1, \ldots, M-1\}\}$ is the set of possible such positions. If $s \in \dot{J}(Z)$, then there is $k \in \{1, \ldots, N-1\}$ 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 (4.6) is now a consequence of

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

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

The second lemma 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 (3.17) as in section 3.3, where we considered convergence of the residence time distributions.

Lemma 4.2. *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*

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

Suppose that the sequences of transition probabilities $(\alpha_M)_{M \in \mathbb{N}}$, $(\gamma_M)_{M \in \mathbb{N}}$ satisfy relation (3.17) *for some transition rates $\alpha, \gamma > 0$. Then for all $\epsilon > 0$*

$$\exists\, \tilde{N}_0 \in \mathbb{N}\; \forall\, N_0 \geq \tilde{N}_0\; \exists\, \tilde{N} \in \mathbb{N}\; \forall\, N \geq \tilde{N}\; \exists\, \tilde{M} \in \mathbb{N}\; \forall\, M \geq \tilde{M}: \quad \psi_1, \tag{4.7}$$

$$\forall\, N_0 \in \mathbb{N}\; \exists\, \tilde{N} \in \mathbb{N}\; \forall\, N \geq \tilde{N}\; \exists\, \tilde{M} \in \mathbb{N}\; \forall\, M \geq \tilde{M}: \quad \psi_2, \tag{4.8}$$

$$\exists\, \tilde{N}_0 \in \mathbb{N}\; \forall\, N_0 \geq \tilde{N}_0\; \exists\, \tilde{N} \in \mathbb{N}\; \forall\, N \geq \tilde{N}: \quad \psi_3. \tag{4.9}$$

Finally, it holds that

$$\forall\, \epsilon > 0\; \exists\, N_0, \tilde{N} \in \mathbb{N}\; \forall\, N \geq \tilde{N}\; \exists\, \tilde{M} \in \mathbb{N}\; \forall\, M \geq \tilde{M}: \quad \psi_1 \;\wedge\; \psi_2 \;\wedge\; \psi_3. \tag{4.10}$$

Proof. Formula (4.10) follows by "putting together" (4.7), (4.8) and (4.9), where $N_0 = N_0(\epsilon)$ can be chosen as the maximum of $\tilde{N}_0$ according to (4.7) and $\tilde{N}_0$ according to (4.9), $\tilde{N} = \tilde{N}(\epsilon, N_0)$ 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 3.1, where the normalizing constant c_M for the probability measure π_M was defined, and equation (3.18) of proposition 3.2, where we obtained an explicit expression for $c_\infty = \lim_{M\to\infty} c_M$. In analogy to propositions 3.1 and 3.2, respectively, we set

$$c_{M,M_0} := 2 \cdot \sum_{k=0}^{M_0} \binom{M}{k} \Big(\frac{\alpha_M}{1-\gamma_M}\Big)^{k \bmod 2} \Big(\frac{\alpha_M \cdot \gamma_M}{(1-\alpha_M)(1-\gamma_M)}\Big)^{\lfloor \frac{k}{2} \rfloor}, \qquad M \in \mathbb{N},\ M_0 \in \{1, \ldots, M\}.$$

$$c_{\infty,M_0} := 2 \cdot \sum_{k=0}^{M_0} \frac{1}{k!} \alpha^{k \bmod 2} (\alpha\gamma)^{\lfloor \frac{k}{2} \rfloor}, \qquad M_0 \in \mathbb{N},$$

Because of relation (3.17) it holds that $\forall\, \tilde{\epsilon} \in (0,1)\ \forall\, M_0 \in \mathbb{N}\ \exists\, \tilde{M} \in \mathbb{N}\ \forall\, M \geq \tilde{M}\ \forall \{0, \ldots, M_0\}$:

$$(4.11) \qquad \begin{aligned} (1+\tilde{\epsilon}) \cdot (\alpha\gamma)^k \ &\geq\ M^{2k} \cdot \big(\tfrac{\alpha_M \cdot \gamma_M}{(1-\alpha_M)(1-\gamma_M)}\big)^k \ \geq\ (1-\tilde{\epsilon}) \cdot (\alpha\gamma)^k \\ \wedge\ \ (1+\tilde{\epsilon})\alpha \ &\geq\ M \cdot \tfrac{\alpha_M}{1-\gamma_M} \ \geq\ (1-\tilde{\epsilon})\alpha \\ \wedge\ \ 1 \ &\geq\ \tfrac{M \cdot \ldots \cdot (M-2k+1)}{M^{2k}} \ \geq\ \tfrac{M \cdot \ldots \cdot (M-2k)}{M^{2k+1}} \ \geq\ 1-\tilde{\epsilon}. \end{aligned}$$

In view of the above we have

$$(4.12) \qquad \begin{aligned} &\forall\, \tilde{\epsilon} \in (0,1)\ \exists\, \tilde{M}_0 \in \mathbb{N}\ \forall\, M_0 \geq \tilde{M}_0\ \exists\, \tilde{M} \in \mathbb{N}\ \forall\, M \geq \tilde{M} : \\ &1 \ \geq\ \frac{c_{\infty,M_0}}{c_\infty} \ \geq\ 1-\tilde{\epsilon} \ \ \wedge\ \ M \ \geq\ M_0 \ \ \wedge\ \ \Big|\frac{c_{M,M_0}}{c_\infty} - 1\Big| + \Big|\frac{c_\infty}{c_{M,M_0}} - 1\Big| \ \leq\ \tilde{\epsilon}. \end{aligned}$$

To conclude the preparations, recall that for $Z \in \mathrm{S}_M(i)$, where $M \in \mathbb{N}$, $i \in \{0, \ldots, M-1\}$, 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 $\forall\, N \geq \tilde{N}\ \forall\, i \in \{0, \ldots, N_0\}$:

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

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

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

where (4.11) 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 3.1. According to the choice of N_0, N, M and because of (4.4) 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

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

Since $\tilde\epsilon \in (0,1)$ was arbitrary, we may set $\tilde\epsilon := \frac{\epsilon}{6}$, thereby establishing (4.7).

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

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

which is possible because of (4.11). 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 3.1.

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 (4.8) is finished.

Let $\tilde\epsilon > 0$. Choose $\tilde N_0 \in \mathbb{N}$ according to (4.12) 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 3.1 we have for $N_0 \geq \tilde N_0$, $N \geq \hat N = \hat N(\tilde\epsilon, N_0)$

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

This establishes (4.9). □

4.3.2 Tightness and uniqueness of the limit point

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 4.2.1.

Proposition 4.3. *Suppose the sequences of transition probabilities $(\alpha_M)_{M\in\mathbb{N}}$, $(\gamma_M)_{M\in\mathbb{N}}$ satisfy relation* (3.17) *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. We will apply lemma 4.2 several times. 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 A.2 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 4.2.1 the definition of $\tilde\imath_M$. For all natural numbers $N_0 < N \leq M$ we have

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

Let $\epsilon > 0$. According to (4.7) we can find natural numbers $N_0 < N \leq \tilde{M}$ such that for all $M \geq \tilde{M}$:

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

Fix N, $\tilde{M}$. In analogy to the definition of U_M^N 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 4.2 we see that $\mathrm{cl}(\tilde{A})$, the closure of $\tilde{A}$, is compact with respect to the Skorokhod topology. By definition we have $\mathrm{U}_M^N \subset \tilde{A}$ for all $M \geq \tilde{M} \geq N$. Define

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

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

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

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 A.1 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 real-valued bounded and uniformly continuous 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 (4.10). 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[3] we see that $d_S(f_Z, f_{\tilde{Z}}) \leq \delta$ for all $Z\in\mathrm{S}_N(i)$ and $\tilde{Z}\in\mathrm{U}_M^N(Z)$ if $i\in\{0,\dots,N_0\}$ 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, $\tilde{Z}$.

[3]The sets $\mathrm{U}_M^N(Z)$ were defined at the beginning of section 4.3.1.

Finally, choose a natural number $\tilde M = \tilde M(\epsilon, N_0, \tilde N, N)$ according to (4.10). Then for $M \geq \tilde M$

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

□

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

Proposition 4.4. *Let $\tilde\pi$ be the weak limit of $(\pi_M)_{M \in \mathbb{N}}$ according to proposition 4.3. 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 $i \in \mathbb{N}_0$

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

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.[4] 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 A.1, proposition 3.1, equations (3.19) and (4.4) of proposition 3.2 and lemma 4.1, respectively, under the scaling condition (3.17). □

4.4 Convergence of the chain distributions on D_∞

Let the notation be that of section 4.2.2, 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 (1987).

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

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

part and ν a random measure, namely the compensator of the jump measure of the semimartingale (Jacod and Shiryaev, 1987: 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 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 (Jacod and Shiryaev, 1987: 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 4.2.2. Recall that $\tilde{\mathrm{P}}_Z^M$ is the distribution of the $\{-1,1\}$-valued cadlag process $(\tilde{Y}^M(t))_{t\geq -r}$ induced by the sequence Y^M of current states of X^M when time discretisation 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 $\tilde{\nu}^{M,Z}\colon D_\infty\times[0,\infty)\times\mathbb{R}\to[0,\infty]$ 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.[5] Set $s(k) := \frac{r}{M}\cdot k$, $k\in\{-M,-M+1,\ldots\}$. According to II.3.11 in Jacod and Shiryaev (1987: p. 94) it holds for all functions ψ, all $t\geq 0$, $\tilde{\omega}\in D_\infty$ (all probabilities with respect to $\tilde{\mathrm{P}}_Z^M$)

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

where the formula of Bayes 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

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

[5] See Jacod and Shiryaev (1987: p. 66) for a definition of the integral process w. r. t. a random measure.

Assume that scaling relation (3.17) is satisfied for some positive transition rates α, γ. Then for all functions $\psi\colon \mathbb{R} \to \mathbb{R}$, $\omega \in D_{\mathbb{R}}$, $t \geq 0$ it holds that

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

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 $(Y(t))_{t\geq 0}$, cf. Jacod and Shiryaev (1987: pp. 68-69). We have for all functions $\psi\colon \mathbb{R}\to\mathbb{R}$, $\omega \in D_{\mathbb{R}}$, $t \geq 0$

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

Theorem IX.2.31 in Jacod and Shiryaev (1987: 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 (1987: p. 80) the second property implies that

- $(\psi * \mu - \psi * \nu^f)_{t\geq 0}$ 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 $\tau_1, \tau_2, \ldots$ which exhaust the jump positions. The above local martingale property must be applied to show that any two solution measures to the semimartingale problem coincide on the sets $\{\tau_n \leq t\}$ for all $t \geq 0$, $n \in \mathbb{N}$. 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 (1987: 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 $\tilde{\mathrm{P}}^f \in \mathcal{M}^1_+(D_\infty)$ as $\tilde{\mathrm{P}}^f := Q^f \circ \theta_f^{-1}$. 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 4.3. As a consequence of proposition 4.4 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

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

If $f_1, \ldots, f_n \in D_0$ with $f_i(0) = f_i(0-)$, $i \in \{1,\ldots,n\}$, then any convex combination of the sequences $(\tilde{\mathrm{P}}^M_{Z(f_1,M)}), \ldots, (\tilde{\mathrm{P}}^M_{Z(f_n,M)})$ converges weakly to the corresponding convex combination of the measures $\tilde{\mathrm{P}}^{f_1}, \ldots, \tilde{\mathrm{P}}^{f_n}$. An approximation argument analogous to that in the proof of proposition 4.3 leads to

Proposition 4.5. *Suppose scaling relation* (3.17) *holds. Let* $\tilde{\mathrm{P}}_M$, $M \in \mathbb{N}$, *be defined as in section 4.2.2, and let* $\tilde\pi$ *be the weak limit of* $(\pi_M)_{M\in\mathbb{N}}$ *according to proposition 4.3. Then there is a probability measure* $\tilde{\mathrm{P}} \in \mathcal{M}^1_+(D_\infty)$ *such that* $\tilde{\mathrm{P}}_M \xrightarrow{w} \tilde{\mathrm{P}}$.

4.5 Residence times revisited

In section 3.3 we calculated the residence time distribution for the two state model of discretisation 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 4.4, 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 3.3.

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.3. We continue to work directly on the canonical space. Define a mapping

$$(4.13) \qquad \tau \colon D_\infty \to [0, \infty], \qquad \tau(f) := \inf\{t \geq 0 \mid f(t) = -1\}.$$

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 cadlag property of the elements of D_∞ has been exploited. Because of this property the infimum in (4.13) 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

$$(4.14) \qquad \tilde{A}_\delta := \big\{ f \in D_\infty \;\big|\; \exists\, \tilde{\delta} \in [0, \delta) : \; f(t) = -1 \; \forall\, t \in [-\delta r, -\tilde{\delta} r) \;\wedge\; f(t) = 1 \; \forall\, t \in [-\tilde{\delta} r, 0] \big\}.$$

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

$$(4.15) \qquad F_\delta(t) := \tilde{\mathrm{P}}\big(\tau \leq t \mid \tilde{A}_\delta\big), \quad t \in [0, \infty),$$

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 determines indeed a probability distribution on $[0, \infty)$.

Let $\tilde{f}_L$ be the residence time distribution density in the limit of discretisation degree M tending to infinity as given by (3.22). Set

$$(4.16) \qquad F(t) := \int_0^t \tilde{f}_L(s) ds, \quad t \in [0, \infty).$$

We have to show that $F_\delta(t)$ tends to $F(t)$ as δ goes to zero for each $t \in [0, \infty)$. From the proof of proposition 4.6 it will become clear that in (4.13) and (4.14), the definitions of τ and $\tilde{A}_\delta$, respectively, instead of time zero we could have chosen any starting time $t_0 \geq 0$.

Proposition 4.6. *Suppose scaling relation* (3.17) *holds. Let the distribution functions* F_δ, $\delta \in (0, 1)$, *and* F *be defined by* (4.15) *and* (4.16), *respectively. Then*

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

Proof. 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 4.2.2. Recall that $\tilde{\mathrm{P}}_M$ is the measure induced by the sequence of current states of 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 4.5 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 A.1 yields

$$\text{(4.17)} \quad \tilde{\mathrm{P}}_M\big(\tau \le t \mid \tilde{A}_\delta\big) \overset{M\to\infty}{\longrightarrow} \tilde{\mathrm{P}}\big(\tau \le t \mid \tilde{A}_\delta\big), \quad \text{i.e.} \quad F_\delta^M(t) \overset{M\to\infty}{\longrightarrow} F_\delta(t) \quad \text{for all } \delta \in (0,1),\ t > 0.$$

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 (4.17) 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

$$\text{(4.18)} \qquad \big|F_\delta^M(t) - F(t)\big| \le \varepsilon \quad \text{for all } \delta \in (0,\delta_0),\ M \ge M_0.$$

As in section 3.3, let $(Y_n^M)_{n\in\{-M,-M+1,\dots\}}$ be the random sequence of current states on $(\Omega,\mathcal{F})$ at discretisation 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

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

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 (3.11), the definition of the residence time distribution $L_M(.)$ of discretisation degree M, we set

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

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 3.3, 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 (3.12), the definition of the tail constant K_M, we set for $\delta \in (0,1)$ and M big enough

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

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

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

As a consequence of proposition 3.1, 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 (3.14) we find that

$$K^j_{\delta,M} = \frac{2}{\Big(1+\sqrt{\frac{\tilde\gamma_M}{\tilde\alpha_M}}\Big)\big(1+\sqrt{\tilde\eta_M}\big)^{M-\lfloor\delta M\rfloor}+\Big(1-\sqrt{\frac{\tilde\gamma_M}{\tilde\alpha_M}}\Big)\big(1-\sqrt{\tilde\eta_M}\big)^{M-\lfloor\delta M\rfloor}} =: K_{\delta,M}. \tag{4.19}$$

In order to calculate $L^j_{\delta,M}$ we apply proposition 3.1 again in a way similar to that of section 3.3. Let $\delta\in(0,1)$, let $M\in\mathbb{M}$ be such that $\lfloor\delta M\rfloor\geq 1$, and $j\in\{0,\dots,\lfloor\delta M\rfloor-1\}$. Then for $k\in\{1,\dots,M-\lfloor\delta M\rfloor\}$

$$\begin{aligned} L^j_{\delta,M}(k) &= \frac{\sqrt{\tilde\gamma_M}}{2}\cdot K_{\delta,M}\cdot\Big(\sqrt{\tilde\gamma_M}\,\big((1+\sqrt{\tilde\eta_M})^{M-\lfloor\delta M\rfloor-k}+(1-\sqrt{\tilde\eta_M})^{M-\lfloor\delta M\rfloor-k}\big)\\ &\qquad +\sqrt{\tilde\alpha_M}\,\big((1+\sqrt{\tilde\eta_M})^{M-\lfloor\delta M\rfloor-k}-(1-\sqrt{\tilde\eta_M})^{M-\lfloor\delta M\rfloor-k}\big)\Big). \end{aligned}\tag{4.20a}$$

While $L^j_{\delta,M}(k)$ in (4.20a) does not vary with j as long as $k\leq M-\lfloor\delta M\rfloor$, for $k\in\{M-\lfloor\delta M\rfloor+1,\dots,M-j\}$ it holds that

$$L^j_{\delta,M}(k) = K_{\delta,M}\cdot\gamma_M\cdot(1-\gamma_M)^{k-M+\lfloor\delta M\rfloor-1}, \tag{4.20b}$$

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

$$L^j_{\delta,M}(k) = K_{\delta,M}\cdot\alpha_M\cdot(1-\gamma_M)^{\lfloor\delta M\rfloor-j}\cdot(1-\alpha_M)^{k-M+j-1}. \tag{4.20c}$$

Now, let the discretisation degree M tend to infinity, where we assume that scaling relation (3.17) holds for some rates α, γ. From (4.19) we see that

$$\begin{aligned} K_{\infty,\delta} := \lim_{M\to\infty} K_{M,\delta} &= \frac{2}{\big(1+\sqrt{\frac{\gamma}{\alpha}}\big)e^{(1-\delta)\sqrt{\alpha\gamma}}+\big(1-\sqrt{\frac{\gamma}{\alpha}}\big)e^{-(1-\delta)\sqrt{\alpha\gamma}}}\\ &= \frac{\sqrt{\alpha}}{\sqrt{\alpha}\cosh\big((1-\delta)\sqrt{\alpha\gamma}\big)+\sqrt{\gamma}\sinh\big((1-\delta)\sqrt{\alpha\gamma}\big)}. \end{aligned}\tag{4.21}$$

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 (4.20a) we find that

$$\lim_{M\to\infty} M\cdot L^{j_M}_{\delta,M}\big(\lfloor qM\rfloor\big) = \sqrt{\gamma}\cdot K_{\delta,\infty}\cdot\Big(\sqrt{\gamma}\,\cosh\big(\sqrt{\alpha\gamma}(1-\delta-q)\big)+\sqrt{\alpha}\,\sinh\big(\sqrt{\alpha\gamma}(1-\delta-q)\big)\Big). \tag{4.22a}$$

If $q\in(1-\delta,1)$, then a rough estimate of (4.20b) and (4.20c), respectively, yields

$$\begin{aligned} \limsup_{M\to\infty} M\cdot L^{j_M}_{\delta,M}\big(\lfloor qM\rfloor\big) &\leq \max\{\alpha,\gamma\}\cdot K_{\delta,\infty},\\ \liminf_{M\to\infty} M\cdot L^{j_M}_{\delta,M}\big(\lfloor qM\rfloor\big) &\geq \min\{\alpha,\gamma\}\cdot K_{\delta,\infty}\cdot e^{-\delta\gamma}\cdot e^{-(q+\delta-1)\alpha}. \end{aligned}\tag{4.22b}$$

On the other hand, if $q\geq 1$, then by (4.20c) we have

$$\begin{aligned} \limsup_{M\to\infty} M\cdot L^{j_M}_{\delta,M}\big(\lfloor qM\rfloor\big) &\leq \alpha\cdot K_{\delta,\infty}\cdot e^{-(q-1)\alpha},\\ \liminf_{M\to\infty} M\cdot L^{j_M}_{\delta,M}\big(\lfloor qM\rfloor\big) &\geq \alpha\cdot K_{\delta,\infty}\cdot e^{-\delta\gamma}\cdot e^{-(q+\delta-1)\alpha}. \end{aligned}\tag{4.22c}$$

Notice that convergence in (4.21) as well as in (4.22) 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 3.2. Taking the time discretisation into account, we obtain $\tilde{f}_L$ as given by (3.22) 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 (4.18) is fulfilled. The assertion then follows. $\square$

Chapter 5

Connection between the reduced and the reference model

The aim of this chapter is to provide a heuristic way of establishing the missing link between our original model, which is given by equation (2.1), and the reduced model developed in chapter 3. What we have to do is calculate the transition rates α, γ of the two state model as functions of the delay parameter β, the noise parameter σ, the delay time r and some quantities related to the shape of the potentials V and U. The situation here is quite similar to the one that was studied by Tsimring and Pikovsky (2001), 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. The Kramers rate is described in section 5.1 and employed in section 5.2, where we calculate escape rates from potentials that should mirror the "effective dynamics" of solutions to equation (2.1).

In section 5.3 we return to the issue of stochastic resonance. The resonance characteristics defined in section 3.4 can now 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 behaviour of the original model.

The last section summarizes what we have found and points out problems we did not resolve. An important question in the context of the approach to stochastic resonance followed here is whether the reduced model – at least in the limit of small noise – really captures the effective dynamics of the original model. This, of course, depends on which feature of the original model is considered as being characteristic and which measure of resonance one chooses. In our case, it was the distribution of interwell transitions that served as the basis for quantifying stochastic resonance.

5.1 Escape rate from a potential well

Our source for this section is Herrmann et al. (2003) and the references therein. 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 1.1 and 2.1. Consider the SDE

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

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 (5.1) 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

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

Since we are interested in the transition behaviour 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 (Freidlin and Wentzell, 1998) 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 5.1 (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*

$$\lim_{\sigma\downarrow 0} \sigma^2 \cdot \ln\left(\mathrm{E}_{\mathrm{P}}\left(\tau_y(X^{x,\sigma})\right)\right) = q_l, \tag{5.2a}$$

$$\lim_{\sigma\downarrow 0} \mathrm{P}\left(\exp\left(\frac{q_l - \delta}{\sigma^2}\right) < \tau_y(X^{x,\sigma}) < \exp\left(\frac{q_l + \delta}{\sigma^2}\right)\right) = 1 \quad \textit{for all} \quad \delta > 0. \tag{5.2b}$$

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 travelling 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 5.1 implies the existence of different time scales for equation (5.1). 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 (5.2a), 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 example, that $L_l < L_r$. According to (5.2b), 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

$$\mathrm{r_K} \sim \exp\Big(-\frac{2\big(\mathcal{U}(x_{max}) - \mathcal{U}(x_{left})\big)}{\sigma^2}\Big). \tag{5.3}$$

Equations (5.2a) and (5.2b) of theorem 5.1 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 (5.3) can be specified as a function of the second derivative of $\mathcal{U}$ at the positions of the minimum and the potential barrier, respectively. The Kramers rate thus reads

$$\mathrm{r_K} = \mathrm{r_K}(\sigma, \mathcal{U}) \;=\; \frac{\sqrt{|\mathcal{U}''(x_{left})\mathcal{U}''(x_{max})|}}{2\pi}\,\exp\left(-\frac{2|\mathcal{U}(x_{left}) - \mathcal{U}(x_{max})|}{\sigma^2}\right). \tag{5.4}$$

Observe that both the assumption of exponentially distributed interwell transition times and formula (5.4) for the Kramers rate are empirical approximations, where the noise parameter σ is supposed to be sufficiently small.

Recent results by Bovier et al. (2002a,b) 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 (5.3). 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 (5.4) will be good enough.

5.2 An approximation formula for the transition rates

The arguments presented in this section are based on those outlined by Tsimring and Pikovsky (2001). In chapter 3 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 (5.4). More precisely, we must have

$$\alpha = \alpha(\sigma) \;=\; r \cdot \mathrm{r_K}(\sigma, \mathcal{U}_\alpha), \qquad \gamma = \gamma(\sigma) \;=\; r \cdot \mathrm{r_K}(\sigma, \mathcal{U}_\gamma). \tag{5.5}$$

Note that the inclusion of the delay time r as a proportionality factor is necessary, because in the construction of our two state model in chapter 3 we took one unit of time as equivalent to the length of the interval $[-r, 0]$.

Assume that $|\beta|$ and σ^2 are small. Recall from section 2.2.3 that there is a unique stationary probability measure for equation (2.1). Let $(X(t))_{t\geq r}$ denote a (weak or strong) stationary solution. Then it is reasonable to expect that the process X spends most of its time near the positions of the two minima of the double well potential which arises as a deformation of V under the influence of the delay force. With $|\beta|$ small these minima are still near the two minima of V, that is we have with high probability $X(t) \approx 1$ or $X(t) \approx -1$ for any $t \geq 0$.

In constructing the desired potentials $\mathcal{U}_\alpha$, $\mathcal{U}_\gamma$ we may – by symmetry – restrict attention to an escape from the right-hand side of the double well potential. We therefore have to distinguish two cases, namely $X(t) \approx 1 \approx X(t-r)$ and $X(t) \approx 1 \approx -X(t-r)$, the former case corresponding to the condition under which interwell transitions occur with rate α, the latter corresponding to interwell transitions with rate γ. Let $\mathcal{U}_\alpha$, $\mathcal{U}_\gamma$ be $\mathbf{C}^2$-functions such that for all $x \in \mathbb{R}$

$$\mathcal{U}_\alpha'(x) \;=\; V'(x) + \beta\, U'(1), \qquad \mathcal{U}_\gamma'(x) \;=\; V'(x) + \beta\, U'(-1),$$

where the term $\beta\, U'(\pm 1)$ gives strength and direction of the delay force. Because of the symmetry of U we may choose $\mathcal{U}_\alpha(x) = V(x) + \beta\, U'(1)x$ and $\mathcal{U}_\gamma(x) = V(x) - \beta\, U'(1)\, x$, $x \in \mathbb{R}$. Denote by $\underline{x}_\alpha$ the position of the right-hand minimum of $\mathcal{U}_\alpha$ and by $\underline{x}_\gamma$ the position of the right-hand minimum of $\mathcal{U}_\gamma$, and let $\overline{x}_\alpha$, $\overline{x}_\gamma$ denote the positions of the respective maxima.

Expanding $\mathcal{U}'_\alpha$, $\mathcal{U}'_\gamma$ around 1 and recalling that $V'(1) = 0$ we obtain a rough, first order approximation of $\underline{x}_\alpha$ and $\underline{x}_\gamma$, respectively. An analogous calculation, this time using a Taylor expansion around 0 and recalling that $V'(0) = 0$, yields first order approximations for $\overline{x}_\alpha$, $\overline{x}_\gamma$, and we obtain

$$\text{(5.6)} \qquad \underline{x}_\alpha \approx 1 - \frac{U'(1)}{V''(1)}\,\beta, \qquad \underline{x}_\gamma \approx 1 + \frac{U'(1)}{V''(1)}\,\beta,$$

$$\text{(5.7)} \qquad \overline{x}_\alpha \approx -\frac{U'(1)}{V''(0)}\,\beta = \frac{U'(1)}{|V''(0)|}\,\beta, \qquad \overline{x}_\gamma \approx \frac{U'(1)}{V''(0)}\,\beta = -\frac{U'(1)}{|V''(0)|}\,\beta.$$

Notice that $U'(1) > 0$, $V''(1) > 0$ and $V''(0) < 0$ as a consequence of (2.2), our assumptions about V and U, cf. section 2.1. There is an important point to be made here. In the above discussion and in section 3.1, where we introduced the two switching rates α and γ, we assumed that $X(t) \approx 1$ or $X(t) \approx -1$. According to (5.6), 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. The presence of the second order minima is discernible even in the figures of section 2.3, which depict numerically simulated solution paths to equation (2.1).[1]

The problem disappears, of course, if U' is constant except on a small symmetric interval $(-\delta, \delta)$ around the origin (see figure 2.1 c) in section 2.1), for in this case the delay force would not depend on the particular value of $X(t-r)$ provided $|X(t-r)| \geq \delta$.

In order to get explicit approximations for the rates α, γ out of Kramers's formula (5.4), we have to calculate the expressions $|\mathcal{U}''_\alpha(\overline{x}_\alpha)\mathcal{U}''_\alpha(\underline{x}_\alpha)|$, $|\mathcal{U}_\alpha(\overline{x}_\alpha) - \mathcal{U}_\alpha(\underline{x}_\alpha)|$, $|\mathcal{U}''_\gamma(\overline{x}_\gamma)\mathcal{U}''_\gamma(\underline{x}_\gamma)|$, $|\mathcal{U}_\gamma(\overline{x}_\gamma)) - \mathcal{U}_\gamma(\underline{x}_\gamma)|$. By construction, $\mathcal{U}''_\alpha = V'' = \mathcal{U}''_\gamma$, and $V'''(0) = 0$ as a consequence of (2.2). Neglecting terms of order higher than one, from (5.6) and (5.7) we obtain

$$|\mathcal{U}''_\alpha(\overline{x}_\alpha)\mathcal{U}''_\alpha(\underline{x}_\alpha)| \approx |V''(0)\,V''(1)|\left(1 - \beta\,\tfrac{V'''(1)\,U'(1)}{(V''(1))^2}\right), \qquad |\mathcal{U}_\alpha(\overline{x}_\alpha) - \mathcal{U}_\alpha(\underline{x}_\alpha)| \approx L - \beta\, U'(1),$$

$$|\mathcal{U}''_\gamma(\overline{x}_\gamma)\mathcal{U}''_\gamma(\underline{x}_\gamma)| \approx |V''(0)\,V''(1)|\left(1 + \beta\,\tfrac{V'''(1)\,U'(1)}{(V''(1))^2}\right), \qquad |\mathcal{U}_\gamma(\overline{x}_\gamma) - \mathcal{U}_\gamma(\underline{x}_\gamma)| \approx L + \beta\, U'(1),$$

where $L := V(0) - V(1)$ is the height of the potential barrier of V. Set $\rho := |V''(0)V''(1)|$, $\eta := \frac{V'''(1)\,U'(1)}{(V''(1))^2}$, $\tilde{\eta} := \frac{U'(1)}{L}$, then substitution in (5.5) yields

$$\text{(5.8a)} \qquad \alpha = \alpha(\sigma) \approx r \cdot \frac{\sqrt{\rho\,(1 - \eta\,\beta)}}{2\pi}\,\exp\left(-\frac{2L\,(1 - \tilde{\eta}\,\beta)}{\sigma^2}\right),$$

$$\text{(5.8b)} \qquad \gamma = \gamma(\sigma) \approx r \cdot \frac{\sqrt{\rho\,(1 + \eta\,\beta)}}{2\pi}\,\exp\left(-\frac{2L\,(1 + \tilde{\eta}\,\beta)}{\sigma^2}\right).$$

Recall from section 5.1 that the Kramers rate is exact only in the small noise limit. Thus, for the formulae (5.8) 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.

[1] Cf. also the numerical results in Curtin et al. (2004).

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. We do not have anything precise to say about the goodness of the above approximation for the escape rates α, γ, so we limit ourselves to a numerical comparison between the resonance point as predicted by our two state model if we choose α, γ according to (5.8) and the behaviour of solutions to the reference equation (2.1).

5.3 The resonance point according to the reduced model

In section 3.4 we defined two measures of resonance, namely the jump height υ_M of the residence time distribution density $\tilde{f}_L$ and the probabilities $\hat{\kappa}_M$, κ_M of transitions within the first and second delay interval, respectively.[2]

Recall that $M \in \mathbb{N} \cup \{\infty\}$ is the degree of discretisation, where $M = \infty$ denotes the limit $M \to \infty$. In view of the discussion of chapter 4 we are justified in restricting attention to the case $M = \infty$, that is to the two state model in continuous time.

Let K_∞ be as defined by (3.19), that is

$$K_\infty \;=\; \frac{2}{\left(1+\sqrt{\frac{\gamma}{\alpha}}\right)e^{\sqrt{\alpha\gamma}} + \left(1-\sqrt{\frac{\gamma}{\alpha}}\right)e^{-\sqrt{\alpha\gamma}}}.$$

According to (3.25) and (3.28) we have

$$\upsilon_\infty \;=\; K_\infty(\alpha-\gamma), \qquad \hat{\kappa}_\infty \;=\; 1-K_\infty, \quad \kappa_\infty \;=\; K_\infty\cdot(1-e^{-\alpha}).$$

Suppose the transition rates α, γ are functions of the reference model parameters as given by (5.8) 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 and the remaining parameters are sufficiently nice. Approximation formula (5.8) makes sense only for $\sigma > 0$ small. This is no serious limitation, though, as we may choose $r > 0$ big enough so that the critical parameter region for σ lies within the scope of formula (5.8).

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

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

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 3.2, 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 $\sqrt{\alpha\gamma}$.

We now turn to definitions 3.1 and 3.2, our two notions of resonance. Clearly, K_∞, α, γ are smooth functions of $\sigma > 0$, K_∞ is strictly decreasing, while α, γ are strictly increasing, and we have

$$\lim_{\sigma\downarrow 0} K_\infty \;=\; 1, \qquad \lim_{\sigma\uparrow\infty} K_\infty \;=\; 0, \qquad \lim_{\sigma\downarrow 0}\alpha \;=\; 0, \qquad \lim_{\sigma\downarrow 0}\gamma \;=\; 0.$$

For every $\epsilon > 0$, $\sigma_\epsilon > 0$ the delay length r can be chosen big enough so that $K_\infty(\sigma) \leq \epsilon$ for all $\sigma \geq \sigma_\epsilon$. If $\beta > 0$ then $\alpha > \gamma$, if $\beta < 0$ then $\alpha < \gamma$, and $\alpha = \gamma$ if $\beta = 0$. Notice that

$$\lim_{\sigma\downarrow 0}\kappa_\infty \;=\; 0 \;=\; \lim_{\sigma\uparrow\infty}\kappa_\infty,$$

[2]The jump height measure corresponds to a measure of resonance proposed by Masoller (2003).

while $\kappa_\infty(\sigma) > 0$ for all $\sigma \in (0, \infty)$, and κ_∞ has a unique global maximum for all admissible choices of β.

The conditions of definitions 3.1 and 3.2 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 (2001), that is V is the standard quartic potential and U a parabola, see figure 2.1 in section 2.1. For the constants appearing in formula (5.8) we have

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

Set $r = 500$ as in the numerical simulations from section 2.3. With $\beta = 0.1$ 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. Both results seem compatible with simulation, see figures 2.5, 2.6 and 2.7.

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.

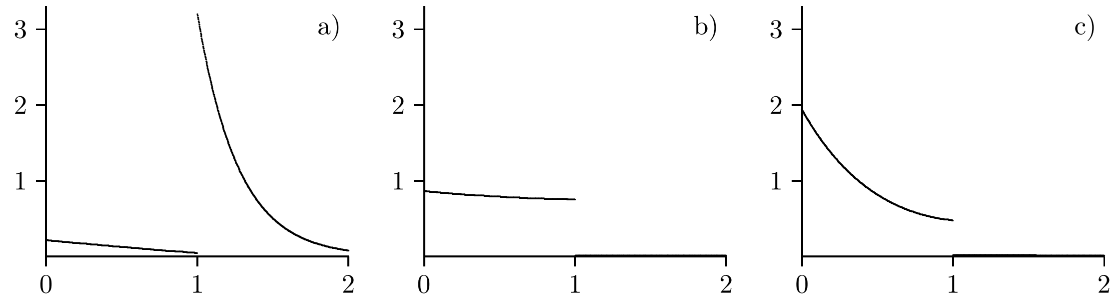

Figure 5.1: Graphs on $[0, 2]$ of the density f_L of the residence time distribution in normalized time. Parameters of the original model: $r = 500$, a) $\sigma = 0.30$, $\beta = 0.1$, b) $\sigma = 0.30$, $\beta = -0.1$, c) $\sigma = 0.35$, $\beta = -0.1$.

There seems to be a discrepancy, now, between the predicted optimal noise level and the level of "most regular" transition behaviour which one would expect from numerical simulation, see figures 2.10 and 2.11. 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 behaviour 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 $4.62r$. Moreover, 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. figure 5.1.

These properties of the residence time distribution support the distinction made in definitions 3.1 and 3.2 between stochastic resonance and pseudo-resonance.

5.4 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 chapter 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. In chapter 4 we studied the passage from discrete to continuous time. 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 the last chapter 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 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 chapter 4 should be a special case of more general convergence results for continuous time Markov chains with delay.

Appendix 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 (1986: ch. 3 §§ 1-3) and Billingsley (1999: § 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

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

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 (Ethier and Kurtz, 1986: 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 $\mathrm{P}_n \xrightarrow{w} \mathrm{P}$, 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 (Ethier and Kurtz, 1986: 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 A.1. *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 (Ethier and Kurtz, 1986: p. 104).

Theorem A.2 (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 (Billingsley, 1999: p. 21).

Theorem A.3. *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 space $D_{\mathbb{R}}^0$

Here, we gather results and definitions from Billingsley (1999: §§ 12-13) on the nature of $D_{\mathbb{R}}^0 := D_{\mathbb{R}}([-r,0])$, the Skorokhod space of all real-valued cadlag 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 cadlag continuity* as

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

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

$$w(f,I) := \sup_{s,t\in I} |f(s) - f(t)|, \quad I \subseteq [-r,0] \text{ an interval.} \tag{A.2}$$

A function $f : [-r,0] \to \mathbb{R}$ is in $D_{\mathbb{R}}^0$ if and only if $\lim_{\delta\downarrow 0} \tilde{w}(f,\delta) = 0$, cf. Billingsley (1999: 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

$$d_S(f,g) := \inf\{\epsilon > 0 \mid \exists \lambda \in \Lambda : \sup_{t\in[-r,0]} |\lambda(t) - t| \le \epsilon \wedge \sup_{t\in[-r,0]} |f(t) - g(\lambda(t))| \le \epsilon\}, \tag{A.3}$$

$$d^\circ_S(f,g) := \inf\{\epsilon > 0 \mid \exists \lambda \in \Lambda : \|\lambda\|_\Lambda \le \epsilon \wedge \sup_{t\in[-r,0]} |f(t) - g(\lambda(t))| \le \epsilon\}. \tag{A.4}$$

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 identy 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 A.4 is a summary of Billingsley (1999: pp. 125-129).

Theorem A.4. *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 cadlag 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 cadlag function. ◊

The following criterion, which is theorem 12.3 in Billingsley (1999: p. 130), is an analogue of the Arzelà-Ascoli theorem for compactness in spaces of continuous functions.

Theorem A.5. *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 space $D^{\infty}_{\mathbb{R}}$

Denote by $D^{\infty}_{\mathbb{R}} := D_{\mathbb{R}}([-r, \infty))$ the space of all real-valued cadlag functions on the interval $[-r, \infty)$. Observe that a cadlag function on $[-r, \infty)$ has at most countably many points of discontinuity (Ethier and Kurtz, 1986: 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 (1986: 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 (A.3) and (A.4) 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 (A.2) 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 (1999: § 16), where a connection is established between $D_{\mathbb{R}}^{\infty}$ and Skorokhod spaces over finite intervals. Observe that all definitions and properties of $D_{\mathbb{R}}^{0} = D_{\mathbb{R}}([-r,0])$ carry over to the space $D_{\mathbb{R}}([-r,t])$ for any $t \in (-r,\infty)$. Set

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

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

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

$$\psi_m:\ D_{\mathbb{R}}^{\infty} \ni f \mapsto \theta_m(f\cdot h_m) \in D_{\mathbb{R}}^{m}.$$

Define a Skorokhod metric d_∞° on $D_{\mathbb{R}}^{\infty}$ by

$$d_\infty^{\circ}(f,g) := \sum_{m=0}^{\infty} 2^{-m}\Big(1 \wedge d_m^{\circ}\big(\psi_m(f),\psi_m(g)\big)\Big). \tag{A.5}$$

An equivalent (but incomplete) Skorokhod metric d_∞ can be defined as in (A.5) by replacing the metrics d_m° with d_m (Billingsley, 1999: p. 168). Now define a metric on the product of the spaces $D_{\mathbb{R}}^{m}$, $m \in \mathbb{N}_0$, that is one sets

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

We have the following embedding theorem (Billingsley, 1999: p. 170).

Theorem A.6. *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 (Billingsley, 1999: 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 (Billingsley, 1999: p. 169).

Proposition A.1. *Let* f_n, $n \in \mathbb{N}$, f *be elements of* $D_{\mathbb{R}}^{\infty}$. *Then* $f_n \overset{n\to\infty}{\longrightarrow} f$ *w. r. t.* d_∞° *if and only if* $\theta_t(f_n) \overset{n\to\infty}{\longrightarrow} \theta_t(f)$ *in* $D_{\mathbb{R}}([-r,t\cdot r])$ *for every continuity point* $t\cdot r$ *of* f.

Bibliography

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

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

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

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

A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability in reversible diffusion processes I. Sharp asymptotics for capacities and exit times. Technical Report 767, Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS), 2002a.

A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability in reversible diffusion processes II. Precise asymptotics for small eigenvalues. Technical Report 768, Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS), 2002b.

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

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.

E. B. Dynkin. *Markov Processes*, volume 122 of *Grundlehren der mathematischen Wissenschaften.* Springer-Verlag, Berlin, 1965.

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

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

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

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.

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.

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

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

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

Bjørnar Larssen. Dynamic programming in stochastic control of systems with delay. *Stochastics Stochastics Rep.*, 74(3-4):651–673, 2002.

Robert Lorenz. Weak approximation of stochastic delay differential equations with time series. 2003.

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

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

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

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.

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

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

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

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.

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

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

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