

Metastability of Markovian systems

A transfer operator based approach in
application to molecular dynamics

Vom Fachbereich Mathematik und Informatik
der Freien Universität Berlin
zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften
genehmigte Dissertation

vorgelegt von

Diplom–Mathematiker Wilhelm Huisinga

Berlin, Mai 2001

TO MY LAST BUT ONE MATH TEACHER AT SCHOOL

Betreuer: Prof. Dr. Christof Schütte
Freie Universität Berlin
Fachbereich Mathematik und Informatik
Arnimallee 2-6,
14195 Berlin

Gutachter: Prof. Dr. Christof Schütte
Prof. Dr. Michael Dellnitz (Paderborn)

Datum der Disputation: 07. November 2001

Contents

Introduction	1
1 Modeling Conformational Dynamics	6
1.1 Thermodynamics and Biomolecular Conformations	6
1.2 Single System Dynamics and Markov Processes	8
1.3 Ensemble Dynamics and Transfer Operators	9
2 The Model Systems	13
2.1 Deterministic Hamiltonian System	13
2.2 Hamiltonian System with Randomized Momenta	14
2.3 Langevin Equation	15
2.4 Smoluchowski Equation	17
3 Metastability	21
3.1 Characterizing Metastability	21
3.2 Identifying Metastable Subsets	22
3.3 Metastable Subsets and Eigenvalues Close to 1	23
4 Analysis of Transfer Operators	31
4.1 The Spectrum and its Parts	31
4.2 Bounds on the Essential Spectral Radius in $L^1(\mu)$	34
4.3 Peripheral Spectrum and Properties in $L^1(\mu)$	39
4.4 Reversibility and Properties in $L^2(\mu)$	45
5 Discretization of Transfer Operators	50
5.1 Galerkin Discretization	50
5.2 Convergence of Discrete Eigenvalues	52
5.3 Evaluating the Stochastic Transition Matrix	53
5.4 The Numerical Identification Algorithm	55
6 Theoretical and Numerical Investigations	56
6.1 Deterministic Hamiltonian System	57
6.2 Hamiltonian System with Randomized Momenta	57
6.3 Langevin Equation	61
6.4 Smoluchowski Equation	68
6.5 Comparison of Model Systems	71
7 Application to Large Systems	73
7.1 Monte Carlo methods	73
7.2 Adaptive Discretization Techniques	73
7.3 Analyzing a Small Biomolecule	75
Summary	78

Introduction

While computer power is becoming ever more increasing, there are many problems in physics, chemistry, biology—like climate modeling, protein folding etc.—where the length and time scales of interest remain entirely beyond the capacity currently available, and will remain out of reach in the foreseeable future. As a consequence, there is an increasing need for simplified, reduced-order descriptions. Reduced-order models may provide insight and numerical simulations for larger length scales and longer time scales, but of course at the cost of discarding some level of detail. Instead of simply neglecting some degrees of freedom, one is rather interested in reduced models that incorporate into their dynamical behavior the effective influence of the neglected coordinates. There are two basic approaches for the reduction of complexity, one is based on *elimination* like, e.g., elimination of fast degrees of freedom [3], or stochastic approaches like Brownian and Langevin models [87], and the other one is based on *remodeling* like, e.g., base pair and rod models [58] in the biomolecular context. The design of new reduced model systems is a growing field of research. In order to verify the approximation quality of reduced models or even to automatically construct reduced model systems the numerical approximation of essential features of dynamical systems becomes an important task.

Recently, efficient techniques for the numerical approximation of the essential statistical behavior of deterministic and stochastic dynamical systems have been developed [13, 14, 69]. They are based on the fact that, when modeling the overall dynamics in terms of some transfer operator, certain features of the dynamics are related to its eigenvalues on and close to the unit circle, and can be identified by exploiting the corresponding eigenfunctions. A theorem based strategy has first been proposed by Dellnitz and Junge [13] to analyze almost invariant subsets, attractors and (almost) cyclic behavior of discrete deterministic dynamical systems subject to small additive noise. It has been successfully applied to examine metastable behavior of deterministic Hamiltonian systems by Deuffhard et al. [14]. Although the numerical results of the latter approach were intriguing and seemed to catch the essential features of the molecular system, the deterministic Hamiltonian model appeared to be unsatisfactory for both theoretical discrepancies and computational complexity [72]. Guided by concepts of statistical physics and Monte Carlo techniques, Schütte et al. introduced in [68, 69] a substantially remodeled stochastic Hamiltonian model. It is based on a special discrete-time Markov process that can be understood as a Hamiltonian systems with randomized momenta. Its reliable application to biomolecular systems is demonstrated in [35, 69].

Modeling, theory and numerics presented herein are motivated by the

successful study of biomolecular systems within the stochastic Hamiltonian model and the desire to extend the concepts to a broader class of systems. It is based on a series of preceding studies, where we investigated in detail the Hamiltonian system with randomized momenta [35, 72] and subsequently extended the approach to other molecular systems [34, 71, 72]. This thesis presents a unified and extended transfer operator based approach to metastability of general Markovian systems. It addresses the question of how modeling, theory and algorithmic aspects should be generalized from the weakly perturbed deterministic and from the Hamiltonian setting to the class of Markovian systems. We contribute a profound analysis of metastability and a theoretical justification of the algorithmic strategy for the identification of metastable subsets. This is achieved by combining results on Markov processes by Meyn & Tweedie [52], Markov operators by Lasota & Mackey [46], Markov semigroups by Davies [9, 10] and Singleton [73, 74], and (weakly) compact operators by Weis [79, 80, 81, 82, 83]. As a consequence, we are able to establish new links between spectral properties of transfer operators and well-established Doeblin and ergodicity conditions on Markov processes and operators. This turns out to be particularly advantageous when aiming at a theoretical justification of the algorithmic approach for new model systems, as we consider herein. This thesis investigates for the first time the essential statistical behavior of the Langevin and the Smoluchowski equation in comparison with the Hamiltonian systems with randomized momenta. All in all, the transfer operator based approach to metastability has proven to be very powerful. Its application to the small biomolecule r(ACC), in comparison with other techniques to study biomolecular conformations, is documented in [35].

Most applications of molecular dynamics are in the context of thermodynamics, not only because most experiments measure thermodynamic quantities, but also since most biomolecular processes can only be understood within a thermodynamical context. In the macroscopic theory of equilibrium thermodynamics, the so-called canonical ensemble describes the distribution of microscopic systems under the condition of constant temperature, volume and number of particles [30]. The canonical ensemble is stationary and hence does not change in time. But at the same time each microscopic single system evolves in time, causing internal fluctuations within the ensemble. The characterization and identification of the most relevant fluctuations is of main interest. From a biochemical point of view, these fluctuations are related to the conformational dynamics of a biomolecule. In this setting, a conformation describes a metastable global state of the molecule, in which the large scale geometric structure is understood to be conserved, whereas on smaller scales the molecule may well vibrate, oscillate or deform (see Figure 1). As a consequence, we model conformations as metastable subsets

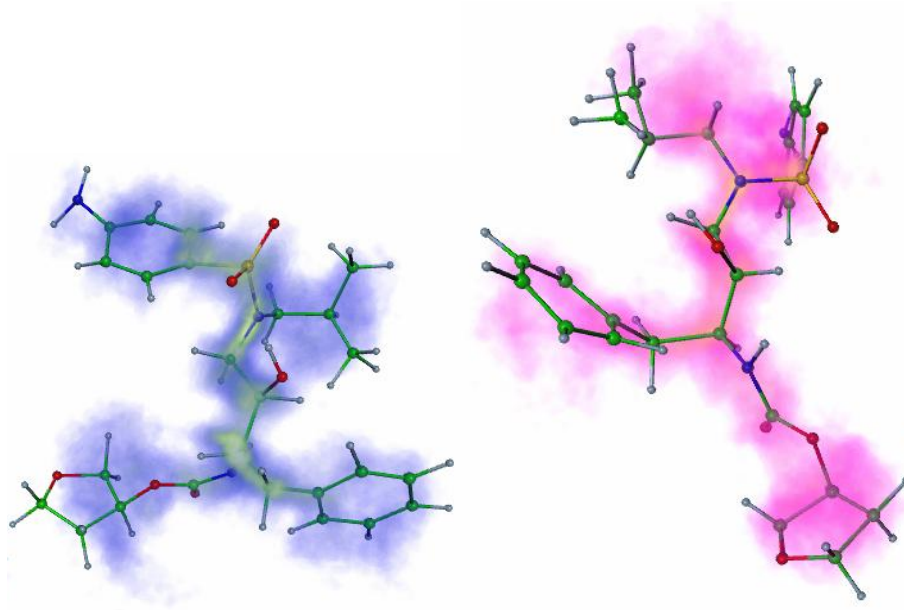


Figure 1: Visualization of two different conformations of a small biomolecule. The so-called ball and stick representations (configurations) correspond to two different conformations of the molecule. The surrounding densities indicate the flexibility within each conformation. Visualization by amira [42].

of the state space and hence include geometric as well as dynamical properties of the system. Both, conformations and their dynamics are expected to gain further insight into the nature of biomolecules and their influence in biochemical reactions.

From a theoretical point of view, the analysis of metastability permits a statistical description of the essential behavior of dynamical systems. The identification of metastable subsets is based on the following idea: Describe internal fluctuations within the invariant distribution by means of a transfer operator defined in terms of the dynamical system. Then the state space can be decomposed into metastable subsets and the essential statistical behavior can be identified by exploiting eigenfunctions of the transfer operator corresponding to eigenvalues close to 1 [13, 17, 69]. Following [72], we give a new theoretical justification of the algorithmic approach in terms of a simple and intriguing relation between the existence of metastable subsets and eigenvalues close to 1 (see Theorem 3.1). We want to emphasize that metastability, as considered herein, is defined w.r.t. some fixed invariant distribution, which in the biomolecular application context is given by the canonical ensemble. This might differ from other approaches to metastability, e.g., the approach via exit times.

The identification strategy requires two particular conditions on the

transfer operator in order to be theoretically justifiable and numerically applicable. Stated in terms of the spectrum of the transfer operator these are (i) the essential spectral radius is less than 1 and (ii) the eigenvalue 1 is simple and dominant (see Sec. 3.2 for a discussion). Within the stochastic Hamiltonian context, Schütte proved in [68] that these two conditions can be reduced to a property of the Hamiltonian flow, which he called momentum–invertibility, and some mixing condition on the dynamics. Having in mind a generalization to Markovian systems, we have to look for alternative conditions independent of particular properties of special model systems. Here the so–called stochastic transition function—a family of probability measures—will play a key role. On the one hand it uniquely characterizes the Markov process representing the microscopic dynamics of single systems. On the other hand it defines the transfer operator modeling the macroscopic evolution of ensembles. In a first step, we are going to relate the two spectral conditions—which are purely functional analytically—to properties of the stochastic transition function. Then, in the second step, these properties will be transformed into more probabilistic conditions on the transfer operator or the Markov process. This enables us to exploit the rich and powerful literature on Markov operators (e.g., Lasota & Mackey [46]) and Markov processes (e.g., Meyn & Tweedie [52]). As a result, we combine results from either of the mathematical theories like, e.g., the fundamental Theorem 4.13, which relates a bound on the essential spectral radius, uniform constrictiveness and the Doeblin–condition.

When dealing with transfer operators, we have to specify the space of functions, the operator is regarded to act on. In the stochastic Hamiltonian approach [68] Schütte considered a weighted Hilbert space of square integrable functions L^2 . For the general Markovian setting, however, the natural space is a weighted Banach space of integrable functions L^1 that includes all probability densities on the state space. Thus, a particular emphasis lies on a detailed analysis of transfer operators acting on L^1 . Nevertheless, we do study transfer operators on L^2 for the special class of reversible Markov processes. Reversibility describes the property that the Markov process and its time–reversed counterpart are statistically the same. It has the advantageous consequence that then the transfer operator is self–adjoint in L^2 .

Within this extended transfer operator based approach to metastability, we analyze theoretically as well as numerically four Markovian systems for molecular dynamics: the deterministic Hamiltonian system, the Hamiltonian system with randomized momenta, the Langevin and the Smoluchowski equation. For the first time we investigate the essential statistical behavior of the Langevin and the Smoluchowski equation in comparison with the Hamiltonian system with randomized momenta. The numerical results give detailed insight into the model systems and prove the transfer operator based

approach to metastability as very powerful.

Acknowledgment. It is a pleasure to thank all those, who supported and encouraged me during the last three and a half years of research. First of all, Christof Schütte and Peter Deuffhard for offering a stimulating research atmosphere, continuous support and for their immense patience. It is their merit to have guided me towards the fascinating field of interdisciplinary research. Many thanks go to the staff at Konrad-Zuse-Zentrum (ZIB), where I spent the first two and a half years during my Ph.D. work. I enjoyed the many discussions, helpful comments and mathematical criticism, whose originated many times from a small cup of espresso in the kitchen. In particular, I would like to thank Frank Cordes, Tobias Galliat, Peter Nettesheim, Daniel Runge, Johannes Schmidt-Ehrenberg and Lin Zschiedrich. At the mathematical institute of the Free University Berlin, where I moved to a year ago, I appreciated the stimulating mixture of teaching and research. Special thanks to Alexander Fischer for fruitful discussions and questioning theory. Finally, I would like to thank Andrew Stuart for illuminating discussions about Markov processes and Dirk Werner for functional analytical advise. Last but for sure not least, I would like to thank Katja and Rosa for simply being there and showing how wonderful live can be.

1 Modeling Conformational Dynamics

The chemically interesting function of many important biomolecules, like proteins or enzymes, results from their dynamical properties, particularly from their ability to undergo so-called conformational transitions [86]. The term **conformations** describes metastable global states of the molecule, in which the large scale geometric structure is understood to be conserved. While on the smallest time scale of about femto- to picoseconds molecular dynamics consists of fast oscillations or fluctuations around equilibrium positions, conformational transitions show up only on a nano- or millisecond time scale and are therefore rare events.

The classical description of molecular processes deals with the microscopic configuration of a molecule, i.e., the positions q and momenta p of all atoms, and leads to a mathematical model in terms of coupled equations of motion for all atoms in the systems (see model systems in Section 2). However, most applications of molecular dynamics are in the context of thermodynamic, not only because most experiments measure thermodynamic quantities, but also since most biomolecular processes can only be understood within a thermodynamical context.

1.1 Thermodynamics and Biomolecular Conformations

Most experiments on biomolecular systems are performed under the conditions of constant temperature T and volume. In equilibrium thermodynamics, the corresponding stationary ensemble is known as the canonical ensemble, whose density w.r.t. the Lebesgue measure dx we denote by f_{can} . To give an explicit formula of f_{can} , we introduce the **Hamiltonian function**

$$H(q, p) = \frac{1}{2} p^T M^{-1} p + V(q), \quad (1)$$

which denotes the internal energy of some single molecular system in state $x = (q, p)$. Here $V : \mathbf{R}^d \rightarrow \mathbf{R}$ is a differentiable potential energy function describing all interactions between the atoms, and M denotes the mass matrix¹. The phase space of a single molecular system is given by $\mathbf{X} \subset \mathbf{R}^{6N}$, where N is the number of atoms. In most cases, it has the simpler form $\Gamma = \Omega \times \mathbf{R}^{3N}$, where Γ is called the phase space and $\Omega \subset \mathbf{R}^{3N}$ is called the position space. Within this setting, the **canonical density** f_{can} associated with the Hamiltonian H is defined as

$$f_{\text{can}}(x) = \frac{1}{Z} \exp(-\beta H(x)), \quad (2)$$

where $Z = \int_{\Gamma} \exp(-\beta H(x)) dx$ denotes the partition function, $\beta = 1/(k_B T)$ the inverse temperature and k_B Boltzmann's constant. Since H separates

¹For simplicity, we assume in the following that M is the identity matrix.

into the sum of two parts depending either only on the momenta or only on the positions, the canonical density factorizes into a product of two densities \mathcal{P} and \mathcal{Q} depending on the momenta and the positions, respectively, only:

$$f_{\text{can}}(x) = \underbrace{\frac{1}{Z_p} \exp\left(-\frac{\beta}{2} p^T M^{-1} p\right)}_{\mathcal{P}(p)} \underbrace{\frac{1}{Z_q} \exp(-\beta V(q))}_{\mathcal{Q}(q)}. \quad (3)$$

For later reference, we denote by μ_{can} and $\mu_{\mathcal{Q}}$ the probability measures induced by the densities f_{can} and \mathcal{Q} , respectively.

Typically, metastability w.r.t. the canonical ensemble is measured via the following two-step experiment:

1. Pre-Selection: Select from the canonical ensemble all such systems with states $x \in C$, where the subset C corresponds to some (measurable) physical property. This selection prepares a **sub-ensemble** μ_C . Physically the sub-ensemble is associated with the property C ; mathematically it is associated with the subset C of the state space.
2. Transition-Counting: Fix some observation time span $\tau > 0$ and determine the relative frequency of systems within the sub-ensemble μ_C that stay in C after the time τ .

A sub-ensemble μ_C will be called **metastable** (on the timescale τ), if the fraction of systems in the sub-ensemble that stays in C after the time τ is close to 1. Hence, metastability depends on C and τ . Identifying metastable sub-ensembles mathematically necessitates the description of internal fluctuation within the canonical ensemble.

Since equilibrium thermodynamics states nothing about the microscopic dynamics of single systems within the ensemble, we have to specify some microscopic dynamics. We will see in the next section that there is no “canonical” single system dynamics but several [31].

We will restrict our considerations to the broad class of microscopic dynamics that can be interpreted as Markov processes. This will enable us to describe internal fluctuations within the stationary ensemble by studying transfer operators induced by the Markov process, as we are going to outline in the next sections.

In the introduction we have characterized conformations as metastable large scale geometric structures. Hence, conformations are thought to be objects in the position space. However, both the canonical ensemble as well as the classical models for microscopic dynamics presented below are

defined in the phase space Γ (positions and momenta). Therefore, we have to specify the relation between metastable sub-ensembles defined in Γ and conformations, characterized in terms of the position space Ω only. In [68] the following relation is proposed:

A conformation $C \subset \Omega$ will be identified with the particular metastable sub-ensemble $\mu_{C \times \mathbf{R}^{3N}}$ corresponding to the particular subset $C \times \mathbf{R}^{3N} \subset \Gamma$. Hence, for every position $q \in C$, the conformation contains all states with $q \in \Omega$ and arbitrary $p \in \mathbf{R}^{3N}$.

In this sense, conformations contain no information on momenta, and are determined within the positions space only. With this characterization of conformations in mind, we could think of reduced microscopic models defined only in the position space Ω . Such simplified models indeed exist and will be presented below. They allow to describe internal fluctuations within the positional canonical density \mathcal{Q} . In agreement with the relation above, metastable sub-ensembles of reduced models are called conformations, too.

1.2 Single System Dynamics and Markov Processes

This section gives a brief mathematical description of Markovian systems. For a detailed introduction see, e.g., [18, 52, 55].

Consider the state space $\mathbf{X} \subset \mathbf{R}^m$ for some $m \in \mathbf{Z}_+$ equipped with the Borel σ -algebra \mathcal{A} on \mathbf{X} . The evolution of a single microscopic system is supposed to be given by a **homogeneous Markov process** $X_t = \{X_t\}_{t \in \mathbf{T}}$ in continuous or discrete time with $\mathbf{T} = \mathbf{R}_+$ or $\mathbf{T} = \mathbf{Z}_+$, respectively. We write $X_0 \sim \mu$, if the Markov process X_t is initially distributed according to μ , i.e., if $\mathbf{P}[X_0 \in A] = \mu(A)$ for every $A \in \mathcal{A}$, and $X_0 \sim x$ if $\mu = \delta_x$ for the dirac measure at x . The motion of X_t is given in terms of a **stochastic transition function** p according to

$$p(t, x, A) = \mathbf{P}[X_{t+s} \in A | X_s = x], \quad (4)$$

for every $t, s \in \mathbf{T}$, $x \in \mathbf{X}$ and $A \in \mathcal{A}$. The map $p : \mathbf{T} \times \mathbf{X} \times \mathcal{B}(\mathbf{X}) \rightarrow [0, 1]$ has the following properties

- (i) $x \mapsto p(t, x, A)$ is measurable for every $t \in \mathbf{T}$ and $A \in \mathcal{B}(\mathbf{X})$,
- (ii) $A \mapsto p(t, x, A)$ is a probability measure for every $t \in \mathbf{T}$ and $x \in \mathbf{X}$.
- (iii) $p(0, x, \mathbf{X} \setminus \{x\}) = 0$ for every $x \in \mathbf{X}$.
- (iv) the Chapman-Kolmogorov equation

$$p(t + s, x, A) = \int_{\mathbf{X}} p(t, x, dz) p(s, z, A) \quad (5)$$

holds for every $t, s \in \mathbf{T}$, $x \in \mathbf{X}$ and $A \in \mathcal{A}$.

The relation between Markov processes and stochastic transition functions is one-to-one, i.e., every homogeneous Markov process defines a stochastic transition function satisfying properties (i) to (iv), and vice versa [52, Chapter 3]. We say that the Markov process X_t admits an **invariant probability measure** μ , or μ is invariant w.r.t. the Markov process, if

$$\int_{\mathbf{X}} p(t, x, A) \mu(dx) = \mu(A) \quad (6)$$

for every $t \in \mathbf{T}$ and $A \in \mathcal{A}$ [52, Chapter 10]. Note that the invariant probability measure needs not be unique. A Markov process X_t is called **reversible** w.r.t. an invariant probability measure μ , if

$$\int_A p(t, x, B) \mu(dx) = \int_B p(t, x, A) \mu(dx) \quad (7)$$

for every $t \in \mathbf{T}$ and $A, B \in \mathcal{A}$. If μ is unique, X_t is simply called reversible. For the special case of a stochastic transition function being absolutely continuous w.r.t. μ , the Markov process X_t is reversible, if $p(t, x, y) = p(t, y, x)$ for every $t \in \mathbf{T}$ and μ -a.e. $x, y \in \mathbf{X}$.

1.3 Ensemble Dynamics and Transfer Operators

Based on the assumption that the microscopic dynamics is given by a homogeneous Markov process we are now able to introduce a Markov operator that allows to describe internal fluctuations within the stationary ensemble.

The basic idea is the following: Consider all systems within the stationary ensemble μ with states in some subset $C \in \mathcal{A}$ (see pre-selection step in Section 1.1). This sub-ensemble of systems is distributed according to the probability measure

$$\nu_0(A) = \frac{1}{\mu(C)} \int_A \mathbf{1}_C(x) \mu(dx),$$

where $\mathbf{1}_C$ denotes the characteristic function of the subset C . In other words, the sub-ensemble corresponds to the density $\hat{\mathbf{1}}_C = \mathbf{1}_C / \mu(C)$ w.r.t. to μ . Since every single microscopic system evolves according to the dynamics of the Markov process, the distribution of the sub-ensemble at time $t \in \mathbf{T}$ is given by the probability measure

$$\begin{aligned} \nu_t(A) &= \int_{\mathbf{X}} \mathbf{P}[X_t \in A | X_0 = x] \nu_0(dx) \\ &= \int_{\mathbf{X}} \hat{\mathbf{1}}_C(x) p(t, x, A) \mu(dx). \end{aligned} \quad (8)$$

Therefore, the relative frequencies of systems within the sub-ensemble that stay in C after the observation time span τ (see transition counting step in

Section 1.1) is determined by $\nu_\tau(C)$. Note that (8) for the special choice of $C = \mathbf{X}$, and hence $\nu_0 = \mu$ becomes $\nu_t(A) = \int_{\mathbf{X}} p(t, x, A) \mu(dx)$. Since the canonical ensemble μ is by assumption stationary, we have to require that $\nu_t = \mu$ for every $t \in \mathbf{T}$ and obtain

$$\int_{\mathbf{X}} p(t, x, A) \mu(dx) = \mu(A) \quad (9)$$

for every $t \in \mathbf{T}$ and $A \in \mathcal{A}$. **Thus, we will henceforth assume that the probability measure μ is invariant w.r.t. the Markov process X_t .**

Our interest is to define an operator P_t that propagates sub-ensembles in time, thus allows to describe the evolution of ν_t . Since invariance of μ implies that $\nu_t \ll \mu$ whenever $\nu_0 \ll \mu$ [60, Chapter 4], we consider the operator on the space of measures that are absolutely continuous w.r.t. μ —or equivalently, on the space of densities w.r.t. μ —rather than on the space of arbitrary probability measures. To do so, we introduce the Banach spaces of equivalence classes of measurable functions

$$L^r(\mu) = \left\{ u : \mathbf{X} \rightarrow \mathbf{C} : \int_{\mathbf{X}} |u(x)|^r \mu(dx) < \infty \right\}. \quad (10)$$

for $1 \leq r < \infty$ and

$$L^\infty(\mu) = \left\{ u : \mathbf{X} \rightarrow \mathbf{C} : \mu\text{-ess sup}_{x \in \mathbf{X}} |u(x)| < \infty \right\}$$

with corresponding norms $\|\cdot\|_r$ and $\|\cdot\|_\infty$, respectively. Due to Hölder's inequality we have $L^r(\mu) \subset L^s(\mu)$ for every $1 \leq s \leq r \leq \infty$. Based on the initially given motivation, we define the **semigroup of propagators** or forward transfer operators $P_t : L^1(\mu) \rightarrow L^1(\mu)$ with $t \in \mathbf{T}$ according to

$$\int_A P_t v(y) \mu(dy) = \int_{\mathbf{X}} p(t, x, A) v(x) \mu(dx) \quad (11)$$

for $A \in \mathcal{A}$. As a consequence of the invariance of μ , the characteristic function $\mathbf{1}_{\mathbf{X}}$ of the entire space is an invariant density of P_t , i.e., $P_t \mathbf{1}_{\mathbf{X}} = \mathbf{1}_{\mathbf{X}}$. Furthermore, P_t is a Markov operator, i.e., P_t conserves norm: $\|P_t v\|_1 = \|v\|_1$ and positivity: $P_t v \geq 0$ if $v \geq 0$, which is a simple consequence of the definition. Using the propagator in the context of the initially given motivation, we see that the sub-ensemble ν_t , originating from ν_0 at time $t = 0$, is distributed according to

$$\nu_t(A) = \int_{\mathbf{X}} P_t \hat{\mathbf{1}}_A(x) \mu(dx).$$

Consequently, $P_t \hat{\mathbf{1}}_A$ is the density of ν_t w.r.t. μ . The semigroup of propagators mathematically models the physical phenomena of evolution of sub-ensembles in time.

In the theory of Markov processes a different semigroup of operators is considered. We will call it² the **semigroup of backward transfer operators** $T_t : L^\infty(\mu) \rightarrow L^\infty(\mu)$ with $t \in \mathbf{T}$, defined by

$$T_t u(x) = \mathbf{E}_x[u(X_t)] = \int_{\mathbf{X}} u(y)p(t, x, dy). \quad (12)$$

As a consequence of property (ii) of the stochastic transition function, we have $T_t \mathbf{1}_{\mathbf{X}} = \mathbf{1}_{\mathbf{X}}$ for every $t \in \mathbf{T}$. Both operators are closely related via the duality bracket

$$\langle v, u \rangle_\mu = \int_{\mathbf{X}} v(x)u(x)\mu(dx)$$

for $v \in L^1(\mu)$ and $u \in L^\infty(\mu)$, namely $\langle P_t v, u \rangle = \langle v, T_t u \rangle$. Thus, the backward transfer operator is the adjoint of the propagator: $P_t^* = T_t$. Since $L^1(\mu)$ is a proper subset of the dual of $L^\infty(\mu)$, we have $P_t \not\subseteq T_t^*$, hence P_t is not the adjoint of T_t . As a consequence, it is much easier to relate properties of P_t to T_t than vice versa. If the Markov process corresponds to a deterministic dynamical system, the propagator and the backward transfer operator are known as the Frobenius–Perron and the Koopman operator, respectively [46].

Propagators associated with reversible Markov processes are of particular interest, since they possess additional structure on the Hilbert space $L^2(\mu)$. Such propagators will be called reversible, too. Note that the propagator is well-defined on any Banach space $L^r(\mu)$ for $1 \leq r < \infty$ with $\|P\|_r \leq 1$, see [62] and cited reference.

Proposition 1.1 *Let $P_t : L^2(\mu) \subset L^1(\mu) \rightarrow L^2(\mu)$ denote the propagator corresponding to the Markov process X_t . Then P_t is self-adjoint w.r.t. the scalar product $\langle \cdot, \cdot \rangle_\mu$ in $L^2(\mu)$, i.e.,*

$$\langle u, P_t v \rangle_\mu = \langle P_t u, v \rangle_\mu; \quad t \in \mathbf{T}$$

for every $u, v \in L^2(\mu)$, if and only if X_t is reversible.

Proof: For $u, v \in L^2(\mu)$ we have

$$\begin{aligned} \langle u, P_t v \rangle_\mu &= \int_{\mathbf{X}} u(y) \overline{P_t v(y)} \mu(dy) = \int_{\mathbf{X}} \int_{\mathbf{X}} u(y) \overline{v(x)p(t, x, dy)} \mu(dx) \\ &= \int_{\mathbf{X}} \int_{\mathbf{X}} u(y) \overline{v(x)} p(t, y, dx) \mu(dy) = \int_{\mathbf{X}} P_t u(x) \overline{v(x)} \mu(dx) \\ &= \langle P_t u, v \rangle_\mu, \end{aligned}$$

²The nomenclature is motivated by the fact that the forward transfer operator is for some model systems related to a reweighted version of the forward Kolmogorov equation, while the backward transfer operator is related to the backward Kolmogorov equation (see the Langevin and the Smoluchowski equation in Sections 2.3 and 2.4, respectively).

where the first identity in the second line is due to reversibility of the Markov process. The reversed implication is just a rearrangement of the calculation. \square

We want to close this section with a remark about the mathematical model used to describe internal fluctuations within the canonical ensemble. Physical experiments on molecular ensembles allow to measure relative frequencies in the canonical ensemble μ . Assuming, as in our case, that μ has the form

$$\mu(dx) = f(x)dx,$$

i.e., is absolutely continuous w.r.t. the Lebesgue measure dx , physical experiments correspond to the densities of the form

$$v_{\text{phys}}(x) = \hat{\mathbf{1}}_C(x)f(x)$$

w.r.t. the Lebesgue measure dx . Whenever physicists use the phrase “probability density” they refer to v_{phys} rather than to densities

$$v_{\text{math}}(x) = \hat{\mathbf{1}}_C(x)$$

w.r.t. probability measure μ , as we do. As will become apparent later, it is mathematically advantageous to consider the semigroup of propagators acting on densities v_{math} rather than on v_{phys} . However, it should be clear that results obtained in either of the two pictures can be transformed into the other.

2 The Model Systems

We introduce four popular models for molecular dynamics: the deterministic Hamiltonian system, the Hamiltonian system with randomized momenta, the Langevin and the Smoluchowski equation. While the deterministic Hamiltonian system and the Langevin equation are classical models on the phase space Γ (positions and momenta), the Hamiltonian system with randomized momenta and the Smoluchowski equation are reduced models acting only on the position space Ω . For each model system we assume that Ω belongs to one of the two fundamentally different cases:

1. **Bounded system:** The position space Ω is unbounded, typically $\Omega = \mathbf{R}^{3N}$, and the potential energy function V is smooth, bounded from below, and satisfies $V \rightarrow \infty$ for $|q| \rightarrow \infty$. Such systems are called bounded, since the energy surfaces $\{(q, p) \in \Gamma : H(q, p) = E\}$ are bounded subsets of Γ for every energy E .
2. **Periodic systems:** The position space Ω is some $3N$ -dimensional torus and the potential energy function V is continuous on Ω and thus bounded. There is an intensive discussion concerning the question of whether V can also be assumed to be smooth as we will do herein, see [68, Sec. 2] for details.

Both cases are typical for molecular dynamics applications. Periodic systems in particular include the assumption of periodic boundaries, which is by far the most popular modeling assumption for biomolecular systems.

2.1 Deterministic Hamiltonian System

The **deterministic Hamiltonian system**

$$\dot{q} = p, \quad \dot{p} = -\nabla_q V(q). \quad (13)$$

defined on the state space $\mathbf{X} = \Gamma$ models an energetically closed system, whose total energy is given by the Hamiltonian H as defined in (1); its energy is conserved under the dynamics. The Markov process $X_t = \{X_t\}_{t \in \mathbf{R}_+}$ defined by the deterministic Hamiltonian system coincides with the flow Φ^τ associated with (13); hence $X_t = \Phi^t x_0$ for the initial distribution $X_0 \sim x_0$. This allows us to denote the stochastic transition function as

$$p(t, x, C) = \mathbf{1}_C(\Phi^t x) = \delta_{\Phi^t x}(C) \quad (14)$$

for every $t \in \mathbf{R}_+$ and $C \in \mathcal{A}$. It is well known that canonical ensemble μ_{can} is invariant w.r.t. X_t . The evolution of densities $v = v(x, t)$ w.r.t. μ_{can} is governed by the **Liouville equation**

$$\partial_t v = \left(\underbrace{-p \cdot \nabla_q + \nabla_q V \cdot \nabla_p}_{i\mathcal{L}} \right) v, \quad (15)$$

where \mathcal{L} denotes the Liouville operator defined on some suitable subspace of $L^1(\mu_{\text{can}})$. Since the solution of (15) satisfies $v(x, t + s) = v(\Phi^{-t}x, s)$ for every $t, s \in \mathbf{R}_+$, the semigroup of propagators $P_t : L^1(\mu_{\text{can}}) \rightarrow L^1(\mu_{\text{can}})$ with $t \in \mathbf{R}_+$ is defined by

$$P_t v(x) = \exp(it\mathcal{L})v(x) = v(\Phi^{-t}x). \quad (16)$$

This is exactly the definition of the Frobenius–Perron operator corresponding to the Hamiltonian flow Φ^t [46]. Since \mathcal{L} is self-adjoint w.r.t. the scalar product in $L^2(\mu)$ the operator P_t is unitary in $L^2(\mu)$. The semigroup of backward transfer operators $T_t : L^\infty(\mu) \rightarrow L^\infty(\mu)$ with $t \in \mathbf{R}_+$ is given by

$$T_t u(x) = u(\Phi^t x), \quad (17)$$

which is identical with the Koopman operator corresponding to Φ^t [46].

2.2 Hamiltonian System with Randomized Momenta

Aiming at a conformational analysis of biomolecular systems, Schütte introduced in [68] some kind of stochastic Hamiltonian system. It is a reduced dynamics defined solely on the position space and derived from the deterministic Hamiltonian system by “randomizing the momenta” and integrating for some fixed observation time span τ .

Let us briefly sketch the derivation. Fix some observation time span $\tau > 0$ and denote by $p_\Gamma(\tau, x, A)$ the stochastic transition function corresponding to the full deterministic Hamiltonian system on Γ (for a comment on the time τ see remark below). In view of the relation between metastable sub-ensembles and conformations, we are interested in a simplified model describing the dynamics between “cylindric” subsets $B \times \mathbf{R}^d$ and $C \times \mathbf{R}^d$. Inserting these special subsets into definition (14) yields

$$\begin{aligned} & p_\Gamma(\tau, B \times \mathbf{R}^d, C \times \mathbf{R}^d) \\ &= \frac{1}{\int_{B \times \mathbf{R}^d} \mu_{\text{can}}(dx)} \int_{B \times \mathbf{R}^d} \mathbf{1}_{C \times \mathbf{R}^d}(\Phi^\tau(x)) \mu_{\text{can}}(dx) \\ &= \frac{1}{\int_B \mu_{\mathcal{Q}}(dq)} \int_B \underbrace{\int_{\mathbf{R}^d} \mathbf{1}_C(\Pi_q \Phi^\tau(q, p)) \mu_{\mathcal{P}}(dp)}_{p_{\bar{\Omega}}(1, q, C)} \mu_{\mathcal{Q}}(dq). \quad (18) \\ &= p_{\bar{\Omega}}^\tau(1, B, C), \end{aligned}$$

where $\Pi_q : \Gamma \rightarrow \Omega$ denotes the projection onto the position space. Equation (18) defines a *one-step* stochastic transition function, whose *n-step* version is determined via the Chapman–Kolmogorov equation (5). The associated discrete time Markov process $Q_n = \{Q_n\}_{n \in \mathbf{Z}_+}$, defined on the state

space $\mathbf{X} = \Omega$, solves the **Hamiltonian system with randomized momenta** [68]

$$Q_{n+1} = \Pi_q \Phi^\tau(Q_n, P_n); \quad n \in \mathbf{Z}_+ \quad (19)$$

where P_n is chosen randomly from the canonical distribution of momenta \mathcal{P} , as defined in (3). As shown in [68] the positional canonical ensemble μ_Q is invariant w.r.t. the Markov process Q_n . The semigroup of propagators $P_n : L^1(\mu_Q) \rightarrow L^1(\mu_Q)$ for $n \in \mathbf{Z}_+$ is given by $P_n = (P_1)^n$ with

$$P_1 v(q) = \int_{\mathbf{R}^d} v(\Pi_q \Phi^{-\tau}(q, p)) \mathcal{P}(p) dp. \quad (20)$$

Exploiting that μ_Q is invariant and Φ^τ is reversible and symplectic, it is shown in [68] that P_1 , and thus the semigroup, is self-adjoint in $L^2(\mu_Q)$.

Remark. For arbitrary, but fixed $\tau > 0$ we have defined in (19) the *one-step* transition function $p^\tau(1, q, D)$. Changing the observation time to $\sigma > 0$ results in a new one-step transition function $p^\sigma(1, q, D)$. In general we will have $p^{2\tau}(1, q, D) \neq p^\tau(2, q, D)$ and, consequently, $P_1^{2\tau} \neq P_2^\tau$, where the superscripts indicate the corresponding observation time spans (for an example see [68, Sec. 3.7.1]). In terms of the Hamiltonian system with randomized momenta, this is not surprising, since $P_1^{2\tau}$ includes only one choice of momenta according to \mathcal{P} , while P_2^τ does include two.

2.3 Langevin Equation

The most popular model for an open system stochastically interacting with its environment is the **Langevin equation**³ [61]

$$\dot{q} = p, \quad \dot{p} = -\nabla_q V(q) - \gamma p + \sigma \dot{W} \quad (21)$$

corresponding to some friction constant $\gamma > 0$ and external force $F_{\text{ext}} = \sigma \dot{W}$ given by a standard $3N$ -dimensional Brownian motion W . Eq. (21) defines a continuous time Markov process $X_t = \{(Q_t, P_t)\}_{t \in \mathbf{T}}$ on the state space $\mathbf{X} = \Gamma$.

In the Langevin model, the effects of solvent molecules not explicitly present in the system being simulated are approximated in terms of a frictional drag on the solute as well as random collisions associated with the thermal motion of the solvent molecules. The Hamiltonian H describes the internal energy of the systems, which is not conserved due to energy

³In our context, the notion \dot{W} is a convenient form of the more common dW . Hence, the Langevin equation (21) should be understood as $dq = p dt$ and $dp = -\nabla_q V(q) dt - \gamma p dt + \sigma dW$, which moreover is just the common abbreviation for the corresponding integral notion, see e.g., [57]. For convenience we will henceforth use the "dot" without further comments.

transfer with the surrounding, the so-called heat bath. Yet, the interplay between stochastic excitation and damping equilibrates the internal energy to $\beta = 2\gamma/\sigma^2$. As a result, the canonical ensemble μ_{can} corresponding to the inverse temperature β is invariant w.r.t. the Markov process defined by (21). The evolution of densities $v = v(t, x)$ w.r.t. μ_{can} is governed by the **Fokker–Planck equation**

$$\partial_t v = \left(\underbrace{\frac{\sigma^2}{2} \Delta_p - p \cdot \nabla_q + \nabla_q V \cdot \nabla_p - \gamma p \cdot \nabla_p}_{\mathcal{L}} \right) v \quad (22)$$

regarded on some suitable subspace of $L^1(\mu_{\text{can}})$. Therefore, \mathcal{L} is the infinitesimal generator of the semigroup of propagators $P_t : L^1(\mu_{\text{can}}) \rightarrow L^1(\mu_{\text{can}})$ with $t \in \mathbf{R}_+$ defined by

$$P_t v = \exp(t\mathcal{L})v \quad (23)$$

for $v \in L^1(\mu_{\text{can}})$. In general, P_t is not self-adjoint in $L^2(\mu)$.

Remark. The evolution of some physical density⁴ $v_{\text{phys}} = v f_{\text{can}} \in L^1(dx)$ with $v \in L^1(\mu_{\text{can}})$ is governed by the so-called **forward Kolmogorov equation** $\partial_t v_{\text{phys}} = \mathcal{A}_{\text{fw}} v_{\text{phys}}$ with

$$\mathcal{A}_{\text{fw}} = \frac{\sigma^2}{2} \Delta_p - p \cdot \nabla_q + \nabla_q V \cdot \nabla_p + \gamma p \cdot \nabla_p + \gamma$$

acting on a suitable subspace of $L^1(dx)$ [38, Chapter 5.1]. It permits to define the semigroup of propagators $P_t^{\text{fw}} : L^1(dx) \rightarrow L^1(dx)$ by

$$P_t^{\text{fw}} v_{\text{phys}} = \exp(t\mathcal{A}_{\text{fw}})v_{\text{phys}}.$$

As a consequence of the invariance of μ , we obtain the relation

$$P_t^{\text{fw}}(v f_{\text{can}}) = (P_t v) f_{\text{can}}; \quad v \in L^1(\mu_{\text{can}}) \quad (24)$$

between the two semigroups of propagators. To derive the evolution equation (22) for v , we insert $v_{\text{phys}} = v f_{\text{can}}$ into the forward Kolmogorov equation and obtain after simple manipulations

$$\partial_t(v f_{\text{can}}) = \mathcal{A}_{\text{fw}}(v f_{\text{can}}) = (\mathcal{L}v) f_{\text{can}},$$

which is the infinitesimal version of (24). Exploiting time-independence and positivity of f_{can} , we finally end up with the Fokker–Planck equation (22).

⁴See remark about our mathematical model at the end of Section 1.3.

In some sense dual to the forward Kolmogorov equation is the **backward Kolmogorov equation** $\partial_t u = \mathcal{A}_{\text{bw}} u$ with

$$\mathcal{A}_{\text{bw}} = \frac{\sigma^2}{2} \Delta_p + p \cdot \nabla_q - \nabla_q V \cdot \nabla_p - \gamma p \cdot \nabla_p$$

acting on a suitable subspace of $L^1(dx)$. For bounded and periodic systems, we have $\langle \mathcal{A}_{\text{fw}} v, u \rangle_2 = \langle v, \mathcal{A}_{\text{bw}} u \rangle_2$ on the Hilbert space $L^2(dx)$ and hence \mathcal{A}_{fw} and \mathcal{A}_{bw} are adjoint to each other. Therefore neither \mathcal{A}_{fw} nor \mathcal{A}_{bw} is self-adjoint in $L^2(dx)$. The generator \mathcal{A}_{bw} permits to define the semigroup of backward transfer operators $T_t^{\text{bw}} : L^1(dx) \rightarrow L^1(dx)$ by

$$T_t^{\text{bw}} u(x) = \mathbf{E}_x[u(X_t)] = \exp(t\mathcal{A}_{\text{bw}})u(x),$$

see, e.g., [38, Chapter 5.1]. We remark that although the formal definition of the two semigroups of backward transfer operators T_t and T_t^{bw} via expectation is the same, they differ in the space of functions regarded to act on.

2.4 Smoluchowski Equation

As a second reduced model system, we introduce the Smoluchowski equation. It is derived from the Langevin equation by considering the high friction limit $\gamma \rightarrow \infty$. In contrast to the Langevin equation it defines a reversible Markov process.

Write the Langevin equation (21) in second order form

$$\ddot{q} = -\nabla_q V(q) - \gamma \dot{q} + \sigma \dot{W}. \quad (25)$$

For the high friction limit, we introduce some smallness parameter $\epsilon > 0$ and transform the friction constant to γ/ϵ ; in order to conserve the inverse temperature $\beta = 2\gamma/\sigma^2$ of the surrounding heat bath and hence the canonical ensemble, we simultaneously have to scale the white noise constant to $\sigma/\sqrt{\epsilon}$. This yields

$$\ddot{q} = -\nabla_q V(q) - \frac{\gamma}{\epsilon} \dot{q} + \frac{\sigma}{\sqrt{\epsilon}} \dot{W}.$$

After rescaling the time according to $t \mapsto \epsilon t$ we finally get

$$\epsilon^2 \ddot{q} = -\nabla_q V(q) - \gamma \dot{q} + \sigma \dot{W}.$$

Since the white noise process \dot{W} is unbounded, we cannot simply assume that the acceleration $\epsilon^2 \ddot{q}$ is small for $\epsilon \ll 1$. However, investigations by Nelson [53] show that the solution $q_{\text{Lan}}^\epsilon(t; q_0, p_0)$ of the Langevin equation (26) and the solution $q_{\text{Smol}}(t; q_0)$ of the **Smoluchowski equation**

$$\dot{q} = -\frac{1}{\gamma} \nabla_q V(q) + \frac{\sigma}{\gamma} \dot{W} \quad (26)$$

are close to each other for high friction γ .

Theorem 2.1 [53, Theorem 10.1] *Assume that $\nabla_q V(q)$ is global Lipschitz. Then, for every p_0 , with probability one*

$$\lim_{\epsilon \rightarrow 0} |q_{\text{Smol}}(t; q_0) - q_{\text{Lan}}^\epsilon(t; q_0, p_0)| = 0$$

uniformly for t in compact subintervals of $[0, \infty)$.

Hence, in the high friction case, the Smoluchowski dynamics is a good approximation of the Langevin dynamics. This fact will be analyzed in the following section.

The stochastic differential equation (26) defines a continuous time Markov process $Q_t = \{Q_t\}_{t \in \mathbf{R}_+}$ on the state space $\mathbf{X} = \Omega$ with transition kernel $p = p(t, q, C)$ and invariant probability measure μ_Q [61]. The evolution of densities $v = v(t, q)$ w.r.t. μ_Q is governed by the **Fokker–Planck equation** (see Remark below for derivation)

$$\partial_t v = \left(\underbrace{\frac{\sigma^2}{2\gamma^2} \Delta_q - \frac{1}{\gamma} \nabla_q V(q) \cdot \nabla_q}_{\mathcal{L}} \right) v \quad (27)$$

regarded on some suitable subspace of $L^1(\mu_Q)$. Therefore, \mathcal{L} is the infinitesimal generator of the semigroup of propagators $P_t : L^1(\mu_Q) \rightarrow L^1(\mu_Q)$ with $t \in \mathbf{R}_+$ defined by

$$P_t v = \exp(t\mathcal{L})v \quad (28)$$

for $v \in L^1(\mu)$.

Proposition 2.2 *The semigroup of propagators $P_t : L^2(\mu_Q) \rightarrow L^2(\mu_Q)$ is self-adjoint w.r.t. the scalar product $\langle \cdot, \cdot \rangle_{\mu_Q}$ in $L^2(\mu_Q)$.*

Proof: We prove that \mathcal{L} is self-adjoint in $L^2(\mu_Q)$. The statement of the theorem then follows from [8, Thm. 4.6]. Consider $v, u \in L^2(\mu_Q)$ in the domain of \mathcal{L} . Then $Q^{\frac{1}{2}}v, Q^{\frac{1}{2}}u \in L^2(dq)$ and

$$\langle \mathcal{L}v, u \rangle_{\mu_Q} = \langle (Q^{\frac{1}{2}}\mathcal{L}Q^{-\frac{1}{2}})(Q^{\frac{1}{2}}v), Q^{\frac{1}{2}}u \rangle_2.$$

A simple calculation proves that $Q^{\frac{1}{2}}\mathcal{L}Q^{-\frac{1}{2}} = \mathcal{L}_s$ for the generator \mathcal{L}_s defined in (30) below. Since \mathcal{L}_s is a so-called Schrödinger operator, it is self-adjoint in $L^2(dq)$ and we obtain

$$\langle \mathcal{L}v, u \rangle_{\mu_Q} = \langle Q^{\frac{1}{2}}v, (Q^{\frac{1}{2}}\mathcal{L}Q^{-\frac{1}{2}})(Q^{\frac{1}{2}}u) \rangle_{\mu_Q} = \langle v, \mathcal{L}u \rangle_{\mu_Q},$$

hence \mathcal{L} is self-adjoint in $L^2(\mu_Q)$. \square

There is a strong relation of our approach to the theory of large deviations and first exit times (see, e.g., Freidlin and Wentzell [23]); a brief outline is given in [72]; further investigations are part of a current diploma thesis [49].

*Remark.*⁵ The evolution of some physical density $v_{\text{phys}} = v\mathcal{Q} \in L^1(\text{d}q)$ with $v \in L^1(\mu_{\mathcal{Q}})$ is governed by the so-called **forward Kolmogorov equation** $\partial_t v_{\text{phys}} = \mathcal{A}_{\text{fw}} v_{\text{phys}}$ with

$$\mathcal{A}_{\text{fw}} = \frac{\sigma^2}{2\gamma^2} \Delta_q + \frac{1}{\gamma} \nabla_q V(q) \cdot \nabla_q + \frac{1}{\gamma} \Delta_q V(q)$$

acting on a suitable subspace of $L^1(\text{d}q)$ [38, Chapter 5.1]. It permits to define the semigroup of propagators $P_t^{\text{fw}} : L^1(\text{d}q) \rightarrow L^1(\text{d}q)$ by

$$P_t^{\text{fw}} v_{\text{phys}} = \exp(t\mathcal{A}_{\text{fw}}) v_{\text{phys}}.$$

As a consequence of the invariance of μ , we obtain the relation

$$P_t^{\text{fw}}(v\mathcal{Q}) = (P_t v)\mathcal{Q}; \quad v \in L^1(\mu_{\mathcal{Q}}) \quad (29)$$

between the two semigroups of propagators. To derive the evolution equation (27) for v , we insert $v_{\text{phys}} = v\mathcal{Q}$ into the forward Kolmogorov equation and obtain after simple manipulations

$$\partial_t(v\mathcal{Q}) = \mathcal{A}_{\text{fw}}(v\mathcal{Q}) = (\mathcal{L}v)\mathcal{Q},$$

which is the infinitesimal version of (29). Exploiting time-independence and positivity of \mathcal{Q} , we finally end up with the Fokker-Planck equation (27). In some sense dual to the forward Kolmogorov equation is the **backward Kolmogorov equation** $\partial_t u = \mathcal{A}_{\text{bw}} u$ with

$$\mathcal{A}_{\text{bw}} = \frac{\sigma^2}{2\gamma^2} \Delta_q - \frac{1}{\gamma} \nabla_q V(q) \cdot \nabla_q$$

acting on a suitable subspace of $L^1(\text{d}q)$. Although \mathcal{L} and \mathcal{A}_{bw} are formally the same, they are regarded to act on different function spaces. For bounded and periodic systems, we have $\langle \mathcal{A}_{\text{fw}} v, u \rangle_2 = \langle v, \mathcal{A}_{\text{bw}} u \rangle_2$ on the Hilbert space $L^2(\text{d}q)$ and hence \mathcal{A}_{fw} and \mathcal{A}_{bw} are adjoint to each other. However, neither \mathcal{A}_{fw} nor \mathcal{A}_{bw} are self-adjoint in $L^2(\text{d}q)$, while \mathcal{L} is self-adjoint in $L^2(\mu_{\mathcal{Q}})$. The generator \mathcal{A}_{bw} permits to define the semigroup of backward transfer operators $T_t^{\text{bw}} : L^1(\text{d}q) \rightarrow L^1(\text{d}q)$ according to

$$T_t^{\text{bw}} u(q) = \mathbf{E}_q[u(X_t)] = \exp(t\mathcal{A}_{\text{bw}}) u(q),$$

⁵We are aware of the fact that large parts of this remark are analogous to those made for the Langevin equation and could hence be abbreviated. Yet, for sake of clarity, we prefer to state everything explicitly, since there are also important differences.

see, e.g., [38, Chapter 5.1]. We remark that although the formal definition of the two semigroups of backward transfer operators T_t and T_t^{bw} via expectation is the same, they differ in the space of functions regarded to act on.

Sometimes it is advantageous to consider yet another generator

$$\mathcal{L}_s = \frac{\sigma^2}{2\gamma^2} \Delta_q - \underbrace{\left(\frac{1}{2\sigma^2} (\nabla V(q))^2 - \frac{1}{2\gamma} \Delta V(q) \right)}_{U(q)} \quad (30)$$

regarded to act on a suitable subspace of $L^2(dq)$. It is defined in terms of the potential function $U : \Omega \rightarrow \mathbf{R}$ and allows to apply the powerful theory of Schrödinger operators (see, e.g., Reed and Simon [59]); for a brief outline see [71]. The **Schrödinger operator** \mathcal{L}_s is related to the generators \mathcal{L} by the identity $\mathcal{Q}^{\frac{1}{2}} \mathcal{L} v = \mathcal{L}_s (\mathcal{Q}^{\frac{1}{2}} v)$ for every $v \in L^1(\mu_{\mathcal{Q}})$.

3 Metastability

We present a mathematical characterization of metastability, its connection to eigenvalues of the transfer operator close to its dominant eigenvalue 1, and a theoretical as well as numerical identification strategy.

3.1 Characterizing Metastability

There is no unique but several definitions of metastability in literature (see, e.g., [4, 9, 72, 73]), each adapted to suit the context. Our definition fits the biomolecular application context and measures metastability w.r.t. the canonical ensemble. Combining the physical characterization of metastable sub-ensembles in Section 1.1 and the mathematical specification in Section 1.3, we aim at defining a transition probability from a subset B into C within the time span τ , denoted by $p(\tau, B, C)$, such that an invariant sub-ensemble C is characterized by $p(\tau, C, C) = 1$, while a metastable sub-ensemble can be characterized by $p(\tau, C, C) \approx 1$. Being "close to 1" is obviously a vague statement—however, in most applications we are interested in a decomposition into the most metastable subsets, which eliminates the problem of interpreting "close to 1". Instead we have to determine the number of subsets, we are looking for. In our approach, this is done by examining the spectrum of the propagator P_τ . Alternatively, we could determine a cascade of decompositions with an increasing number of metastable subsets.

Define the **transition probability** $p(t, B, C)$ from $B \in \mathcal{A}$ to $C \in \mathcal{A}$ within the time span t as the conditional probability

$$p(t, B, C) = \mathbf{P}_\mu[X_t \in C | X_0 \in B] = \frac{\mathbf{P}_\mu[X_t \in C \text{ and } X_0 \in B]}{\mathbf{P}_\mu[X_0 \in B]}, \quad (31)$$

where \mathbf{P}_μ indicates that the initial distribution of the Markov process X_t is due to μ , hence $X_0 \sim \mu$. The similar symbols for both the transition probability between subsets $p(t, B, C)$ as well as for the stochastic transition function corresponding to the Markov process emphasize the strong relation to the definition of $p(t, x, C)$ in (4), which allows to rewrite (31) as

$$p(t, B, C) = \frac{1}{\mu(B)} \int_B p(t, x, C) \mu(dx). \quad (32)$$

The transition probability quantifies the dynamical fluctuations within the stationary ensemble μ . For later reference, we state the following two properties:

- (i) using the duality bracket $\langle \cdot, \cdot \rangle_\mu$ between $L^1(\mu)$ and $L^\infty(\mu)$, we get

$$p(t, B, C) = \frac{\langle P_t \mathbf{1}_B, \mathbf{1}_C \rangle_\mu}{\langle \mathbf{1}_B, \mathbf{1}_B \rangle_\mu}. \quad (33)$$

- (ii) metastability of C may equivalently be characterized by the condition that $p(t, C, \mathbf{X} \setminus C) \approx 0$, which yields [73]:

$$p(t, C, \mathbf{X} \setminus C) = \frac{1}{2\mu(C)} \|P_t \mathbf{1}_C - \mathbf{1}_C\|_1. \quad (34)$$

In Section 1.1 we have seen that metastability of sub-ensembles can experimentally be measured w.r.t. some observation time τ . Therefore, we will fix some $\tau > 0$ and concentrate in the sequel on the single propagator P_τ rather than on the entire semigroup of propagators $\{P_t\}_{t \in \mathbf{T}}$.

3.2 Identifying Metastable Subsets

We now present the fundamental algorithmic strategy used to identify metastable subsets. The basic idea is to interpret metastability as almost invariance. Since invariant subsets are associated with the eigenvalue $\lambda = 1$ and can be identified exploiting the corresponding eigenfunctions [17], metastable subsets are thought to be associated with almost $\lambda = 1$ eigenvalues and can be identified by exploiting the corresponding eigenfunctions.

Consider the propagator $P_\tau : L^r(\mu) \rightarrow L^r(\mu)$ with $r = 1, 2$; its spectrum is contained in the unit disc $\{\lambda \in \mathbf{C} : |\lambda| \leq 1\}$. Whenever a proper subset $C \subset \mathbf{X}$ is invariant under the Markov process, i.e., $p(t, x, \mathbf{X} \setminus C) = 0$ for all $x \in C$, the probability density $\hat{\mathbf{1}}_C = \mathbf{1}_C / \mu(C)$ is an eigenfunction corresponding to $\lambda = 1$. In particular, since μ is assumed to be invariant, $\hat{\mathbf{1}}_{\mathbf{X}}$ is an eigenfunction corresponding to $\lambda = 1$. Loosely speaking, a characterization of metastability according to (34) suggests that C is metastable if $\hat{\mathbf{1}}_C$ is an approximate eigenfunction corresponding to an eigenvalue close to $\lambda = 1$. This motivates the following **algorithmic strategy**:

Metastable subsets (on the time scale $\tau > 0$) can be identified via eigenfunctions of the propagator P_τ corresponding to eigenvalues $|\lambda| < 1$ close to the Perron root $\lambda = 1$. In doing so, the number of metastable subsets is equal to the number of eigenvalues close to 1, including $\lambda = 1$ and counting multiplicity.

The strategy mentioned above has first been proposed by Dellnitz and Junge [13] for discrete dynamical systems with weak random perturbations and has been successfully applied to molecular dynamics in different contexts [69, 71, 68]; a justification is given by Theorem 3.1. The algorithmic strategy necessitates the following two **conditions on the propagator P_τ** :

- (C1) The essential spectral radius of P_τ is less than one, i.e., $r_{\text{ess}}(P_\tau) < 1$.
- (C2) The eigenvalue $\lambda = 1$ of P_τ is simple and dominant, i.e., $\eta \in \sigma(P_\tau)$ with $|\eta| = 1$ implies $\eta = 1$.

It is important to remember that we defined transition probabilities between subsets, and therefore metastability, w.r.t. the invariant measure μ . Assume that the Markov process X_t admits another invariant measure ν , which, for sake of simplicity, is absolutely continuous w.r.t. μ with density $f \in L^1(\mu)$. Then, f is an eigenfunction of P_τ corresponding to $\lambda = 1$. As a consequence, we will not be able to decide in general whether an eigenfunction corresponding to some eigenvalue $\lambda \approx 1$ is related to metastable behavior of the ensemble represented by μ or by ν . Thus, the algorithmic strategy requires *uniqueness* of the invariant measure. Additionally, the physical interpretation of the ensemble excludes other eigenvalues than $\lambda = 1$ on the unit circle. Hence, $\lambda = 1$ has to be simple and dominant. For the numerical realization and discretization of the eigenvalue problem, we moreover need that the relevant eigenvalues are *isolated and of finite multiplicity*. For those eigenvalues convergence results of the numerical discretization algorithm can be established. This implies that the essential spectral radius has to be less than 1, hence permitting the existence of isolated eigenvalues of finite multiplicity close to $\lambda = 1$.

3.3 Metastable Subsets and Eigenvalues Close to 1

We now give a mathematical justification of the algorithmic strategy introduced above. The main result is stated in Theorem 3.1. It illuminates the strong relation between the existence of a cluster of eigenvalues close to 1 and a possible decomposition of the state space into metastable subsets. We state the theorem under the additional assumption of reversibility of the Markov process X_t and comment on how the results can be applied to non-reversible Markov processes.

Consider the propagator $P_\tau : L^2(\mu) \rightarrow L^2(\mu)$ satisfying the two conditions (C1) and (C2), and assume that the Markov process is reversible. Due to Proposition 1.1, P_t is self-adjoint and its spectrum has the form

$$\sigma(P_\tau) \subset [l, r] \cup \{\lambda_2\} \cup \{1\},$$

with $-1 < l \leq r \leq \lambda_2 < \lambda_1 = 1$. We restrict our considerations to the case that the Perron root is “nearly two-fold degenerate”: We assume that λ_2 is a simple isolated eigenvalue, hence $r = \lambda_3 < \lambda_2$, and further that the corresponding eigenfunction v_2 is normalized by $\langle v_2, v_2 \rangle_\mu = 1$ and satisfies $v_2 \in L^\infty(\mu)$. Note that $\langle v_2, \mathbf{1}_\mathbf{X} \rangle_\mu = 0$, since P_τ is self-adjoint.

A **decomposition** $\mathcal{D} = \{D_1, \dots, D_n\}$ of the state space \mathbf{X} is a collection of subsets $D_k \subset \mathbf{X}$ with the properties:

- (i) positivity: $\mu(D_k) > 0$ for every k ,
- (ii) disjointness: $D_k \cap D_l = \emptyset$ for $k \neq l$, and

(iii) covering property: $\cup_{k=1}^n D_k = \mathbf{X}$.

For a decomposition $\mathcal{D} = \{B, C\}$ of \mathbf{X} into two subsets, we define the following function

$$v_{BC} = \sqrt{\frac{\mu(C)}{\mu(B)}} \mathbf{1}_B - \sqrt{\frac{\mu(B)}{\mu(C)}} \mathbf{1}_C, \quad (35)$$

which is constant on either of the two sets B and C , and is normalized to $\|v_{BC}\|_2 = 1$. Under the assumptions on the propagator P_τ stated above we obtain the following relation between the existence of metastable subsets and eigenvalues close to 1.

Theorem 3.1 *Let $\mathcal{D} = \{B, C\}$ be an arbitrary decomposition of \mathbf{X} into two subsets. Then*

$$1 + \kappa \lambda_2 \leq [p(\tau, B, B) + p(\tau, C, C)] \leq 1 + \lambda_2, \quad (36)$$

with $\kappa = \langle v_2, v_{BC} \rangle_\mu^2 \leq 1$. In addition, choosing

$$B = \{x \in \mathbf{X} : v_2(x) \geq 0\} \quad \text{and} \quad C = \{x \in \mathbf{X} : v_2(x) < 0\},$$

we have $1 - 8c^2\epsilon \leq \kappa$ with constants $\epsilon = (1 - \lambda_2)/(1 - \lambda_3)$ and $c = \|v_2\|_\infty$.

Proof: The proof is based on results by Davies [9, 10] and a subsequent paper of Singleton [73]. In order to be applicable to our situation, we have to extend their results, since in general, we cannot assume the existence of an infinitesimal generator as in [9, 10, 73]. In [73] we have to replace the strongly continuous semigroup $\exp(-Ht)$ by P_τ . Furthermore, to match the assumptions on the spectrum in [73] with ours, we have to rescale the time of the semigroup. Interpreting $p(\tau, B, C)$ as $\mu(C) \langle P_\tau \mathbf{1}_B / \mu(B), \mathbf{1}_C / \mu(C) \rangle_\mu$, as stated by property (i) in Section 3.1, Lemma 4 of [73] and its subsequent remark state that

$$p(\tau, B, C) = \frac{1}{2\mu(B)} \|\mathbf{1}_B - P_\tau \mathbf{1}_B\|_1 = \mu(C) \langle v_{BC} - P_\tau v_{BC}, v_{BC} \rangle_\mu.$$

We then exploit Theorem 5 of [73] to bound the scalar product by

$$1 - \lambda_2 \leq \langle v_{BC} - P_\tau v_{BC}, v_{BC} \rangle_\mu \leq 1 - \kappa \lambda_2$$

with $\kappa = \langle v_2, v_{BC} \rangle_\mu^2 \leq 1$. Putting everything together, we end up with

$$\mu(B)(1 - \lambda_2) \leq p(\tau, B, C) \leq \mu(B)(1 - \kappa \lambda_2). \quad (37)$$

Repeating the calculation with $v_{CB} = -v_{BC}$ and exchanged roles of B and C , we see that inequality (37) holds in an analogous way for $p(\tau, C, B)$.

Hence, summing up both inequalities and exploiting $\mu(B) + \mu(C) = 1$, this finally gives

$$\begin{aligned} 1 - \lambda_2 &\leq [p(\tau, B, C) + p(\tau, C, B)] \leq 1 - \kappa\lambda_2 \\ \Leftrightarrow 1 + \kappa\lambda_2 &\leq [p(\tau, B, B) + p(\tau, C, C)] \leq 1 + \lambda_2, \end{aligned}$$

which is the first statement of Theorem 3.1. For the second statement on the lower bound on κ , we conclude from Theorem 3 of [73] that

$$\kappa = 1 - \frac{1}{2}\|v_2 - v_{BC}\|_2^2. \quad (38)$$

In the following we determine, analogous to Theorem 3 in [9], an upper bound on $\|v_2 - v_{BC}\|_2$, which implies a lower bound on κ . Once again, $\exp(-Ht)$ has to be replaced by P_τ and furthermore, the infinitesimal generator $-H$ has to be substituted by $\text{Id} - P_\tau$. Define the function

$$\psi = \sqrt{\frac{\|v_2^-\|_2}{\|v_2^+\|_2}} v_2^+ + \sqrt{\frac{\|v_2^+\|_2}{\|v_2^-\|_2}} v_2^-,$$

where v_2^+ and v_2^- denote the positive and negative part of v_2 , respectively. Note that $\langle P_\tau \psi, \psi \rangle_\mu \geq \lambda_1$ implies

$$\langle (\text{Id} - P_\tau) \psi, \psi \rangle_\mu \geq 1 - \lambda_2. \quad (39)$$

Now, define $\xi = \psi - \langle \psi, \mathbf{1}_X \rangle \mathbf{1}_X$. Since ξ is orthogonal to $\mathbf{1}_X$ and v_2 , we obtain by means of Cauchy–Schwarz $\langle P_\tau \xi, \xi \rangle \leq \lambda_3 \|\xi\|_2^2$ and therefore

$$(1 - \lambda_3) \|\xi\|_2^2 \leq 1 - \lambda_3 \|\xi\|_2^2 \leq 1 - \langle P_\tau \xi, \xi \rangle \leq \langle (\text{Id} - P_\tau) \xi, \xi \rangle. \quad (40)$$

Assembling (39) and (40) results in

$$\begin{aligned} (1 - \lambda_3) \|\xi\|_2^2 &\leq \langle (\text{Id} - P_\tau) \xi, \xi \rangle \\ &= \langle (\text{Id} - P_\tau) \psi, \psi \rangle - \langle \psi, \mathbf{1} \rangle \langle (\text{Id} - P_\tau) \psi, \mathbf{1}_X \rangle \\ &\quad - \langle \psi, \mathbf{1} \rangle \langle (\text{Id} - P_\tau) \mathbf{1}_X, \psi \rangle + \langle \psi, \mathbf{1} \rangle^2 \langle (\text{Id} - P_\tau) \mathbf{1}_X, \mathbf{1}_X \rangle \\ &= \langle (\text{Id} - P_\tau) \psi, \psi \rangle \leq 1 - \lambda_2, \end{aligned}$$

which implies $\|\xi\|_2^2 \leq \epsilon$ with $\epsilon = (1 - \lambda_2)/(1 - \lambda_3)$. With this modification in the proof of Theorem 3 in [9], we finally get

$$\|v_2 - v_{BC}\|_2^2 \leq 16 \|v_2\|_\infty^2 \epsilon,$$

which together with (38) gives the lower bound on κ . \square

Theorem 3.1 highlights the strong relation between a decomposition of the state space into two metastable subsets and a second eigenvalue close

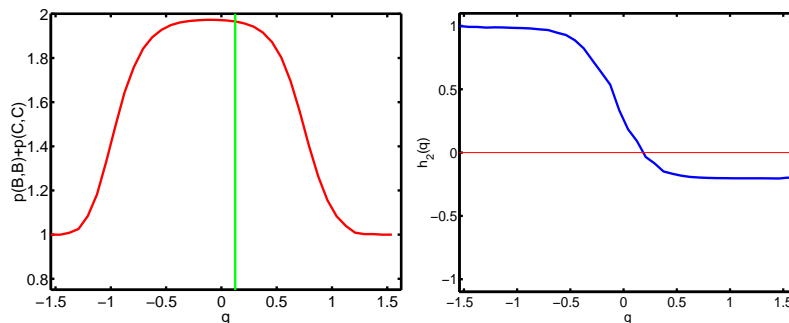


Figure 2: Data based on Smoluchowski equation. Left: metastability of a decomposition $\mathcal{D} = \{B, C\}$ with $B = (-\infty, q]$ and $C = (q, \infty]$ in dependence on q . The problem of finding the maximal value corresponding to the optimal decomposition is ill-conditioned. The vertical line corresponds to the decomposition $B = \{x \in \mathbf{X} : h_2(x) \geq 0\}$ and $C = \mathbf{X} \setminus B$ defined by the second eigenfunction h_2 (right). We obtain $\kappa = 0.984$ according to Theorem 3.1.

to the dominant eigenvalue 1. For an arbitrary decomposition $\mathcal{D} = \{B, C\}$ let us call $p(\tau, B, B) + p(\tau, C, C)$ the metastability of the decomposition \mathcal{D} . Then Theorem 3.1 states that the metastability of a decomposition $\mathcal{D} = \{B, C\}$ cannot be larger than $1 + \lambda_2$, while it is at least $1 + \kappa\lambda_2$. The upper bound is “large” whenever the eigenfunction v_2 corresponding to λ_2 is almost constant on the two metastable subsets B and C . As stated by the second part of Theorem 3.1, we can guarantee metastability for the particular decomposition into $B = \{x : v_2(x) \geq 0\}$ and $C = \{x : v_2(x) < 0\}$ whenever (i) the gap between the second and third eigenvalue is large, hence $\epsilon = (1 - \lambda_2)/(1 - \lambda_3)$ is small, and (ii) the essential maximum $c = \|v_2\|_\infty$ of the second eigenfunction v_2 is small. In [10] Davies proved, that in the case of a strongly continuous positive semigroup of self-adjoint propagators, e.g., in the case of the Smoluchowski dynamics, the lower bound on κ is in fact independent of $c = \|v_2\|_\infty$, whenever $c < \infty$. Nevertheless, the lower bound on $p(\tau, B, B) + p(\tau, C, C)$ via $\epsilon \ll 1$ implies the quite restrictive assumption: $\lambda_3 \ll \lambda_2$ on the spectrum P_τ . In numerical experiments we have observed intriguing results of the identification strategy even for situations corresponding to large ϵ -values [17].

In view of Theorem 3.1, it is natural to ask, whether there is an *optimal* decomposition with highest possible metastability. The answer is illustrated by Figure 2: *Even if there exists an optimal decomposition, the problem of finding it might be ill-conditioned.* The graph shows the metastability of a family of decompositions. It is based on the propagator P_τ corresponding to the Smoluchowski equation for the double-well potential on $\mathbf{X} = \mathbf{R}$. The conditions (C1) and (C2) on P_τ are justified by Proposition 6.5, while the assumption on the spectrum can be fulfilled by choosing an appropriate in-

verse temperature β . We identify a whole domain of decompositions that are nearly optimal. In this case the problem of finding the maximum is ill-conditioned. We also observe that the decomposition suggested by our identification algorithm is nearly optimal. The phenomenon illustrated by Figure 2 is believed to be typical in our application context, which is due to the fact that the canonical ensemble has large regions of almost vanishing probability.

Having an application to more complicated dynamical behavior in mind, we claim the following generalization of Theorem 3.1 for a decomposition into more than two subsets: Assume that the propagator P_τ , acting on $L^2(\mu)$, is associated with a reversible Markov process and satisfies the conditions (C1) and (C2). Moreover, assume that its spectrum is of the form $\sigma(P_\tau) \subset [l, r] \cup \{\lambda_n\} \cup \dots \cup \{\lambda_2\} \cup \{1\}$ with simple, isolated eigenvalues $\lambda_n < \dots < \lambda_2 < \lambda_1 = 1$ and corresponding eigenfunctions in $L^\infty(\mu)$. Given a decomposition $\mathcal{D} = \{D_1, \dots, D_n\}$ of \mathbf{X} , denote by $\{v_{D_1}, \dots, v_{D_n}\}$ some μ -orthonormal basis of $\text{span}\{\mathbf{1}_{D_1}, \dots, \mathbf{1}_{D_n}\}$.

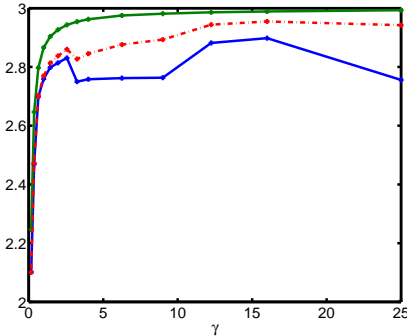


Figure 3: Illustration of the Conjecture for a decomposition of the state space into three subsets. The graph is based on the propagator P_τ corresponding to the Smoluchowski equation. The test system is the three-well potential defined in (63) for $\beta = 2$ and different values of γ . For details on the discretization see Section 6. The top solid line represents the upper bound $1 + \lambda_2 + \lambda_3$ in the Conjecture, while the dashed line corresponds to the metastability $p(\tau, D_1, D_1) + p(\tau, D_2, D_2) + p(\tau, D_3, D_3)$ of the decomposition obtained by applying the identification algorithm. The bottom solid line represents the lower bound $\kappa_1 + \kappa_2\lambda_2 + \kappa_3\lambda_3$.

Conjecture. Let $\mathcal{D} = \{D_1, \dots, D_n\}$ be an arbitrary decomposition of \mathbf{X} into n subsets. Then

$$\kappa_1 + \dots + \kappa_n \lambda_n \leq [p(\tau, D_1, D_1) + \dots + p(\tau, D_n, D_n)] \leq 1 + \dots + \lambda_n$$

with $\kappa_j = \langle v_j, v_{D_j} \rangle_\mu^2 \leq 1$.

For a numerical verification of the Conjecture for a decomposition of the state space into $n = 3$ subsets see Figure 3. The conjecture is in agreement

with all numerical experiments performed in Section 6.

In the following, we want to comment on an extension to *non-reversible* Markov processes. In his PhD thesis [74], Singleton proved results about metastable states for non-self-adjoint strongly continuous Markov semi-groups, which might be used to extend Theorem 3.1. The main problem in the absence of self-adjointness is to control both, the propagator and its adjoint at the same time. This necessitates increasing technical effort, since we have to pose conditions on spectral projections and resolvents rather than on the distribution of eigenvalues to handle non-self-adjointness. As Singleton stated in [74], the results obtained for suitable non-self-adjoint operators are of the same order of magnitude.

We will proceed in a different way based on an idea due to Froyland and Dellnitz [25]. They made the fascinating observation that we can associate to every non-reversible Markov chain a reversible Markov chain that possesses the same invariant measure and the same metastable subsets. We will extend this approach for our purpose from the finite state space to the general state space. Consider a Markov process $X_t = \{X_t\}_{t \in \mathbf{T}}$ and assume that its stochastic transition function p is absolutely continuous, hence $p(t, x, dy) = p(t, x, y)\mu(dy)$ with density jointly measurable in x and y . For some fixed observation time span $\tau > 0$, consider the discrete-time Markov process $X_n = \{X_{\tau n}\}_{n \in \mathbf{Z}_+}$; its stochastic transition function is given by $p_\tau(n, x, dy) = p(\tau n, x, dy)$. We define the **time-reversed Markov process** $Y_n = \{Y_n\}_{n \in \mathbf{Z}_+}$ via its stochastic transition function q_τ given by

$$q_\tau(n, x, dy) = q_\tau(n, x, y)\mu(dy) = p_\tau(n, y, x)\mu(dy),$$

which by definition is again absolutely continuous and discrete in time. The map q_τ satisfies the requirements for a stochastic transition function, since $q_\tau(n, x, \mathbf{X}) = 1$ due to invariance of μ , and $\int_{\mathbf{X}} q_\tau(n, x, dz)q_\tau(1, z, A) = q_\tau(n+1, x, A)$, which implies the Chapman–Kolmogorov equation. If p_τ is reversible then $q_\tau = p_\tau$, as we would expect. The time-reversed Markov process Y_n has two important properties (analogous to [25]):

- (i) The probability measure μ is invariant w.r.t. Y_n , since

$$\begin{aligned} \int_{\mathbf{X}} q_\tau(n, x, A)\mu(dx) &= \int_{\mathbf{X}} \int_A p(\tau n, y, x)\mu(dy)\mu(dx) \\ &= \int_A \int_{\mathbf{X}} p(\tau n, y, x)\mu(dx)\mu(dy) = \mu(A). \end{aligned}$$

- (ii) The time-reversed Markov process Y_n has the same metastable subsets as the original process X_n , i.e.,

$$p_\tau(n, C, C) = q_\tau(n, C, C)$$

for arbitrary $C \in \mathcal{A}$ and every $n \in \mathbf{Z}_+$. This is a special case of the general identity $\mu(C) p_\tau(n, C, D) = \mu(D) q_\tau(n, D, C)$, which is due to

$$\begin{aligned} \mu(C) p_\tau(n, C, D) &= \int_C \int_D p_\tau(n, x, y) \mu(dy) \mu(dx) \\ &= \int_C \int_D q_\tau(n, y, x) \mu(dy) \mu(dx) \\ &= \mu(D) q_\tau(n, D, C). \end{aligned}$$

Although neither X_n nor Y_n need to be reversible, we can use them to construct a reversible Markov process (analogous to [25]):

Theorem 3.2 *Define the **time-symmetrized Markov process** $Z_n = \{Z_n\}_{n \in \mathbf{Z}_+}$ via its one-step stochastic transition function*

$$r_\tau(1, x, dy) = \frac{1}{2} [p_\tau(1, x, y) + q_\tau(1, x, y)] \mu(dy).$$

Its n -step version given by the Chapman–Kolmogorov equation. Then, Z_n is invariant w.r.t. to μ , reversible and possesses the same one step metastability as X_n , i.e.,

$$p_\tau(1, C, C) = r_\tau(1, C, C)$$

for arbitrary $C \in \mathcal{A}$.

Proof: The statements about the invariance of μ and the metastability are obvious. Now, let us prove that Z_n is reversible w.r.t. μ , hence we have to prove condition (7). For $A, B \in \mathcal{A}$ we have

$$\begin{aligned} \int_A r_\tau(1, x, B) \mu(dx) &= \int_A \int_B \frac{1}{2} [p_\tau(1, x, y) + q_\tau(1, x, y)] \mu(dx) \mu(dy) \\ &= \int_A \int_B \frac{1}{2} [q_\tau(1, y, x) + p_\tau(1, y, x)] \mu(dx) \mu(dy) \\ &= \int_B r_\tau(1, x, A) \mu(dx); \end{aligned}$$

hence Z_n is reversible. \square

If the original Markov process X_t is reversible, then $r_\tau = p_\tau$ and the time-symmetrized Markov process coincides with the original one sampled at rate τ . For the interesting case of an originally non-reversible Markov process this is quite different. In general, we have:

- (i) The n -step transition probability $r_\tau(n, \cdot, \cdot)$ is not defined via the sum of the n -step transition probabilities $p_\tau(n, \cdot, \cdot)$ and $q_\tau(n, \cdot, \cdot)$, since already

$$r_\tau(2, x, A) \neq \int_{\mathbf{X}} \frac{1}{2} [p_\tau(2, x, A) + q_\tau(2, x, A)].$$

This is not surprising, because the left hand side, defined via the Chapman–Kolmogorov equation, involves the product of $p_\tau(1, \cdot, \cdot)$ and $q_\tau(1, \cdot, \cdot)$, while the right hand side does not.

- (ii) Due to (i) we have $p_\tau(n, C, C) \neq r_\tau(n, C, C)$ for $n > 1$.
- (iii) There exists no continuous–time Markov process $\hat{Z}_t = \{\hat{Z}_t\}_{t \in \mathbf{R}}$ such that Z_n is obtained by sampling \hat{Z}_t at rate τ , i.e., such that $Z_n = \hat{Z}_{n\tau}$ for $n \in \mathbf{Z}_+$. Hence, even if the original Markov process is defined via a stochastic differential equation, this is not the case for the time–symmetrized Markov process.

Sums of transition probabilities are frequently encountered in the Markov chain Monte Carlo theory, where it is well known that a realization of the time–symmetrized Markov process can be performed in two steps: (i) choose randomly one of the two transition functions p_τ or q_τ with equal probability $1/2$. (ii) proceed according to the chosen transition function. Repeat this procedure for every discrete time step. However, in order to discretize the propagator corresponding to the time–symmetrized Markov process, we will proceed in a different way, as outlined in Section 5.3.

In view of Theorem 3.2, we conclude that the original, *possibly non-reversible* Markov process X_t possesses a decomposition into metastable subsets (on the timescale τ), if the time–symmetrized Markov process Z_n does. This allows us to apply Theorem 3.1. In particular, the eigenfunctions related to the time–symmetrized Markov process can be used to identify the metastable subsets of the original, non–reversible Markov process. We will exemplify the time–symmetrization approach for the Langevin equation in Section 6.3.

4 Analysis of Transfer Operators

In the preceding section we have presented an algorithmic approach to the identification of metastable subsets under the two conditions (C1) and (C2), which are functional analytical statements on the spectrum of the propagator P_τ . In this section, we want to transform these conditions into a more probabilistic language, which will result in establishing equivalent conditions on the stochastic transition function. For general Markov processes it is natural to consider P_τ acting on $L^1(\mu)$, the Banach space that includes all probability densities w.r.t. μ . Yet, for *reversible* Markov processes it is advantageous to restrict the analysis to $L^2(\mu)$, since the propagator will then be self-adjoint. Therefore, in the first two sections we start with analyzing the conditions (C1) and (C2) in $L^1(\mu)$, while in the third section, we then concentrate on $L^2(\mu)$. For convenience we use the abbreviation $P = P_\tau$ and $p(x, C) = p(\tau, x, C)$ for some fixed time $\tau > 0$. As a consequence, $P^n = P_{n\tau}$ corresponds to the Markov process sampled at rate τ with stochastic transition function given by $p^n(\cdot, \cdot) = p(n\tau, \cdot, \cdot)$. The results presented in this section mainly follow [34].

4.1 The Spectrum and its Parts

Consider a complex Banach space E with norm $\|\cdot\|$ and denote the spectrum⁶ of a bounded linear operator $P : E \rightarrow E$ by $\sigma(P)$. For an eigenvalue $\lambda \in \sigma(P)$, the multiplicity of λ is defined as the dimension of the generalized eigenspace; see e.g., [40, Chap. III.6]. Eigenvalues of multiplicity 1 are called simple. The set of all eigenvalues $\lambda \in \sigma(P)$ that are isolated and of finite multiplicity is called the **discrete spectrum**, denoted by $\sigma_{\text{discr}}(P)$. The **essential spectral radius** $r_{\text{ess}}(P)$ of P is defined as the smallest real number, such that outside the ball of radius $r_{\text{ess}}(P)$, centered at the origin, are only discrete eigenvalues, i.e.,

$$r_{\text{ess}}(P) = \inf\{r \geq 0 : \lambda \in \sigma(P) \text{ with } |\lambda| > r \text{ implies } \lambda \in \sigma_{\text{discr}}(P)\}.$$

This definition of $r_{\text{ess}}(P)$ is unusual in the sense that it does not involve any definition of the essential spectrum; yet, it is the way we will exploit $r_{\text{ess}}(P)$ and it will be justified by Theorem 4.1 below. Usually, the essential spectral radius is related to the smallest disc containing the entire essential spectrum $\sigma_{\text{ess}}(P)$ of P . Unfortunately, there are many different characterizations of essential spectra (see e.g., [66, 47] or [32, Chapter 107]). The definition that results in the smallest set is due to Kato [40, Chapter IV.5.6] who defines $\sigma_{\text{ess}}^{\text{Kato}}(P)$ as the complement of $\{\lambda \in \mathbf{C} : \lambda - P \text{ is semi Fredholm}\}$. The

⁶For common functional analytical terminology see, e.g., [19, 32, 40, 84].

⁷A bounded linear operator $P : E \rightarrow E$ on a Banach space E is said to be semi Fredholm, if its range $R(P) = \{y = Px : x \in E\}$ is closed and the dimension of its kernel

definition that results in the largest set is due to Browder [5] according to whom $\sigma_{\text{ess}}^{\text{Browder}}(P)$ is the complement of the discrete spectrum, as defined above. According to Lebow and Schechter [47] we get the surprising result that all other known definitions of essential spectra fall between those of Kato and Browder and lie inside the ball with radius $r_{\text{ess}}(P)$ centered at the origin:

Theorem 4.1 *For every bounded linear operator $P : E \rightarrow E$ on a complex Banach space E holds*

$$\sup\{|\lambda| : \lambda \in \sigma_{\text{ess}}^{\text{Kato}}(P)\} = r_{\text{ess}}(P) = \sup\{|\lambda| : \lambda \in \sigma_{\text{ess}}^{\text{Browder}}(P)\}.$$

Loosely speaking, Theorem 4.1 states that the essential spectral radius is invariant under the definition of the essential spectrum.

As a guiding example for our strategy to bound the essential spectral radius, consider the following semi-norm $\|\cdot\|_c$ defined by

$$\|P\|_c = \inf\{\|P - S\| : S \text{ compact}\}.$$

Then the essential spectral radius is characterized by

$$r_{\text{ess}}(P) = \lim_{n \rightarrow \infty} \|P^n\|_c^{1/n}.$$

Note the analogy to the spectral radius $r(P)$ of P , defined as the smallest upper bound for all elements of the spectrum: $r(P) = \sup\{|\lambda| : \lambda \in \sigma(P)\}$. In terms of the operator norm $\|\cdot\|_1$, the representation $r(P) = \lim_{n \rightarrow \infty} \|P^n\|_1^{1/n}$ as $n \rightarrow \infty$ is well-known [19, Chap. VII.3.5]. The above characterization of $r_{\text{ess}}(P)$ is closely related to **quasi-compactness**:

Definition 4.2 ([32]) *A bounded linear operator $P : E \rightarrow E$ is called quasi-compact, if there exist some $m \in \mathbf{Z}_+$ and a compact operator $S : E \rightarrow E$ such that $\|P^m - S\| < 1$.*

Combining quasi-compactness with the characterization of $r_{\text{ess}}(P)$ yields:

Corollary 4.3 *For bounded linear operator $P : E \rightarrow E$ holds*

- (i) *if $r_{\text{ess}}(P) < 1$ then P is quasi-compact*
- (ii) *if P is quasi-compact for some $m \in \mathbf{Z}_+$ and compact operator S with $\|P^m - S\| = 1 - \eta < 1$, then $r_{\text{ess}}(P) \leq (1 - \eta)^{1/m} < 1$.*

$N(P) = \{x \in E : Px = 0\}$ or the codimension of its range, i.e., $\dim E/R(P)$, are finite [40, Chapter IV.5]. If both, the dimension of the kernel and the codimension of the range are finite, then P is called a *Fredholm operator*.

We conclude that the essential spectral radius can be bound by using compact operators:

Find for some power P^m with $m \in \mathbf{Z}_+$ a decomposition into a compact part S and the remaining part $P^m - S$. Then, we have the upper bound: $r_{\text{ess}}(P) \leq \|P^m - S\|^{1/m}$.

In other words, the “larger” the compact part of P^m is, the smaller the essential spectral radius of P will be. Our goal is to relate compactness of S to properties of the stochastic transition function that defines P . Due to the various possible definitions of essential spectra, this approach might not be restricted to compact operators. This is indeed the case, as we will see below. The crucial point will be to find the class of operators that fits best both the Banach space as well as the propagator and the Markov process. In $L^1(\mu)$ weakly compact operators are better adapted for our purpose, while in $L^2(\mu)$ the compact ones will do a good job. This is basically due to the fact that in either case we can characterize the property of being (weakly) compact in terms of the underlying probability space, which finally enables us to relate bounds on the essential spectral radius to properties of the stochastic transition function. For relations between the essential spectral radius and measures of non-compactness, see [56, 83].

Spectral conditions can be quite sensitive to the Banach space of functions the operator is regarded to act on. This is illustrated by the following example due to Davies [11, Chapter 4.3].

Example 4.4 *Consider the Smoluchowski equation*

$$\dot{q} = -q + \dot{W} \quad (41)$$

on the state space $\mathbf{X} = \mathbf{R}$. It corresponds to the harmonic potential $V(q) = q^2/2$ with $\gamma = \sigma = 1$ and invariant probability measure

$$\mu_{\mathcal{Q}}(dq) = \frac{1}{\sqrt{2\pi}} \exp(-q^2/2) dq.$$

The Markov process defined by (41) is known as the Ornstein–Uhlenbeck process [38]. The evolution of densities $v = v(t, q)$ w.r.t. $\mu_{\mathcal{Q}}$ is governed by the Fokker–Planck equation

$$\partial_t v = \underbrace{\left(\frac{1}{2} \Delta - q \cdot \nabla_q \right)}_{\mathcal{L}} v, \quad (42)$$

which defines a strongly continuous contraction semigroup $P_t = \exp(t\mathcal{L})$ on $L^r(\mu)$ for every $1 \leq r < \infty$. The spectra of \mathcal{L} and P_t have the following properties:

- (i) In $L^1(\mu)$ it is $\sigma(\mathcal{L}) = \{z \in \mathbf{C} : \operatorname{Re}(z) \leq 0\}$, with every $z \in \sigma(\mathcal{L})$ satisfying $\operatorname{Re}(z) < 0$ being an eigenvalue of multiplicity two. This implies for the propagator that

$$\sigma(P_t) = \{z \in \mathbf{C} : |z| \leq 1\},$$

with every $z \in \sigma(P_t)$ satisfying $|z| < 1$ being an eigenvalue of infinite multiplicity, hence $r_{\text{ess}}(P_t) = 1$.

- (ii) In $L^2(\mu)$ it is $\sigma(\mathcal{L}) = \{z \in \mathbf{C} : z = 0, -1, -2, \dots\}$, with the n th Hermite polynomial being the eigenfunction corresponding to $\lambda_n = -n$. Hence, the entire spectrum is discrete. This implies for the propagator

$$\sigma(P_t) = \{z \in \mathbf{C} : z = e^{-tn} \text{ for } n = 0, 1, 2, \dots\},$$

with $r_{\text{ess}}(P_t) = 0$.

From a numerical point of view, we would like to consider the space of functions that is “generated” by the discretization procedure for finer and finer decompositions of the state space. This, however, is believed to be a very tough question.

4.2 Bounds on the Essential Spectral Radius in $L^1(\mu)$

This section analyzes the essential spectral radius of an arbitrary propagator $P : L^1(\mu) \rightarrow L^1(\mu)$ in terms of its stochastic transition function. In doing so, weakly compact operators will play an important role. The main result is stated in Theorem 4.13, which relates the essential spectral radius, uniform constrictiveness and a certain Doeblin-condition.

Definition 4.5 ([19, 50]) *A bounded linear operator $S : L^1(\mu) \rightarrow L^1(\mu)$ is called **weakly compact** if it maps the closed unit ball $B_1(X)$ onto a relatively weakly compact set, i.e., the closure of $S(B_1(X))$ is compact in the weak topology.*

Obviously, every compact operator is weakly compact; the converse is not true. The next theorem characterizes the essential spectral radius of an arbitrary bounded linear operator in terms of weakly compact operators.

Theorem 4.6 ([81, 83]) *Let $P : L^1(\mu) \rightarrow L^1(\mu)$ denote a bounded linear operator. Define the semi-norm $\Delta(P)$ according to*

$$\Delta(P) = \min \{\|P - S\|_1 : S \text{ is weakly compact}\}.$$

Then the essential spectral radius of P is characterized by

$$r_{\text{ess}}(P) = \lim_{n \rightarrow \infty} \Delta(P^n)^{1/n}. \quad (43)$$

In particular, $r_{\text{ess}}(P) \leq \Delta(P)$.

The theorem states that the larger the weakly compact part of P is, the less the essential spectral radius will be. Hence, good upper bounds on $r_{\text{ess}}(P)$ require a detailed analysis of weak compactness. It should be clear from the introductory statements of this section that we could also apply Corollary 4.3 to characterize the essential spectral radius in $L^1(\mu)$ by compact operators. The utility of weakly compact operators will become apparent by the next theorem that relates this particular class of operators to the underlying measure space $(\mathbf{X}, \mathcal{A}, \mu)$.

Theorem 4.7 ([81, 82]) *Let $P : L^1(\mu) \rightarrow L^1(\mu)$ denote a bounded linear operator. Then*

$$\Delta(P) = \limsup_{\mu(A) \rightarrow 0} \|\mathbf{1}_A \circ P\|_1, \quad (44)$$

where the limit is understood to be taken over all sequences of subsets whose μ -measure converges to zero, and $\mathbf{1}_A$ is interpreted as a multiplication operator: $(\mathbf{1}_A v)(x) = \mathbf{1}_A(x)v(x)$. In particular,

$$\limsup_{\mu(A) \rightarrow 0} \|\mathbf{1}_A \circ P\|_1 = 0,$$

if and only if P is weakly compact.

As a consequence of Theorem 4.7, we will deduce in the following that absolutely continuous stochastic transition functions may give rise to weakly compact operators, while transition functions that are singular w.r.t. μ never do so. This will finally enable us to characterize the essential spectral radius in terms of properties of the stochastic transition function.

Corollary 4.8 *Consider some propagator $S : L^1(\mu) \rightarrow L^1(\mu)$ defined by*

$$Sv(y) = \int_{\mathbf{X}} v(x)p(x, y)\mu(dx) \quad (45)$$

associated with some absolutely continuous stochastic transition function $p(x, dy) = p(x, y)\mu(dy)$. Then S is weakly compact if there exists some $s > 1$ such that $\|p(x, \cdot)\|_s \in L^\infty(\mu)$ as a function of x , i.e.,

$$\text{ess sup}_{x \in \mathbf{X}} \int_{\mathbf{X}} p(x, y)^s \mu(dy) < \infty$$

holds. In particular, S is weakly compact if $\text{ess sup}_{x, y \in \mathbf{X}} p(x, y) < \infty$.

Proof: For $A \in \mathcal{B}(\mathbf{X})$, we have

$$\|\mathbf{1}_A \circ S\|_1 = \sup_{\|v\|_1 \leq 1} \int_A \int_{\mathbf{X}} v(x)p(x, y)\mu(dx)\mu(dy).$$

Applying Hölder's inequality twice, we finally get

$$\|\mathbf{1}_A \circ S\|_1 \leq \operatorname{ess\,sup}_{x \in \mathbf{X}} \int_A p(x, y) \mu(dy) \leq \|\mathbf{1}_A\|_r \operatorname{ess\,sup}_{x \in \mathbf{X}} \|p(x, \cdot)\|_s$$

with $1 \leq r, s \leq \infty$ and $1/s + 1/r = 1$. For $1 < s$, the limit of $\|\mathbf{1}_A \circ S\|_1$ as $\mu(A) \rightarrow 0$ tends to zero, since $\|\mathbf{1}_A\|_r = \sqrt[r]{\mu(A)}$. \square

For analyzing propagators corresponding to not necessarily absolutely continuous stochastic transition functions, consider the **Lebesgue decomposition** of $p(x, dy) = p_a(x, y)\mu(dy) + p_s(x, dy)$, where p_a and p_s represent the absolutely continuous and the singular part w.r.t. μ , respectively [45]. Furthermore, define the (not necessarily stochastic) transition function

$$r_n(x, y) = \begin{cases} p_a(x, y) & \text{if } p_a(x, y) \geq n \\ 0 & \text{otherwise} \end{cases}.$$

With this notation, we are ready to state the important

Theorem 4.9 ([82]) *For an arbitrary propagator $P : L^1(\mu) \rightarrow L^1(\mu)$ the equality*

$$\Delta(P) = \inf_{n \in \mathbf{Z}_+} \operatorname{ess\,sup}_{x \in \mathbf{X}} \{r_n(x, \mathbf{X}) + p_s(x, \mathbf{X})\}$$

holds.

In the particular case, where p_a gives rise to a weakly compact operator, Theorem 4.9 states that

$$\Delta(P) = \operatorname{ess\,sup}_{x \in \mathbf{X}} p_s(x, \mathbf{X}) = 1 - \operatorname{ess\,inf}_{x \in \mathbf{X}} \int_{\mathbf{X}} p_a(x, y) \mu(dy).$$

If only some decomposition $P = R + S$ with weakly compact S is known, we may still apply Theorem 4.6 to get an upper bound on $\Delta(P)$. Assume that the stochastic transition function can be decomposed according to $p(x, dy) = p_R(x, dy) + p_W(x, dy)$ such that S , defined via $Sv(y) = \int_{\mathbf{X}} v(x) p_W(x, dy)$, is weakly compact. Then

$$\Delta(P) \leq \operatorname{ess\,sup}_{x \in \mathbf{X}} p_R(x, \mathbf{X}) \leq 1 - \operatorname{ess\,inf}_{x \in \mathbf{X}} p_W(x, \mathbf{X})$$

by Theorem 4.6. Using one of the inequalities involving $\Delta(P)$, we are able to bound the essential spectral radius due to Theorem 4.6. This is illustrated by the following example due to Schütte [68, Chapter 4.1].

Example 4.10 *Consider the Hamiltonian system with randomized momenta for the harmonic potential $V(q) = q^2/2$ on some position space $\Omega \subset \mathbf{R}$ with*

inverse temperature β and positional canonical distribution $\mu_{\mathcal{Q}}$. Choose the observation time span $\tau = 2\pi$ and decompose the stochastic transition function according to

$$p_{\tau}(q, dy) = p_a(q, y)\mu_{\mathcal{Q}}(dy) + p_s(q, dy)$$

into an absolutely continuous and singular part w.r.t. $\mu_{\mathcal{Q}}$. Depending on the position space, we distinguish two cases

- (i) Consider $\Omega = \mathbf{R}$, the bounded system case. Since $\tau = 2\pi$ is the period of the harmonic oscillator, we deduce that $P_{\tau} = \text{Id}$ and hence $r_{\text{ess}}(P_{\tau}) = 1$. In terms of the stochastic transition function this means that $p_a = 0$ and $p_s(q, dy) = \delta_q(dy)$ for every $q \in \Omega$.
- (ii) Consider $\Omega = [-1, 1]$ with periodic boundary conditions. It can be shown that in this case the density p_a is bounded and satisfies

$$\inf_{q \in \Omega} \int_{\Omega} p_a(q, y)\mu_{\mathcal{Q}}(dy) = 2\Phi(-\sqrt{\beta})$$

where Φ denotes the distribution function of the standard normal distribution. Setting

$$\eta = 2\Phi(-\sqrt{\beta}) = 2(1 - \Phi(\sqrt{\beta}))$$

we have $0 \leq \eta \leq 1$ and finally $r_{\text{ess}}(P_{\tau}) \leq \Delta(P_{\tau}) = 1 - \eta$ due to⁸ Theorem 4.9. In other words, the (upper bound on the) essential spectral radius depends on the inverse temperature and therefore on the mean energy of the ensemble. The lower the mean energy (and hence the higher the inverse temperature) is, the larger the essential spectral radius will be. This corresponds to the intuition that the periodic system behaves more and more like the bounded system for decreasing mean energy.

So far we have shown how to prove $r_{\text{ess}}(P) < 1$ in terms of the stochastic transition function p . The properties imposed on p emerged from functional analytical requirements on the propagator P . We now link these results to the theory of Markov processes and Markov operators. An important property of Markov operators is constrictiveness [46]; it rules out the possibility that for some initial density v the iterates $P^n v$ eventually concentrate on a set of very small or vanishing measure.

⁸For the propagator regarded to act on $L^2(\mu_{\mathcal{Q}})$ the stronger statement $r_{\text{ess}}(P_{\tau}) = 1 - \eta$ is proved in [68].

Definition 4.11 A propagator $P : L^1(\mu) \rightarrow L^1(\mu)$ is called *constrictive* if there exist constants $\epsilon, \delta > 0$ such that for every density $v \in L^1(\mu)$ there exists $m = m(v) \in \mathbf{Z}_+$ with

$$\mu(A) \leq \epsilon \Rightarrow \int_A P^n v(y) \mu(dy) \leq 1 - \delta, \quad (46)$$

for every $n \geq m$. We call a propagator **uniformly constrictive** if there exists $m \in \mathbf{Z}_+$ such that (46) holds for $n \geq m$ uniformly in $L^1(\mu)$.

For arbitrary $v \in L^1(\mu)$, uniform constrictiveness can be restated as $\mu(A) \leq \epsilon \Rightarrow \|\mathbf{1}_A \circ P^n\|_1 \leq 1 - \delta$ for every $n \geq m$. Moreover, it is sufficient to assume that the condition holds for $n = m$ only, since due to $\|P^k\|_1 = 1$ for $k \in \mathbf{Z}_+$ this already implies (46) for all $n \geq m$. In view of the characterization of $\Delta(P)$ in (44), uniform constrictiveness seems to be closely related to $\Delta(P) < 1$ and thus to some bound on the essential spectral radius; this is indeed the case, as we will see below. Furthermore, there should exist a similar condition involving the backward transfer operator T . This, in turn, is closely related to the *Doebelin-condition*, which is well-known in the theory of Markov processes [18, 52, 60]. It states that there exists a probability measure ν , constants $\epsilon, \delta > 0$ and $m \in \mathbf{Z}_+$ such that $\nu(A) \leq \epsilon \Rightarrow \sup_{x \in \mathbf{X}} p^m(x, A) \leq 1 - \delta$. To suit our context, we slightly adapt the Doebelin-condition in the way that we require $\nu = \mu$ and that the implication holds for μ -a.e. points only:

Definition 4.12 The stochastic transition function p is said to fulfill the μ -a.e. *Doebelin-condition* if there exist constants $\epsilon, \delta > 0$ and $m \in \mathbf{Z}_+$ such that

$$\mu(A) \leq \epsilon \Rightarrow p^m(x, A) \leq 1 - \delta \quad (47)$$

for μ -a.e. $x \in \mathbf{X}$ and every $A \in \mathcal{B}(\mathbf{X})$.

Using the backward transfer operator, we deduce that (47) is equivalent to $\mu(A) < \epsilon \Rightarrow \|T^m \mathbf{1}_A\|_\infty = \text{ess sup}_{x \in \mathbf{X}} p^m(x, A) \leq 1 - \delta$. In fact, the condition is true for all $n \geq m$, since $\|T^{m+k} \mathbf{1}_A\|_\infty \leq \|T^k\|_\infty \|T^m \mathbf{1}_A\|_\infty$ and $\|T^k\|_\infty = 1$ holds for $k \geq 1$. The next theorem states the main result of this section. It relates the functional-analytical, the Markov operator theoretical and the Markov process theoretical point of view.

Theorem 4.13 Let $P : L^1(\mu) \rightarrow L^1(\mu)$ denote the propagator corresponding to a stochastic transition function $p : \mathbf{X} \times \mathcal{B}(\mathbf{X}) \rightarrow [0, 1]$. Then, the following statements are equivalent:

- (i) The essential spectral radius of P is less than one: $r_{\text{ess}}(P) < 1$.
- (ii) The propagator P is uniformly constrictive.

(iii) The stochastic transition functions fulfills the μ -a.e. Doeblin-condition.

If conditions (ii) or (iii) are satisfied for some $\epsilon, \delta > 0$ and $m \in \mathbf{Z}_+$, then condition (i) holds with $r_{\text{ess}}(P) \leq (1 - \delta)^{1/m}$.

Proof: Assume (i) holds, i.e., $r_{\text{ess}}(P) < 1$. Due to Eqs. (43) and (44), there exists $m \in \mathbf{Z}_+$ such that $\Delta(P^m) < 1$, which implies the μ -a.e. Doeblin-condition (4.12) due to $\|\mathbf{1}_A \circ P^m\|_1 = \|T^m \mathbf{1}_A\|_\infty$ (see Lemma 4.1 in [34]). As just stated, (iii) is equivalent to (ii). Using the note following Def. 4.11, it is obvious that (ii) and (i) are equivalent. The bound on $r_{\text{ess}}(P)$ follows from (43) and (44). \square

In view of the established equivalence, the essential spectral radius is related to the possibility of the system to eventually concentrate on a set of small or vanishing measure. In other words, the more the dynamics is smeared over the entire state space, the less is the essential spectral radius, while irregular or singular behavior may give rise to a large essential spectral radius.

4.3 Peripheral Spectrum and Properties in $L^1(\mu)$

This section analyzes the peripheral spectrum and its relation to properties of propagators P acting on $L^1(\mu)$. Due to our particular interest—cf. condition (C1)—we restrict the analysis to uniformly constrictive propagators, i.e., we assume that $r_{\text{ess}}(P) < 1$. We will see that under this assumption the peripheral spectrum completely characterizes the asymptotic properties of P , as it is known from the finite dimensional case.

Recall that we require throughout this thesis that the probability measure μ is invariant w.r.t. the Markov process. This is equivalent to the condition $P\mathbf{1}_X = \mathbf{1}_X$. A subset $E \subset X$ is called non-null if $\mu(E) > 0$. A non-null subset $E \subset X$ is called **invariant** if $P\mathbf{1}_E = \mathbf{1}_E$. Parts of the following two theorems are scattered over the literature see, e.g., [19, 46, 85].

Theorem 4.14 (Invariant Decomposition) *Let $P : L^1(\mu) \rightarrow L^1(\mu)$ denote a uniformly constrictive propagator. Then*

- (i) *there are only finitely many eigenvalues $\lambda \in \sigma_{\text{discr}}(P)$ with $|\lambda| = 1$, each being a root of unity. The dimension of each eigenspace is finite and equal to the multiplicity of the corresponding eigenvalue;*
- (ii) *the eigenvalue $\lambda = 1$ is of multiplicity d , if and only if there exists a decomposition of the state space*

$$\mathbf{X} = E_1 \cup \dots \cup E_d \cup F$$

into d mutually disjoint invariant subsets E_j and a set $F = \mathbf{X} \setminus \bigcup_j E_j$ of μ -measure zero.

Proof: Direct application of Thm. 4.13 of this thesis and Thm. 3 of [19, VIII.8] proves the first part. For the second statement, we exploit the fact that $Pv = v$ implies $Pv^+ = v^+$ and $Pv^- = v^-$, where $v^{+/-}$ denotes the positive or negative part of v , respectively [46]. Assume that the multiplicity of $\lambda = 1$ is d . Then, as a consequence of the first part, there exist d linear independent eigenfunctions v_1, \dots, v_d . Due to the decomposition result for v , we can also choose d linear independent *densities*, which we again denote by v_1, \dots, v_d . We now show that the densities can be chosen in such a way that their supports $E_j = \text{supp}(v_j)$ are mutually disjoint, i.e., $\mu(E_j \cap E_k) = 0$ for $j \neq k$. If for some choice of linear independent densities v_1, \dots, v_d there exist v_j, v_k such that $\mu(E_j \cap E_k) > 0$, we simply substitute v_j, v_k by $(v_j - v_k)^+, (v_j - v_k)^-$. This is possible, since $\text{span}\{(v_j - v_k)^+, (v_j - v_k)^-\} = \text{span}\{v_j, v_k\}$ and $\text{span}\{(v_j - v_k)^+, (v_j - v_k)^-, v_j, v_k\} > 2$ would be in contradiction to the fact that the multiplicity of $\lambda = 1$ is d . Due to $P\mathbf{1}_{\mathbf{X}} = \mathbf{1}_{\mathbf{X}}$, we have $v_j = \mathbf{1}_{E_j}/\mu(E_j)$ and $\sum_j \mu(E_j) = 1$. Finally, define $F = \mathbf{X} \setminus \bigcup_j E_j$. Since any decomposition into d mutually disjoint invariant subsets results in a multiplicity of $\lambda = 1$ of at least d , the second statement is proved. \square

The decomposition of the state space given by the theorem is unique up to μ -equivalence. There is an analogous decomposition result for the stochastic transition function p , since for every invariant subset E the identity

$$\mu(E) = \int_E \mathbf{1}_E(y) \mu(dy) = \int_E P\mathbf{1}_E(y) \mu(dy) = \int_E p(x, E) \mu(dx)$$

implies $p(x, E) = 1$ for μ -a.e. $x \in E$. Thus, the decomposition of Theorem 4.14 induces a decomposition of the stochastic transition function, which again is unique up to μ -equivalence. For a “strong” decomposition holding everywhere see, e.g., [85]. For some root of unity $\omega = \exp(2\pi i/m)$ with $m \in \mathbf{Z}_+$, we call $\sigma_{\text{cycle}}(\omega) = \{\omega, \omega^2, \dots, \omega^m\}$ an **eigenvalue cycle** associated with ω . A further subdecomposition of an invariant subset E into m mutually disjoint, non-null subsets $\{E_1, \dots, E_m\}$ is called a **subset cycle** of length m if $P\mathbf{1}_{E_j} = \mathbf{1}_{E_{j+1}}$ for $j = 1, \dots, m$ with the convention $E_{m+1} = E_1$. For the next theorem, an eigenvalue of multiplicity ν is interpreted as ν equal eigenvalues $\lambda_1, \dots, \lambda_\nu$ of multiplicity 1.

Theorem 4.15 (Cycle Decomposition) *Let $P : L^1(\mu) \rightarrow L^1(\mu)$ denote a uniformly contractive propagator. Then*

- (i) *each discrete eigenvalue $\lambda \in \sigma_{\text{discr}}(P)$ of unit modulus is part of some eigenvalue cycle, i.e., there exists $m \in \mathbf{Z}_+$ such that $\lambda \in \sigma_{\text{cycle}}(\omega)$ with $\omega = \exp(2\pi i/m)$;*
- (ii) *there is a one-to-one correspondence between eigenvalue cycles and subset cycles. More precisely, let d denote the multiplicity of $\lambda = 1$.*

Then the set of all eigenvalues of unit modulus can be decomposed into d eigenvalue cycles $\sigma_{\text{cycle}}(\omega_j)$ with $\omega_j = \exp(2\pi i/m_j)$, $m_j \in \mathbf{Z}_+$ and $j = 1, \dots, d$, if and only if the state space \mathbf{X} can be decomposed into d subset cycles $\{E_{j1}, \dots, E_{jm_j}\}$ of length m_j for $j = 1, \dots, d$.

Proof: Use Theorem 4.14 of this thesis and Theorem 11 in [85], which also holds in our case, to show that each invariant subset E can be decomposed into a subset cycles $\{E_1, \dots, E_m\}$ of length m . Consider the restricted propagator $P_E = \mathbf{1}_E \circ P \circ \mathbf{1}_E$, which is well-defined by Theorem 4.14. Then, the length m is equal to the multiplicity of $\lambda = 1$ of P_E^m . Thus, it remains to show that $\sigma(P_E) \cap \{|\lambda| = 1\} = \sigma_{\text{cycle}}(\omega)$ with $\omega = \exp(2\pi i/m)$. But every subset cycle $\{E_1, \dots, E_m\}$ of P is also a subset cycle of P_E and allows us to define m linear independent eigenfunctions $v_{k+1} = \sum_{j=0}^{m-1} \omega^{-kj} P_E^j \mathbf{1}_{E_1}$, see e.g. [13], which correspond to the eigenvalues ω^k for $k = 1, \dots, m$. This completes the proof. \square

From a functional analytical point of view, the decomposition results are related to a partial spectral decomposition of P , as we will see in the next result due to Dunford and Schwartz [19, Chapter VIII]. It exploits the fact that uniform constrictiveness is equivalent to quasi-compactness of the propagator (Thm. 4.13 and Cor. 4.3).

Theorem 4.16 (Spectral Decomposition) *Let $P : L^1(\mu) \rightarrow L^1(\mu)$ denote a uniformly constrictive propagator and let Π_λ denote the spectral projection corresponding to the discrete eigenvalue λ . Then, for every $n \in \mathbf{Z}_+$,*

$$P^n = \sum_{\lambda \in \sigma(P), |\lambda|=1} \lambda^n \Pi_\lambda + D^n$$

with some strict contraction $D : L^1(\mu) \rightarrow L^1(\mu)$ satisfying $\|D^n\|_1 \leq Mq^n$ for some $M > 0$ and $0 < q < 1$. Furthermore, the projections fulfill

$$\Pi_\lambda = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \frac{1}{\lambda^k} P^k, \quad (48)$$

where the limit is understood to be uniform.

Now, we exploit the above results to analyze properties of the propagator P and the underlying Markov process given by its stochastic transition function p .

Definition 4.17 *Let $P : L^1(\mu) \rightarrow L^1(\mu)$ denote a uniformly constrictive propagator.*

- (i) P is said to be **ergodic** if every invariant subset E is of μ -measure 1. Equivalently, $P\mathbf{1}_E = \mathbf{1}_E$ implies $\mu(E) = 0$ or $\mu(E) = 1$.

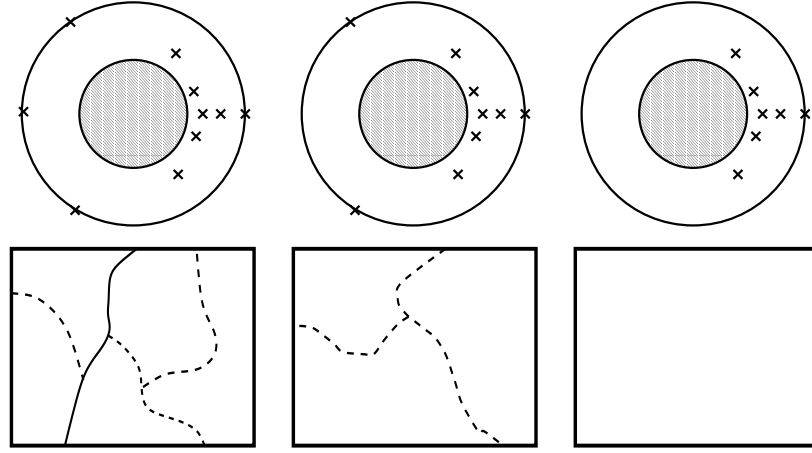


Figure 4: Top: idealized spectra of uniformly constrictive propagators. All eigenvalues are assumed to be simple except for $\lambda = 1$ in the left graphic, which must be at least two fold. Outer disc of radius $r = 1$ containing the entire spectrum and inner disc with radius $r_{\text{ess}} < 1$ containing the essential spectrum. Bottom: decomposition of the state space (rectangle) into invariant sets (separated by solid lines) and subset cycles (separated by dashed lines) corresponding to the spectra above. Left: two eigenvalue cycles with $m = 3$ and $m = 2$, respectively resulting in a decomposition of the state space into two invariant subsets that can be further decomposed into subset cycles of length $m = 3$ and $m = 2$, respectively. Middle: one eigenvalue cycle with $m = 3$ resulting in a decomposition of the state space into a subset cycle of length $m = 3$. Right: The eigenvalue 1 is simple and dominant. Hence, there neither exists a decomposition of the state space into invariant subsets nor subset cycles.

(ii) P is called **periodic** with period p if it is ergodic and p is the largest integer for which a subset cycle of length p occurs. If $p = 1$, then P is called **aperiodic**.

According to [46], a Markov operator $P : L^1(\mu) \rightarrow L^1(\mu)$ satisfying $P\mathbf{1}_{\mathbf{X}} = \mathbf{1}_{\mathbf{X}}$ is said to be ergodic if $P^n v$ converges in the sense of Cesàro for every density $v \in L^1(\mu)$ weakly to $\mathbf{1}_{\mathbf{X}}$. Anticipating the results of the next corollary and using Thm. 5.5.1 from [46, Sec. 5.5], it can easily be shown that for uniformly constrictive propagators this definition is equivalent to Def. 4.17 (i). In the theory of Markov processes, the term ergodicity is used slightly different, since it requires aperiodicity. Corollary 4.18 may be used to establish the relation. The next corollary states how these properties are related to the decomposition results previously obtained.

Corollary 4.18 Let $P : L^1(\mu) \rightarrow L^1(\mu)$ denote a uniformly constrictive propagator. Then

- (i) P is ergodic if and only if the eigenvalue $\lambda = 1$ is simple.
- (ii) P is aperiodic if and only if the eigenvalue $\lambda = 1$ is simple and dominant, i.e., $\eta \in \sigma(P)$ satisfying $|\eta| = 1$ implies $\eta = 1$.

Ergodicity is related to the fact that it is impossible to decompose the state space into independent parts. The analogue in the theory of Markov processes is *irreducibility* expressing that it is possible to move from (almost) every state to every “relevant” subset within a finite time:

Definition 4.19 ([52, 60]) *A stochastic transition function p is said to be μ -a.e. irreducible if*

$$\mu(A) > 0 \Rightarrow p^m(x, A) > 0 \quad (49)$$

for μ -a.e. $x \in \mathbf{X}$, every $A \in \mathcal{B}(\mathbf{X})$ and some $m = m(x, A) \in \mathbf{Z}_+$. If (49) holds for every $x \in \mathbf{X}$ then p is called μ -irreducible.

The next theorem relates the two statements about indecomposability:

Theorem 4.20 *Let $P : L^1(\mu) \rightarrow L^1(\mu)$ denote a uniformly contractive propagator corresponding to the stochastic transition function p . Then P is ergodic if and only if p is μ -a.e. irreducible.*

Proof: Due to the remark following Def. 4.17, P is ergodic if and only if $P(\mathbf{1}_B/\mu(B))$ converges to $\mathbf{1}_{\mathbf{X}}$ in the sense of Cesàro for every $B \in \mathcal{B}(\mathbf{X})$ with $\mu(B) > 0$. For arbitrary $A \in \mathcal{B}(\mathbf{X})$ with $\mu(A) > 0$ this is equivalent to

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \int_{\mathbf{X}} P^k \mathbf{1}_B(y) \mathbf{1}_A(y) \mu(dy) = \mu(A) \mu(B) \\ \Leftrightarrow & \lim_{n \rightarrow \infty} \int_B \frac{1}{n} \sum_{k=1}^n p^k(y, A) \mu(dy) = \int_B \mu(A) \mu(dy) \\ \Leftrightarrow & \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n p^k(y, A) = \mu(A); \quad \mu\text{-a.e.}, \end{aligned}$$

where we used Lebesgue’s dominated convergence theorem. Since by assumption $\mu(A) > 0$, this is equivalent to μ -a.e. irreducibility according to Def. 4.19. \square

Often, we are interested in dynamical systems—deterministic or stochastic—that exhibit a *unique* invariant density and guarantee that for every initial density v the iterates $P^n v$ converge to the invariant density. In view of Corollary 4.18, these systems are necessarily connected to ergodic propagators, but due to possible cyclic behavior, ergodicity is not sufficient.

Definition 4.21 ([46, Chap. 5.6]) *A propagator $P : L^1(\mu) \rightarrow L^1(\mu)$ is called **asymptotically stable** if*

$$\lim_{n \rightarrow \infty} \|P^n v - \mathbf{1}_{\mathbf{X}}\|_1 = 0 \quad (50)$$

for every density $v \in L^1(\mu)$.

Define the **limit propagator** $P_\infty : L^1(\mu) \rightarrow L^1(\mu)$ by

$$P_\infty v(y) \equiv \int_{\mathbf{X}} v(x) \mu(dx) \quad (51)$$

for arbitrary $v \in L^1(\mu)$, which corresponds to the projection onto the eigenspace spanned by $\mathbf{1}_{\mathbf{X}}$. In terms of P_∞ we can restate (50) in the equivalent form: $\lim_{n \rightarrow \infty} \|P^n v - P_\infty v\|_1 = 0$ for $v \in L^1(\mu)$. Finally, we get [34]

Corollary 4.22 *Let $P : L^1(\mu) \rightarrow L^1(\mu)$ denote a uniformly contractive propagator. Then P is asymptotically stable if and only if P is ergodic and aperiodic. In either case,*

$$\|P^n - P_\infty\|_1 \leq Mq^n \quad n \in \mathbf{Z}_+$$

for some constants $q < 1$ and $M < \infty$.

An analogous result to Cor. 4.22 for the backward transfer operator is well established in the theory of Markov chains. It is related to a property of the stochastic transition function called uniform ergodicity [52]. To state it, we introduce the **total variation norm** on measures:

$$\|\nu\|_{\text{TV}} = \sup_{|u| \leq 1} \int_{\mathbf{X}} u(y) \nu(dy).$$

Definition 4.23 *A stochastic transition function p is said to be μ -a.e. **uniformly ergodic** if*

$$\|p^n(x, \cdot) - \mu\|_{\text{TV}} \leq Mq^n \quad n \in \mathbf{Z}_+ \quad (52)$$

for μ -a.e. $x \in \mathbf{X}$ and some constants $q < 1$ and $M < \infty$. If (52) holds for every $x \in \mathbf{X}$ then p is called **uniformly ergodic**.

In terms of the backward transfer operator and its **limit backward transfer operator** $T_\infty : L^\infty(\mu) \rightarrow L^\infty(\mu)$ defined by

$$T_\infty u(x) \equiv \int_{\mathbf{X}} u(y) \mu(dy),$$

we can restate (52) in the equivalent form $\lim_{n \rightarrow \infty} \|T^n - T_\infty\|_\infty = 0$. Exploiting the duality $P_\infty^* = T_\infty$, we can relate asymptotically stable propagators and μ -a.e. uniformly ergodic stochastic transition functions as follows:

Theorem 4.24 *Let $P : L^1(\mu) \rightarrow L^1(\mu)$ denote some propagator. Then P is uniformly contractive and asymptotically stable if and only if its corresponding stochastic transition function p is μ -a.e. uniformly ergodic.*

Proof: The result follows from the fact that μ -a.e. uniform ergodicity is equivalent to $\lim_{n \rightarrow \infty} \|T^n - T_\infty\|_\infty = 0$, which due to duality is equivalent to uniform constrictiveness and asymptotic stability due to Cor. 4.22. \square

As a result, we can reformulate the two conditions (C1) and (C2) imposed on the propagator P_τ regarded to act on $L^1(\mu)$ in the equivalent form:

(C1) The propagator P_τ is uniformly constrictive. Equivalently, the stochastic transition function $p(x, A) = p(\tau, x, A)$ fulfills the μ -a.e. Doeblin-condition.

(C2) Condition (C1) holds and P_τ is asymptotically stable.

Moreover, the propagator P_τ satisfies conditions (C1) and (C2) if the stochastic transition function is μ -a.e. uniformly ergodic. Since the reformulated conditions are stated in the language of Markov operators and Markov processes, we can exploit the rich literature on these topics (see [46, 52] and cited reference therein) to verify the conditions (C1) and (C2) for different model systems in Section 6.

4.4 Reversibility and Properties in $L^2(\mu)$

The basic idea in analyzing reversible propagators on $L^2(\mu)$ will be to follow along the lines of the $L^1(\mu)$ approach. In doing so, compact operators will replace the role previously played by weakly compact operators. Both cases are special situations of a much more general Δ -calculus introduced by Schechter [67] in 1972. His aim was to study strictly singular operators⁹, which play an important role as admissible perturbations of Fredholm operators¹⁰ [39, 79]. These, moreover, are closely related to essential spectra and in particular to the essential spectral radius [83]. Schechter introduced his quantity for an arbitrary bounded linear operator on some Banach space. For the $L^1(\mu)$ case, Weis proved in [81] the very useful identity $\Delta(P) = \limsup_{\mu(A) \rightarrow \infty} \|\mathbf{1}_A \circ P\|_1$, which played the key role for the subsequent analysis in Section 4.2. As we will see, this characterization of Δ does unfortunately not carry over to $L^2(\mu)$ in general—but it remains true for integral operators [80].

Before we start studying propagators on $L^2(\mu)$, we want to recall that due to Hölder's inequality we have $\|v\|_1 \leq \|v\|_2$ for every $v \in L^2(\mu)$. Hence,

⁹A closed bounded linear operator $P : E \rightarrow E$ on some Banach space E is called *strictly singular*, if it does not possess a bounded inverses on any infinite dimensional subspace M of E [32]. Equivalently, the existence of some constant $\gamma > 0$ such that $\|Px\| \geq \gamma\|x\|$ for every $x \in M \subset E$ implies that M is finite dimensional [39, Chapter 4.5].

¹⁰For a definition see footnote on page 31.

any convergence rate obtained in $L^2(\mu)$ will imply the same rate in the $L^1(\mu)$ norm, when restricted to square integrable functions, i.e., whenever $\|P^n v - P_\infty v\|_2 \leq Mq^n$ holds, then also

$$\|P^n v - P_\infty v\|_1 \leq Mq^n$$

for every $v \in L^2(\mu)$. This way we obtain probabilistic interpretations of results established in $L^2(\mu)$.

Theorem 4.25 ([83]) *Let $P : L^2(\mu) \rightarrow L^2(\mu)$ denote a bounded linear operator. Define the semi-norm $\Delta(P)$ according to*

$$\Delta(P) = \min \{ \|P - S\|_2 : S \text{ is compact} \}.$$

Then the essential spectral radius of P is characterized by

$$r_{\text{ess}}(P) = \lim_{n \rightarrow \infty} \Delta(P^n)^{1/n}. \quad (53)$$

In particular, $r_{\text{ess}}(P) \leq \Delta(P)$. If additionally P is positive¹¹ and self-adjoint, then $r_{\text{ess}}(P) = \Delta(P)$.

Note that Corollary 4.3 applies to our situation, hence $r_{\text{ess}}(P) < 1$ if and only if P is quasi-compact. This was the path followed in [68] by Schütte to prove that the essential spectral radius is less than 1. Our aim in the following is to relate the property of quasi-compactness and hence $r_{\text{ess}}(P) < 1$ to properties of the stochastic transition function and the corresponding Markov process. We start by giving a characterization of compact operators comparable to Theorem 4.7. To do so, we have to introduce the notion of compactness in measure.

Definition 4.26 ([43, Chapter 1.3.3]) *Let $S : L^2(\mu) \rightarrow L^2(\mu)$ denote a bounded linear operator. Then S is called **compact in measure** if it maps weakly convergent sequences to sequences converging in measure. More precisely, if $\{f_n\}_{n \in \mathbf{Z}_+} \subset L^2(\mu)$ is weakly convergent, then for every $\epsilon > 0$ there is $n_0 \in \mathbf{Z}_+$ such that $\mu(\{|Sf_n - Sf_m| \geq \epsilon\}) < \epsilon$ for every $n, m > n_0$.*

An important class of operators being compact in measure are positive integral operators [43], and hence all propagators corresponding to absolutely continuous transition functions. We are now able to give a characterization of compact operators in terms of the probability measure μ .

¹¹Here, positivity is understood in the Markov operator sense: $Pv \geq 0$ if $v \geq 0$ as stated on page 10. This is different from positivity of self-adjoint operators on a Hilbert space: $\langle v, P_t v \rangle_\mu \geq 0$ for every v .

Lemma 4.27 ([43, Thm. 3.1]) *Let $S : L^2(\mu) \rightarrow L^2(\mu)$ denote a bounded linear operator. Then S is compact, if and only if it is compact in measure and satisfies*

$$\limsup_{\mu(A) \rightarrow 0} \|\mathbf{1}_A \circ P\|_2 = 0, \quad (54)$$

where the limit is understood to be taken over all sequences of subsets whose μ -measure converges to zero and $\mathbf{1}_A$ is interpreted as a multiplication operator: $(\mathbf{1}_A v)(x) = \mathbf{1}_A(x)v(x)$.

Weis proved that for an arbitrary *integral* operators, the expression of the left hand side of (54) is identical to the Δ semi-norm and therefore allows to bound the essential spectral radius.

Theorem 4.28 ([80]) *Let $P : L^2(\mu) \rightarrow L^2(\mu)$ denote a bounded linear integral operator. Then*

$$\Delta(P) = \limsup_{\mu(A) \rightarrow 0} \|\mathbf{1}_A \circ P\|_2. \quad (55)$$

In particular,

$$\limsup_{\mu(A) \rightarrow 0} \|\mathbf{1}_A \circ P\|_2 = 0,$$

if and only if P is compact.

As in the $L^1(\mu)$ case, we now want to link the results concerning the Δ semi-norm to properties of the stochastic transition function, in terms of which the propagator is defined. The next lemma is comparable to Cor. 4.8.

Lemma 4.29 *Consider the reversible propagator $S : L^2(\mu) \rightarrow L^2(\mu)$ defined by*

$$Sv(y) = \int_{\mathbf{X}} v(x)p(x,y)\mu(dx) \quad (56)$$

associated with some absolutely continuous stochastic transition function $p(x, dy) = p(x,y)\mu(dy)$, and assume that p is jointly measurable in x and y . Then, S is compact, if the stochastic transition function satisfies the **Kontorovic condition**:

there exist $1 \leq r, s \leq \infty$ with $1/r + 1/s = 1$ such that $\|p(x, \cdot)\|_s \in L^r(\mu)$ as a function of x , i.e.,

$$\int_{\mathbf{X}} \int_{\mathbf{X}} p(x,y)^s \mu(dy)^{r/s} \mu(dx) < \infty. \quad (57)$$

In addition, S is compact if the stochastic transition function satisfies the condition $p(\cdot, \cdot) \in L^r(\mu \times \mu)$ for some $2 \leq r \leq \infty$.

Proof: The first statement is due to Theorem 7.2 of Krasnoselskii et al. [43, Chapter 2], where we have to choose $\tau = (1/2 - 1/r)/(1/s - 1/r)$ for $r \neq s$ and $\tau = 1/2$ for $r = s$. The second statement is a consequence of the first and Hölder's inequality, since $L^r(\mu \times \mu) \subset L^2(\mu \times \mu)$ for $2 \leq r \leq \infty$. \square

For $s = r = 2$ the Kontorovic condition is equivalent to the statement that S is a Hilbert–Schmidt operator, which is known to be compact [6].

Due to investigations initiated by Roberts and Rosenthal [62], quasi-compactness of P is related to certain stability properties of Markov processes. To state them, let $M : \mathbf{X} \rightarrow \mathbf{R}_+$ denote an integrable function, i.e., $M \in L^1(\mu)$ and define the induced M -norm on measures by

$$\|\nu\|_M = \sup_{|v| \leq M} \left| \int_{\mathbf{X}} v(x) \nu(dx) \right|,$$

where $|v| \leq M$ is understood to hold pointwise for every $x \in \mathbf{X}$. For the special case $M \equiv 1$, the M -norm coincides with the total variation norm.

Definition 4.30 *Let p denote some stochastic transition function. Then*

(i) *p is called μ -a.e. geometrically ergodic if*

$$\|p^n(x, \cdot) - \mu\|_{\text{TV}} \leq M(x)q^n; \quad n \in \mathbf{Z}_+ \quad (58)$$

for μ -a.e. $x \in \mathbf{X}$, some constant $q < 1$, and some function $M : \mathbf{X} \rightarrow \mathbf{R}$ satisfying $M < \infty$ pointwise.

If inequality (58) holds for every $x \in \mathbf{X}$ and some function $M \in L^1(\mu)$, then p is called geometrically ergodic.

(ii) *p is called V -uniformly ergodic¹² if*

$$\|p^n(x, \cdot) - \mu\|_M \leq CM(x)q^n; \quad n \in \mathbf{Z}_+$$

for every $x \in \mathbf{X}$, constants $q < 1$ and $C \leq \infty$, and some function $M \in L^1(\mu)$ satisfying $1 \leq M$ pointwise.

The relation between the stability properties defined above is as follows: By definition, V -uniform ergodicity implies geometric ergodicity, which in turn implies μ -a.e. geometric ergodicity. On the other hand, for irreducible and aperiodic stochastic transition functions μ -a.e. geometric ergodicity implies V -uniform ergodicity according to [62, Prop. 2.1]. We now get the following important result:

¹²The notion V -uniform ergodicity is due to the fact that the function M involved in its definition is usually called V . However, in this thesis we already used V to denote the potential energy function.

Theorem 4.31 *Let $P : L^2(\mu) \rightarrow L^2(\mu)$ denote a reversible propagator. Then P satisfies conditions (C1) and (C2) in $L^2(\mu)$, if and only if its stochastic transition function is μ -irreducible and μ -a.e. geometrically ergodic. The latter two conditions on the stochastic transition function p are particularly satisfied, if p is geometrically or V -uniformly ergodic.*

Proof: If P satisfies the two conditions (C1) and (C2), then p is μ -a.e. geometrically ergodic due to Theorem 1 of [63]. On the other hand if p is reversible, μ -irreducible and μ -a.e. geometrically ergodic, then P satisfies the conditions (C1) and (C2) as an immediate result of Theorem 2 of [63] and Theorem 2.1 of [62]. The second statement follows directly from the remark preceding the theorem. \square

The assumption of μ -irreducibility of the stochastic transition function in Theorem 4.31 seems to be artificial. One would rather expect μ -a.e. irreducibility, which furthermore would be a consequence of μ -a.e. geometric ergodicity. Hence, we expect Theorem 4.31 to hold without the assumption of μ -irreducibility. For reversible propagators we finally get the following relation between the conditions (C1) and (C2) in $L^1(\mu)$ and those in $L^2(\mu)$:

Theorem 4.32 *Let $P : L^1(\mu) \rightarrow L^1(\mu)$ denote some propagator satisfying conditions (C1) and (C2) in $L^1(\mu)$. If P is reversible and its stochastic transition function is μ -irreducible then $P : L^2(\mu) \subset L^1(\mu) \rightarrow L^2(\mu)$ also satisfies the conditions (C1) and (C2) in $L^2(\mu)$.*

Proof: In $L^1(\mu)$ the conditions (C1) and (C2) are equivalent to μ -a.e. uniform ergodicity of the associated Markov process (see Theorem 4.24). Since μ -a.e. uniform ergodicity implies μ -a.e. geometric ergodicity, P satisfies (C1) and (C2) in $L^2(\mu)$ due to Theorem 4.31. \square

We finally obtain the useful

Corollary 4.33 *If $P : L^r(\mu) \rightarrow L^r(\mu)$ with $r = 1, 2$ is reversible and its stochastic transition function p is uniformly ergodic, then P satisfies the conditions (C1) and (C2) both in $L^1(\mu)$ and $L^2(\mu)$.*

As a result of this section, we can state the conditions (C1) and (C2) in a more probabilistic language. Particularly, Theorem 4.31 will be very useful when verifying conditions (C1) and (C2) for new model systems.

5 Discretization of Transfer Operators

If we want to identify metastable subsets we have to compute certain eigenfunctions of the propagator P_τ . In the following we describe the discretization procedure of the eigenvalue problem $P_\tau v = \lambda v$. Throughout this section we assume that P_τ satisfies the conditions (C1) and (C2) defined in Section 3.2. Part of this section follows from [68, 69].

5.1 Galerkin Discretization

Let $\mathcal{D} = \{D_1, \dots, D_n\}$ denote a decomposition of the state space and define the associated finite dimensional ansatz space by $\mathcal{V}_n = \text{span}\{\mathbf{1}_{D_1}, \dots, \mathbf{1}_{D_n}\}$. Then, the **Galerkin projection** $\Pi_n : L^1(\mu) \rightarrow \mathcal{V}_n$ of $v \in L^1(\mu)$ is defined by

$$\Pi_n v = \sum_{k=1}^n \frac{\langle v, \mathbf{1}_{D_k} \rangle_\mu}{\langle \mathbf{1}_{D_k}, \mathbf{1}_{D_k} \rangle_\mu} \mathbf{1}_{D_k},$$

where $\langle \cdot, \cdot \rangle_\mu$ is the duality bracket between $L^1(\mu)$ and $L^\infty(\mu)$. The resulting discretized propagator $\Pi_n P_\tau \Pi_n$ induces an approximate eigenvalue problem $\Pi_n P_\tau \Pi_n v = \lambda \Pi_n v$ in \mathcal{V}_n . Using $v = \sum_{k=1}^n \nu_k \mathbf{1}_{D_k}$, the discretized eigenvalue problems reads in coordinate representation

$$\sum_{l=1}^n \langle \mathbf{1}_{D_l}, P_\tau \mathbf{1}_{D_k} \rangle_\mu \nu_l = \lambda \langle \mathbf{1}_{D_k}, \mathbf{1}_{D_k} \rangle_\mu \nu_k \quad (59)$$

for $k = 1, \dots, n$. After division of (59) by $\langle \mathbf{1}_{D_k}, \mathbf{1}_{D_k} \rangle_\mu = \mu(D_k) > 0$, we obtain the convenient form

$$S\nu = \lambda\nu$$

with $\nu = (\nu_1, \dots, \nu_n) \in \mathbf{C}^n$ and $n \times n$ stochastic transition matrix $S = (S_{kl})$, whose entries are given by the one-step transition probabilities from D_k to D_l within the time τ :

$$S_{kl} = \frac{\langle P_\tau \mathbf{1}_{D_k}, \mathbf{1}_{D_l} \rangle_\mu}{\langle \mathbf{1}_{D_k}, \mathbf{1}_{D_k} \rangle_\mu} = p(\tau, D_k, D_l). \quad (60)$$

Since P_τ is a Markov operator, its Galerkin discretization S is a (row) stochastic matrix, i.e., $S_{kl} \geq 0$ and $\sum_{l=1}^n S_{kl} = 1$ for every $k = 1, \dots, n$. Hence, all its eigenvalues λ satisfy $|\lambda| \leq 1$. Moreover, we have the following three important properties [68, 69]:

- (i) The row vector $\pi = (\pi_1, \dots, \pi_n)$ with $\pi_k = \mu(D_k)$ represents the discretized invariant probability measure μ . It is a left eigenvector corresponding to the eigenvalue $\lambda = 1$, i.e., $\pi S = \pi$.

- (ii) S is *irreducible and aperiodic*. As a consequence, the eigenvalue $\lambda = 1$ is *simple and dominant*. In particular, the discretized invariant density π is the *unique* invariant density of S .
- (iii) If P_τ is reversible then S is *self-adjoint* w.r.t. the discrete scalar product $\langle u, v \rangle_\pi = \sum u_i \bar{v}_i \pi_i$. Equivalently, S satisfies the *detailed balance condition* $\pi_k S_{kl} = \pi_l S_{lk}$ for every $k, l \in \{1, \dots, n\}$. Consequently, all eigenvalues of S are real-valued and contained in the interval $[-1, 1]$.

The discretization of the propagator can be interpreted as a *coarse graining* procedure: Coarse graining the state space $\{x \in \mathbf{X}\} \rightarrow \{D_1, \dots, D_n\}$ results in a coarse graining of the propagator $P_\tau \rightarrow S$ corresponding to a coarse graining of the Markov process $p(\tau, x, C) \rightarrow p(\tau, D_k, D_l)$ with invariant measures $\mu \rightarrow \pi$. In doing so, the discretization inherits the most important properties of the propagator.

In numerical experiments, it is desirable to estimate the essential spectral radius $r_{\text{ess}}(P_\tau)$. Since $r_{\text{ess}}(P_\tau) \leq \Delta(P_\tau)$, we suggest the following heuristics to define an indicator $[\Delta(P_\tau)]$ for some upper bound on $r_{\text{ess}}(P_\tau)$. The basic idea is to use a decomposition $\mathcal{D} = \{D_1, \dots, D_n\}$ of the state space and a “discretized” version of

$$\Delta(P_\tau) = \limsup_{\mu(A) \rightarrow 0} \sup_{0 \neq v \in L^1(\mu)} \frac{1}{\|v\|_1} \|\mathbf{1}_A \circ P_\tau v\|_1.$$

Hence, replacing suprema by maxima w.r.t. the decomposition \mathcal{D} we get

$$\begin{aligned} [\Delta(P_\tau)] &= \max_{D \in \mathcal{D}} \max_{0 \neq v \in \mathcal{V}_n} \frac{1}{\|v\|_1} \|\mathbf{1}_D \circ P_\tau v\|_1 \\ &= \max_{j,k} \frac{1}{\mu(D_k)} \|\mathbf{1}_{D_j} \circ P_\tau \mathbf{1}_{D_k}\|_1 \\ &= \max_{j,k} S_{jk}. \end{aligned}$$

Therefore, the maximal entry of the stochastic transition matrix S , obtained from discretizing the propagator P_τ , can be used as an indicator for $r_{\text{ess}}(P_\tau)$. In order to better capture the nature of the limit process $\mu(A) \rightarrow 0$, we suggest to use a sequence of decompositions $\mathcal{D}_1, \dots, \mathcal{D}_m$ that get finer and finer, and consider the corresponding sequence of indicators $[\Delta(P_\tau)]_{\mathcal{D}_1}, \dots, [\Delta(P_\tau)]_{\mathcal{D}_m}$. A proof about the reliability of the indicator seem to be possible under additional regularity conditions on the stochastic transition function. However, it should be clear that these regularity conditions conflict the fact that a non-vanishing essential spectral radius is related to singular and therefore irregular behavior of the underlying dynamics (see Theorem 4.9 and Lemma 4.29).

For a fixed decomposition, the indicator may still advantageous be exploited in a hierarchical context to indicate regions of further refinement: If $[\Delta(P_\tau)] = S_{jk} \approx 1$ for some pair (j, k) , then we may suggests

- (i) a refinement of the sampling corresponding to the j th box, since the approximation quality for the j th box was too bad, or
- (ii) a further subdivision of the k th box, since the statistical weight $\mu(D_k)$ was too big.

If neither of the two strategies decreases the value of $[\Delta(P_\tau)]$, then the essential spectral radius may indeed be close to 1. The indicator works quite well for Markovian systems, as we are going to demonstrate it in Section 6; it has also already been successfully applied to biomolecular systems [7]. As it was pointed out by G. Froyland [24], it might be less useful for deterministic systems, in particular for hyperbolic deterministic systems, where we would expect the maximal entry of S to be approximately 0.5 due to expansion and contraction of the dynamics.

5.2 Convergence of Discrete Eigenvalues

We restrict our considerations to the important class of reversible propagators $P_\tau : L^2(\mu) \rightarrow L^2(\mu)$ satisfying the conditions (C1) and (C2). Under these assumptions, convergence results have been proved (see [68] for details).

Denote by $\sigma(P_\tau)$ the spectrum of P_τ in $L^2(\mu)$ and by $\sigma_{\text{discr}}(P_\tau) \subset \sigma(P_\tau)$ the discrete spectrum. We are interested in approximating a cluster of (real-valued) discrete eigenvalues $\lambda_c, \dots, \lambda_1 \in \sigma_{\text{discr}}(P_\tau)$ close to 1 and “outside” the disc with radius $r_{\text{ess}}(P_\tau)$. Assume that the eigenvalues are repeated according to their multiplicity with

$$r_{\text{ess}}(P_\tau) < \lambda_c \leq \dots \leq \lambda_2 < \lambda_1 = 1.$$

and corresponding eigenfunctions v_c, \dots, v_1 , orthogonal w.r.t. $\langle \cdot, \cdot \rangle_\mu$. Furthermore, we require that the sequence of the Galerkin ansatz spaces $\mathcal{V}_1 \subset \mathcal{V}_2 \subset \dots$ is dense in $L^2(\mu)$ and the corresponding decompositions $\mathcal{D}_1, \mathcal{D}_2, \dots$ are getting finer and finer, i.e., $\max_{D \in \mathcal{D}_n} \text{diam}(D) \rightarrow 0$ as $n \rightarrow \infty$. Denote by $S(\mathcal{V}_n)$ the stochastic transition matrix obtained from discretizing the propagator P_τ w.r.t. the ansatz space \mathcal{V}_n . Furthermore, denote the eigenvalues and corresponding eigenvectors of $S(\mathcal{V}_n)$ by $\lambda_i(\mathcal{V}_n)$ and $u_i(\mathcal{V}_n)$, respectively (ordered in decreasing magnitude and taken into account multiplicity). Under these assumptions the dominant eigenvalues of $S(\mathcal{V}_n)$ are good approximations of the dominant eigenvalues of P_τ , whenever the discretization is fine enough. In this case $S(\mathcal{V}_n)$ also has a cluster of eigenvalues $\lambda_c(\mathcal{V}_n) \leq \dots \leq \lambda_2(\mathcal{V}_n) < \lambda_1(\mathcal{V}_n) = 1$ close to 1. More precisely, for every $j = 1, \dots, c$, we get

$$\lambda_j(\mathcal{V}_n) \longrightarrow \lambda_j \quad \text{and} \quad u_j(\mathcal{V}_n) \longrightarrow u_j$$

in modulus and in the $L^2(\mu)$ -norm, respectively, as $n \rightarrow \infty$ [68].

5.3 Evaluating the Stochastic Transition Matrix

We consider the evaluation of the stochastic transition matrix S obtained from discretizing P_τ .

Consider two elements B, C of some decomposition of the state space. Combining $p(\tau, x, C) = \mathbf{E}_x[\mathbf{1}_C(X_\tau)]$ with Eq. (32) we get

$$p(\tau, B, C) = \frac{1}{\mu(B)} \int_B \mathbf{E}_x[\mathbf{1}_C(X_\tau)] \mu(dx), \quad (61)$$

which can be approximated within two steps:

(A1) approximation of the integral

$$\int_B g(x) \mu(dx) \approx \sum_{k=1}^N \alpha_k g(x_k)$$

by some deterministic or stochastic integration scheme with partition points or random variables x_1, \dots, x_N , respectively, and weights $\alpha_1, \dots, \alpha_N$ [16, 29];

(A2) approximation of the expectation value

$$\mathbf{E}_x[\mathbf{1}_C(X_\tau)] \approx \frac{1}{M} \sum_{j=1}^M \mathbf{1}_C(X_\tau(\omega_j, x))$$

by relative frequencies, where $X_\tau(\omega_k, x)$ denotes a realization of the Markov process at time τ with initial distribution $X_0 \sim x$ [52, Chapter 17].

A combination of the two steps (A1) and (A2) with $g(x) = \mathbf{E}_x[\mathbf{1}_C(X_\tau)]$ results in

$$p(\tau, B, C) \approx \frac{1}{M} \sum_{k=1}^N \sum_{j=1}^M \alpha_k \mathbf{1}_C(X_\tau(\omega_{kj}, x_k));$$

hence, for each initial point x_k , the Markov process X_τ is realized M times. This allows us to approximate the entries of the stochastic transition matrix S due to $S_{jk} = p(\tau, D_j, D_k)$. The approximation quality of $p(\tau, B, C)$ depends on the interplay between the two approximation steps (A1) and (A2). Numerical experiments in low dimensions show that it is even possible to take $M = 1$, if the number of partition points N is chosen in such a way that the number of points per subset of the decomposition \mathcal{D} is reasonable large. For high-dimensional problems, we in general will be forced to use stochastic integration schemes, such as Monte Carlo methods, to approximate the integral in (A1); for further details see Section 7.1. For low-dimensional problems, we may also use deterministic integration schemes:

Example 5.1 Assume that $\mathbf{X} = [a, b] \subset \mathbf{R}$ and the invariant measure is absolutely continuous w.r.t. the Lebesgue measure, i.e., $\mu(dx) = f(x)dx$ for some density f . For $N \in \mathbf{Z}_+$ define the partition points $x_k = a + kh$ with $k = 1, \dots, N$ and $h = (b - a)/N$. Using the Trapezoid rule in (A1), and M realization of the Markov process in (A2) we get

$$p(\tau, B, C) \approx \frac{1}{M \#[x_k \in B]} \sum_{j,k} \mathbf{1}_B(x_k) \mathbf{1}_C((X_\tau(\omega_j, x_k))) f(x_k),$$

where $\#[x_k \in B]$ denotes the total number of partition points x_k in B .

Already in moderately low dimensions the strategy presented in Example 5.1 may cause serious memory problems and unacceptable numerical effort. Here the adaptive discretization technique by Dellnitz and Junge [12, 13] can be of significant use. Developed to study in particular hyperbolic dynamical systems, its successful application to a small molecular system is described in [14]. For the analysis of biomolecules, e.g, small peptides, a different approach has to be chosen, as will be outlined in Section 7. Performing a realization of the discrete time Markov process $X_n = \{X_n\}_{n \in \mathbf{Z}_+}$, we may also combine the two approximation steps (A1) and (A2). Recall that we defined for some fixed $\tau > 0$ the time-reversed Markov process $Y_n = \{Y_n\}_{n \in \mathbf{Z}_+}$ with stochastic transition function q_τ and the time-symmetrized Markov process $Z_n = \{Z_n\}_{n \in \mathbf{Z}_+}$ with stochastic transition function r_τ , see Section 3.3.

Example 5.2 Let x_0, \dots, x_N denote a sequence of sampling points obtained from a realization of the discrete time Markov process X_n . Then

$$p(\tau, B, C) \approx \frac{\#[x_k \in B \text{ and } x_{k+1} \in C]}{\#[x_k \in B]}, \quad (62)$$

where convergence is guaranteed for μ -a.e. initial points x_0 by conditions (C1) and (C2) and the law of large numbers [52]. Since the reversed sampling x_N, \dots, x_0 is a realization of the time-reversed Markov process Y_n with stochastic transition function q_τ , we have

$$q_\tau(\mathbf{1}, B, C) \approx \frac{\#[x_k \in B \text{ and } x_{k-1} \in C]}{\#[x_k \in B]}.$$

Hence, we can approximate the stochastic transition function r_τ corresponding to the time-symmetrized Markov process Z_n by

$$r_\tau(\mathbf{1}, B, C) \approx \frac{\#[x_k \in B \text{ and } x_{k+1} \in C] + \#[x_k \in B \text{ and } x_{k-1} \in C]}{2 \#[x_k \in B]}.$$

For a reversible Markov process the identity $q_\tau(\mathbf{1}, B, C) = p_\tau(\mathbf{1}, B, C)$ and consequently $r_\tau(\mathbf{1}, B, C) = p_\tau(\mathbf{1}, B, C)$ holds. Thus, for reversible Markov processes we may “double” the information by considering both the original as well as the reversed sampling.

Intuitively, it should be clear that the technique described in Example 5.2 becomes less efficient, if the Markov process admits a decomposition of the state space into very metastable subsets. Then, convergence to equilibrium is very slow so that we have to use a different approach (see Section 7).

5.4 The Numerical Identification Algorithm

We briefly outline the identification algorithm, a detailed description can be found in Deuffhard et al. [17].

The aim is to completely decompose the state space \mathbf{X} into metastable subsets. Hence, given a decomposition $\mathcal{D} = \{D_1, \dots, D_n\}$ of \mathbf{X} , we have to determine a clustering $\{C_1, \dots, C_c\}$ into c metastable clusters by assigning each D_j with $j = 1, \dots, n$ to some cluster C_k with $k = 1, \dots, c$. This is done by exploiting the *almost constant level structure* of the dominant eigenvectors, which shows up in Theorem 3.1. Denote by v_1, \dots, v_c the eigenvectors corresponding to a cluster of c eigenvalues close to 1. Then, they are almost constant on each metastable subset, i.e., if D_i and D_j belong to the same metastable subset, then $v_k(D_i) \approx v_k(D_j)$ for $k = 1, \dots, c$. Associate to each subset D_j the c -tuple of eigenvector components

$$D_j \longmapsto (v_1(D_j), \dots, v_c(D_j))$$

and define a clustering $\{C_1, \dots, C_c\}$ by collecting subsets D_j with almost identical c -tuples into the same cluster. Then, as it is shown in [17], this is sufficient to define a clustering into metastable clusters in the case of weak coupling. The identification of metastable clusters is reduced to cluster c -tuples w.r.t. geometrical similarity. We have implemented an algorithm, which also copes with larger perturbations in the eigenvector components due to stronger coupling between metastable subsets; for details see [17].

6 Theoretical and Numerical Investigations

Based on the results of the last two sections, we want to theoretically and numerically analyze our four model systems. We will see that under suitable conditions on the potential energy function all model systems but the deterministic Hamiltonian system satisfy the two basic conditions (C1) and (C2). The application to a three-well potential show significant similarities for the model systems on a mesoscopic level.

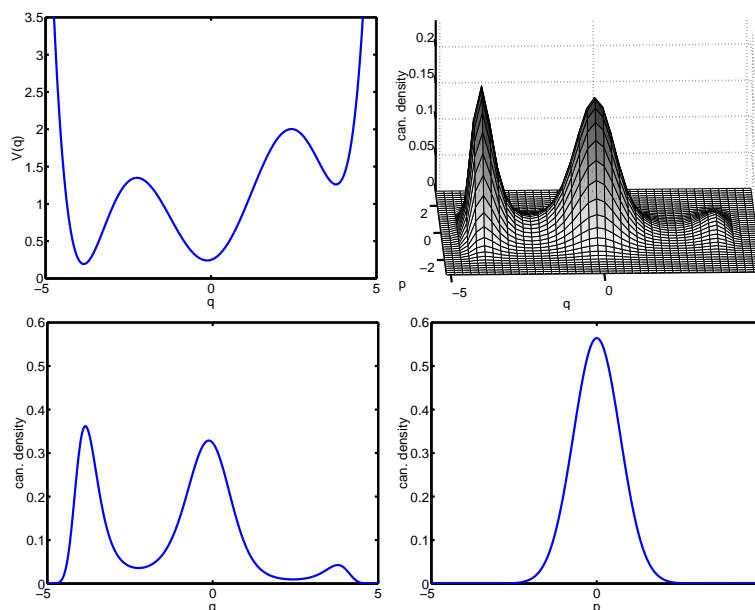


Figure 5: Three well potential (top left) and corresponding canonical density f_{can} for $\beta = 2$ (top right). The positional canonical density Q (bottom left) and canonical density of momenta \mathcal{P} (bottom right).

In the following we consider the three well potential $V : \Omega \rightarrow \mathbf{R}$ with

$$V(q) = \frac{1}{400}(q^6 - 30q^4 + 238q^2 + 56q + 100) \quad (63)$$

as our test system. We choose $\Omega = [-5, 5]$ as the position space with periodic boundary conditions, and modify the potential function at the boundary ± 5 in such a way that it is smooth (cf. Remark below). Its graph, the canonical distribution f_{can} corresponding to $\beta = 2$ and its two marginal distributions Q and \mathcal{P} are shown in Figure 5. We choose $\tau = 1$ as the observation time span. Intuitively, we would expect to exist three metastable subsets around the (local) minima of the potential function for moderate temperature.

Remark. The positional canonical density Q already indicates that it is very unlikely to stay near the periodic boundary at ± 5 ; this can also be

observed from the realization shown in Figure 6. Hence, numerically it might be difficult to resolve the difference between an unbounded position space and one with periodic boundary conditions. Yet, in special situations this is possible, as we are going to exemplify for the Hamiltonian system with randomized momenta in application to a harmonic potential test system (see end of Section 6.2).

6.1 Deterministic Hamiltonian System

Denote by P_τ the propagator corresponding to the deterministic Hamiltonian system.

Proposition 6.1 *For both bounded systems and periodic systems, the propagator $P_\tau : L^r(\mu_{\text{can}}) \rightarrow L^r(\mu_{\text{can}})$ neither satisfies condition (C1) nor condition (C2) in $L^r(\mu_{\text{can}})$ with $r = 1, 2$.*

Proof: For the statement on $r = 1$, we observe that the stochastic transition function $p_\tau(x, dy) = \delta_{\Phi^\tau x}(dy)$ is singular w.r.t. the invariant measure $\mu_{\text{can}}(dy) = f_{\text{can}}(y)dy$ implying $\Delta(P_\tau) = 1$ by Theorem 4.9. Since $p_\tau^n(x, dy) = \delta_{\Phi^{n\tau}x}(dy)$, we more generally have $\Delta(P_\tau^n) = 1$ for every $n \in \mathbf{Z}_+$ and hence $r_{\text{ess}}(P_\tau) = 1$ by Theorem 4.6. This violates condition (C1). Now, define for an arbitrary smooth function $\mathcal{F} : \mathbf{R} \rightarrow \mathbf{R}_+$ the density $f : \mathbf{X} \rightarrow \mathbf{R}$ by $f(x) = \mathcal{F}(H(x))$ and the thereby induced measure μ_f by $\mu_f(dy) = f(y)dy$. Since the Hamiltonian flow is energy-preserving, μ_f is a finite invariant measure for every $f \in L^1(\mu_{\text{can}})$. This way we may construct arbitrarily many invariant probability measures, which violates condition (C2). The statement for $r = 2$ follows from the fact that P_τ is unitary in $L^2(\mu_{\text{can}})$, as stated in Section 2.1. \square

Within the proof of Proposition 6.1 we have shown that the deterministic Hamiltonian system admits infinitely many invariant probability measures. Due to this ambiguity, in our context pure Hamiltonian dynamics seems not to be appropriate for modeling internal fluctuations within one specific stationary ensemble, in our case the canonical ensemble.

6.2 Hamiltonian System with Randomized Momenta

Denote by P_τ the propagator corresponding to the Hamiltonian system with randomized momenta. We first state under which conditions on the potential function the two requirements (C1) and (C2) on P_τ hold. Then, we numerically analyze the induced essential statistical behavior w.r.t. the positional canonical ensemble.

Proposition 6.2 *For periodic systems with position space $\Omega \subset \mathbf{R}^d$, some fixed observation time span $\tau > 0$ and smooth periodic potential function*

$V : \Omega \rightarrow \mathbf{R}$, the propagator $P_\tau : L^r(\mu_Q) \rightarrow L^r(\mu_Q)$ satisfies the conditions (C1) and (C2) in $L^r(\mu_Q)$ with $r = 1, 2$.

Proof: The statement for $r = 1$ is based on results in [68]. Due to [68, Lem. 4.51 and Prop. 4.18] the Lebesgue decomposition of the stochastic transition function $p_\tau(q, dy) = p_a(q, y)\mu_Q(dy) + p_s(q, dy)$ has the following two properties:

- (i) the absolutely continuous part satisfies: $\text{ess sup}_{(q,y) \in \Omega} p_a(q, y) < \infty$.
- (ii) the singular part satisfies: $\text{ess sup}_{q \in \Omega} p_a(q, \Omega) < 1$.

Application of Theorem 4.9 proves that condition (C1) holds. Since P_τ is asymptotically stable according to [68, Lem. 4.51], condition (C2) is a consequence of Corollary 4.22. In order to prove the statement for $r = 2$ we note that due to [68, Lem. 4.31] the Markov process corresponding to the Hamiltonian system with randomized momenta is μ_Q -irreducible. Therefore, the statement is a consequence of Theorem 4.32. Using a different approach, the statement for $r = 2$ was already proved by Schütte in [68]. \square

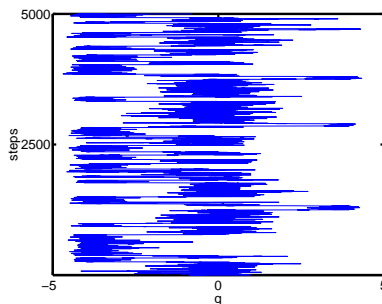


Figure 6: Typical realization of the Hamiltonian system with randomized momenta for $\beta = 2$, observation time span $\tau = 1$ and initial distribution $Q_0 \sim 1$.

By Proposition 6.2 the application of our algorithmic approach to the test systems is theoretically justified. In order to discretize the propagator, we proceed according to Example 5.1 using the Trapezoid rule with $N = 300000$, $M = 1$ and the Leapfrog discretization [65] of the Hamiltonian flow with internal step size $\Delta t = 0.02$. A typical realization is shown in Figure 6. We observe that the Markov process stays for some time close to one of the three (local) minima, then suddenly jumps close to another minimum, stays there for a while, jumps again and so on. Hence, by looking at the realization we visually identify three metastable subsets. Discretizing the state space $\Omega = [-5, 5]$ with periodic boundary conditions into 30 equal-sized intervals, we obtain a 30×30 stochastic transition matrix S . Solving the eigenvalue

problem for S yields:

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	\dots	λ_{30}
1.000	0.975	0.958	0.599	0.490	0.369	\dots	-0.435

Evaluating the indicator for $r_{\text{ess}}(P_\tau)$ we get $[\Delta(P_\tau)] = 0.557$ —for a further analysis of the indicator within a hierarchical context see below. Looking at the spectrum of S , we identify a cluster of three eigenvalues $\{\lambda_1, \lambda_2, \lambda_3\}$ close to 1 that is well separated from the remaining part of the spectrum by a gap. Hence, in view of our algorithmic strategy we look for a decomposition into three metastable subsets. The eigenfunctions corresponding to the largest eigenvalues are shown in Figure 7 (left). We observe *almost constant levels* around the three minima for the first three eigenfunctions, while the fourth eigenfunction does not show this particular structure. This almost constant level structure is exploited by the identification algorithm outlined in Section 5.4. Application to our example yields a clustering $\{C_1, C_2, C_3\}$ with $C_1 = \{q \leq -2.1\}$, $C_2 = \{-2.1 < q \leq 1.8\}$ and $C_3 = \{1.8 < q\}$. The statistical weights $\mu(C_k)$ within the canonical ensemble μ_Q and the metastabilities $p(\tau, C_k, C_k)$ are given by the following table:

metastable subset	C_1	C_2	C_3
statistical weight	0.324	0.616	0.060
metastability	0.966	0.973	0.908

The essential statistical behavior, i.e., the probability of transitions between the metastable subsets, is described by the coupling matrix $C = (c_{jk})_{j,k=1,2,3}$ with $c_{jk} = p(\tau, C_j, C_k)$. For our example, we obtain

$$C = \begin{pmatrix} 0.966 & 0.034 & 0 \\ 0.018 & 0.973 & 0.009 \\ 0 & 0.092 & 0.908 \end{pmatrix}.$$

Analyzing only the coupling matrix C and the metastability of the clusters, we would predict that a typical realization of the Markov process would stay most of the time in C_2 , sometimes moving to C_1 , stay there for some time, then moving back and so on. Rarely, there will be transitions to C_3 , where in addition the Markov process is unlikely to stay for a while. This is what we observed for the realization shown in Figure 6. In this sense the clustering, its metastabilities and the corresponding coupling matrix allow to describe the essential statistical behavior of the model system.

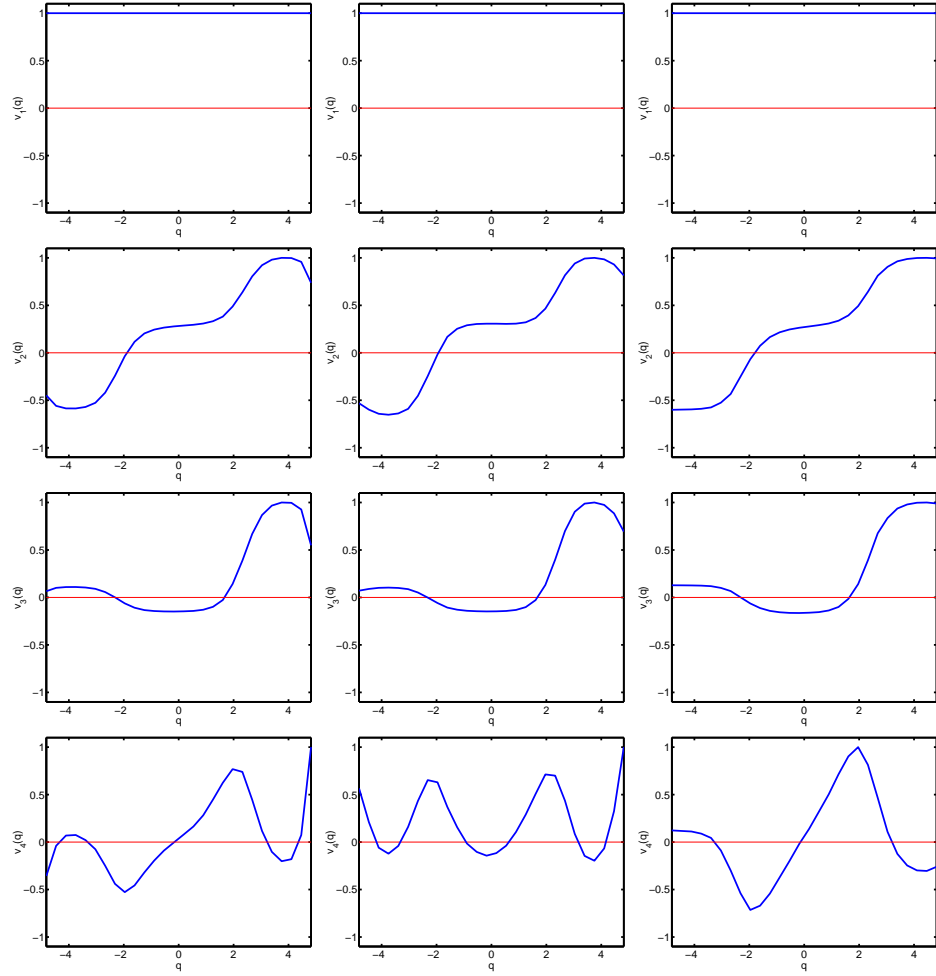


Figure 7: The four dominant eigenfunctions of the propagator P_τ for different model systems. Left: Hamiltonian system with randomized momenta corresponding to the eigenvalues 1.000, 0.975, 0.958, 0.599 (from top to bottom). Middle: Langevin equation for $\gamma = 1.0$ corresponding to the eigenvalues 1.000, 0.969, 0.949, 0.430; eigenfunctions *projected* on the position space. Right: Smoluchowski equation for $\gamma = 1.0$ corresponding to the eigenvalues 1.000, 0.950, 0.915, 0.387.

Let us have a closer look at the indicator $[\Delta(P_\tau)]$. As suggested in Section 5.1, the indicator should be used within some hierarchy of decompositions. The following table shows the indicator for our test system evaluated in a hierarchy of decompositions into m equal-sized intervals:

m	30	50	100	200	500	1000
$[\Delta(P_\tau)]$	0.557	0.423	0.315	0.207	0.152	0.117

We deduce that the essential spectral radius is likely to be much less than the initially value of $[\Delta(P_\tau)] = 0.557$ indicated. Moreover, while $[\Delta(P_\tau)]$ drastically decays for finer decompositions of the state space, the dominant eigenvalues of P_τ are quite insensitive w.r.t. refinements. This was already illustrated for a small molecule in [69]. However, we do not know the exact value of $r_{\text{ess}}(P_\tau)$ in order to validate the results.

This is possible for the systems presented in Example 4.10 (ii), where we considered a harmonic potential on the position space $\Omega = [-1, 1]$ with periodic boundary conditions. Let us choose $\beta = 2$ and consider the corresponding propagator P_τ . Analytically, we have $\Delta(P_\tau) = 0.841$ according to Example 4.10. The following table shows the indicator $[\Delta(P_\tau)]$ based on different decompositions of the state space into m equal-sized intervals:

m	30	50	100	200
$[\Delta(P_\tau)]$	0.846	0.849	0.850	0.859

We observe that a finer decomposition of the state space does not significantly change the value of $[\Delta(P_\tau)]$; the results are in good agreement with the analytical value. For the unbounded state space $\Omega = \mathbf{R}$ described in Example 4.10 (i) we have $P_\tau = \text{Id}$ and hence $r_{\text{ess}}(P_\tau) = 1$. This perfectly illustrates the influence of periodic boundary conditions and also demonstrates that it is possible to resolve the difference to the unbounded state space numerically.

6.3 Langevin Equation

Let P_τ denote the propagator corresponding to the Langevin Markov process and p_τ its stochastic transition function. Up to now, the stability properties of the Langevin Markov process are only partially understood. While (μ -a.e.) geometric or V -uniform ergodicity can be proved for both bounded systems and periodic systems under reasonable conditions on the potential function [51, 76], little is known about (μ -a.e.) uniform ergodicity. It is believed that the Markov process is not (μ -a.e.) uniformly ergodic for bounded systems, while the periodic case is even less understood [33]. These topics are subject to current investigations. Since the Langevin Markov process

is not reversible, it is an ideal test system to demonstrate the effects of time-symmetrization. Hence, we analyze both the “original” as well as the time-symmetrized Markov process.

Proposition 6.3 *The following holds for the Langevin equation:*

- (i) *Fix some observation time space $\tau > 0$ and consider a periodic system with position space $\Omega \subset \mathbf{R}^d$ and smooth potential function $V : \Omega \rightarrow \mathbf{R}$. Then, μ_{can} is the unique invariant probability measure of the Langevin Markov process $X_n = \{X_{n\tau}\}_{n \in \mathbf{Z}_+}$ and its stochastic transition function p_τ is (μ_{can} -a.e.) geometrically and V -uniformly ergodic.*
- (ii) *Fix some observation time space $\tau > 0$ and consider a bounded system with position space $\Omega = \mathbf{R}^d$ and smooth potential function $V : \Omega \rightarrow \mathbf{R}$ satisfying $V(q) \geq 0$ for $q \in \Omega$ and growing at infinity like $\|q\|^{2l}$ for some positive integer l . Then, μ_{can} is the unique invariant probability measure of the Langevin Markov process $X_n = \{X_{n\tau}\}_{n \in \mathbf{Z}_+}$ and its stochastic transition function p_τ is (μ_{can} -a.e.) geometrically and V -uniformly ergodic.*

Proof: Statement (i) is an immediate result of Theorem 3.1 by Stuart [76], while statement (ii) immediately follows from Theorem 3.2 by Mattingly et al. [51]. \square

We remark that according to Mattingly et al. [51] the condition on the growth rate of the potential function in the case of bounded systems can be further weakened (see cond. 3.1 in [51]). While we are able to prove (μ_{can} -a.e.) geometric and V -uniform ergodicity for the Langevin equation, reversibility fails to hold, since the infinitesimal generator \mathcal{L} and therefore the propagator P_τ is not self-adjoint in $L^2(\mu_{\text{can}})$. In this case, Theorem 4.31 states nothing about the validity of the conditions (C1) and (C2) in $L^2(\mu_{\text{can}})$. Although the application of our algorithmic approach is not (yet) theoretically justified, the numerical results presented below are very promising.

In order to discretize the propagator in application to our test system, we proceed according to Example 5.2, hence exploit a realization of the discrete time Markov process X_n . We use the Leapfrog discretization for the deterministic part with internal step size $\Delta t = 0.02$ and random variables $N_n \sim \sigma\sqrt{\Delta t}\mathcal{N}(0, 1)$ for a realization of the white noise (within each internal step). We discretize the state space $\mathbf{X} = \Omega \times \mathbf{R}$ with $\Omega = [-5, 5]$ and periodic boundary conditions into $m = 900$ subsets as follows: discretize the position space $\Omega = [-5, 5]$ into 30 equal-sized intervals and the momenta space \mathbf{R} by partitioning $[-3, +3]$ into 28 equal-sized intervals and adding the two infinite intervals $(-\infty, -3]$ and $[3, \infty)$. Note that in view of the momenta

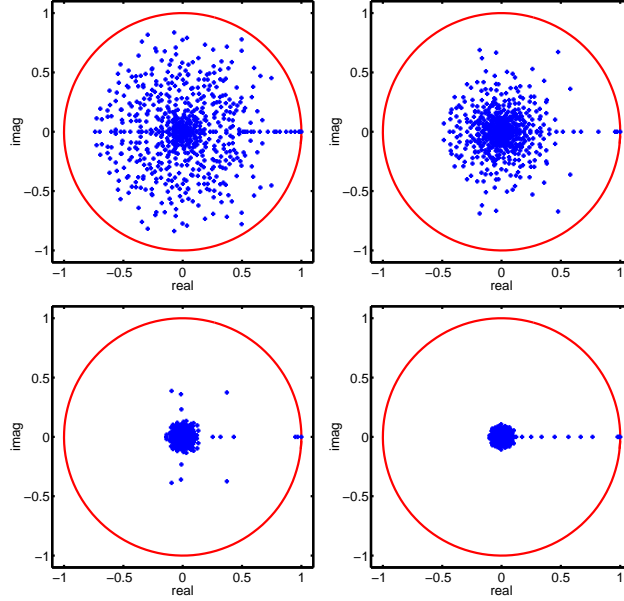


Figure 8: Spectrum of P_τ for the Langevin equation for $\gamma = 0.01$ (upper left), $\gamma = 0.16$ (upper right), $\gamma = 1.0$ (lower left) and $\gamma = 4.0$ (lower right).

distribution shown in Figure 5, the statistical weight of the infinite intervals is negligible. This way, we obtain a 900×900 non self-adjoint stochastic transition matrix S . Solving the eigenvalue problem for S yields

λ_1	λ_2	λ_3	λ_4	$\lambda_{5/6}$...
1.000	0.969	0.949	0.430	$0.369 \pm 0.372i$...

with $[\Delta(P_\tau)] = 0.750$ for $m = 400$ discretization subsets and $[\Delta(P_\tau)] = 0.667$ for $m = 900$. Hence, $[\Delta(P_\tau)]$ indicates that the first three dominant eigenvalues belong to the discrete spectrum and are separated by a spectral gap from the remaining part of the spectrum. This is in agreement with the fact that the dominant eigenvalues are rather insensitive w.r.t. further refinements of the decomposition. The next table show the eigenvalues of *maximal modulus* for different decompositions into m subsets:

m	λ_1	λ_2	λ_3	λ_4
400	1.000	0.968	0.947	0.412
900	1.000	0.969	0.949	0.430
1600	1.000	0.969	0.949	0.436
2500	1.000	0.969	0.950	0.440

We deduce from this results that numerically the two conditions (C1) and (C2) are satisfied and proceed with the case $m = 900$. The eigenfunctions corresponding to the largest eigenvalues are shown in Figure 10 (middle),

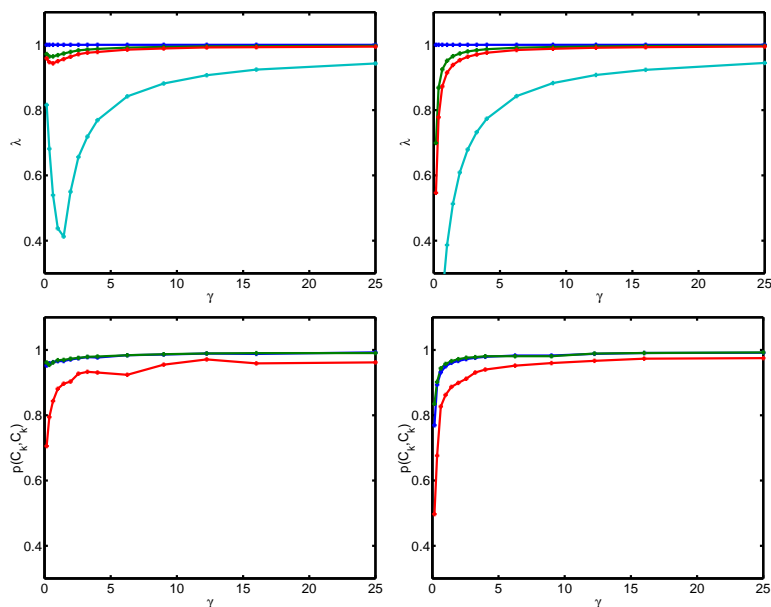


Figure 9: First four dominant eigenvalues (maximal real part) of P_τ for the Langevin equation (top left) and Smoluchowski equation (top right) for different values of γ . Metastabilities of the three subsets for the Langevin equation (bottom left) and Smoluchowski equation (bottom right) for different values of γ .

their projections onto the position space are depicted in Figure 7 (middle). We remark that, since the first three eigenvalues are real, the corresponding eigenfunctions can be chosen real-valued. Moreover, because the largest scalar product between either two different normalized eigenfunctions is of order 10^{-3} , the first three eigenfunctions are almost orthogonal, although the propagator is not reversible. The same holds for the projected eigenfunctions. Hence, the dominant part of P_τ is nearly self-adjoint and we would identify metastable subsets by application of the identification algorithm. However, we proceed in a slightly different way, since our aim is to identify conformations induced by the Langevin equation.

Recall from Section 1.1 that conformations are thought to be objects in the position space. For reduced models acting only on the position space Ω like, e.g., the Hamiltonian system with randomized momenta, the notions of conformations and metastable subsets coincide (see end of Sec. 1.1). This is different for models acting on the phase space $\Gamma = \Omega \times \mathbf{R}$ (positions and momenta) like, e.g., the Langevin equation. In this case we characterized conformations as special metastable subsets of the form $C = \hat{C} \times \mathbf{R} \subset \Gamma$ with $\hat{C} \subset \Omega$. Hence, for every position $q \in \hat{C}$, the conformation C contains all states of the form (q, p) with $p \in \mathbf{R}$. Our strategy to identify conformations induced by the Langevin equation is therefore as follows: In the position space Ω , run the identification algorithm based on the first three *projected*

eigenfunctions and extend the identified metastable subsets $\hat{C}_j \subset \Omega$ to the “cylindrical” subsets $C_j = \hat{C}_j \times \mathbf{R} \subset \mathbf{X}$. Finally check for metastability of the C_j w.r.t. the Langevin Markov process in Γ .

Applying this strategy¹³, we end up with a clustering $\{C_1, C_2, C_3\}$ with $C_1 = [q \leq -2.0] \times \mathbf{R}$, $C_2 = [-2.0 < q \leq 1.7] \times \mathbf{R}$ and $C_3 = [1.7 < q] \times \mathbf{R}$. The statistical weights $\mu(C_k)$ within the canonical ensemble μ_{can} and the metastabilities $p(\tau, C_k, C_k)$ are given by the following table:

metastable subset	C_1	C_2	C_3
statistical weight	0.350	0.594	0.055
metastability	0.963	0.966	0.884

(64)

Calculating the coupling matrix yields

$$C = \begin{pmatrix} 0.963 & 0.037 & 0 \\ 0.023 & 0.966 & 0.011 \\ 0 & 0.116 & 0.884 \end{pmatrix}. \quad (65)$$

So far we have analyzed the Langevin equation for the specific choice of friction constant $\gamma = 1$ and stochastic excitation $\sigma = 1$, resulting in the inverse temperature $\beta = 2$. Recalling that $\beta = 2\gamma/\sigma^2$, we find that there is a one-parameter family of Langevin equations corresponding to the same inverse temperature: choose $\sigma = \sqrt{2\gamma/\beta}$ for arbitrary $\gamma > 0$. Figure 8 illustrates the dependence of the spectrum on γ . It shows the spectrum of the stochastic transition matrix S obtained from discretizing the propagator P_τ corresponding to $\beta = 2$ and different values of γ . We observe that for small values of γ the spectrum of S is spread all over the unit disc, while it concentrates more and more on the interval $[0, 1]$ for larger values of γ . This reflects the fact that the Langevin equation is similar to the deterministic Hamiltonian system for $\gamma \approx 0$ [23], while it is similar to the Smoluchowski equation for $\gamma \gg 1$. Hence, the Langevin equation exhibits different behavior depending on the friction constant γ . This might also explain the dependence of the dominant eigenvalues on γ , as shown in Figure 9 (top). The eigenvalues accumulate in $\lambda = 1$ for $\gamma \approx 0$ as well as for $\gamma \gg 1$. The former effect reflects the fact that the deterministic Hamiltonian system admits infinitely many invariant probability distributions (see Sec. 6.1), while the latter effect is due to the behavior of the Smoluchowski equation for large γ .

¹³We could think of different strategies like, e.g., projecting the Langevin Markov process onto the position space Ω and considering the corresponding propagator. Alternatively, we could modify the Langevin model similar to the Hamiltonian system with randomized momenta in order to obtain a Langevin system with randomized momenta in the position space only. Our strategy is motivated by the fact that it fits best our context, as we are going to see below and in Section 6.5.

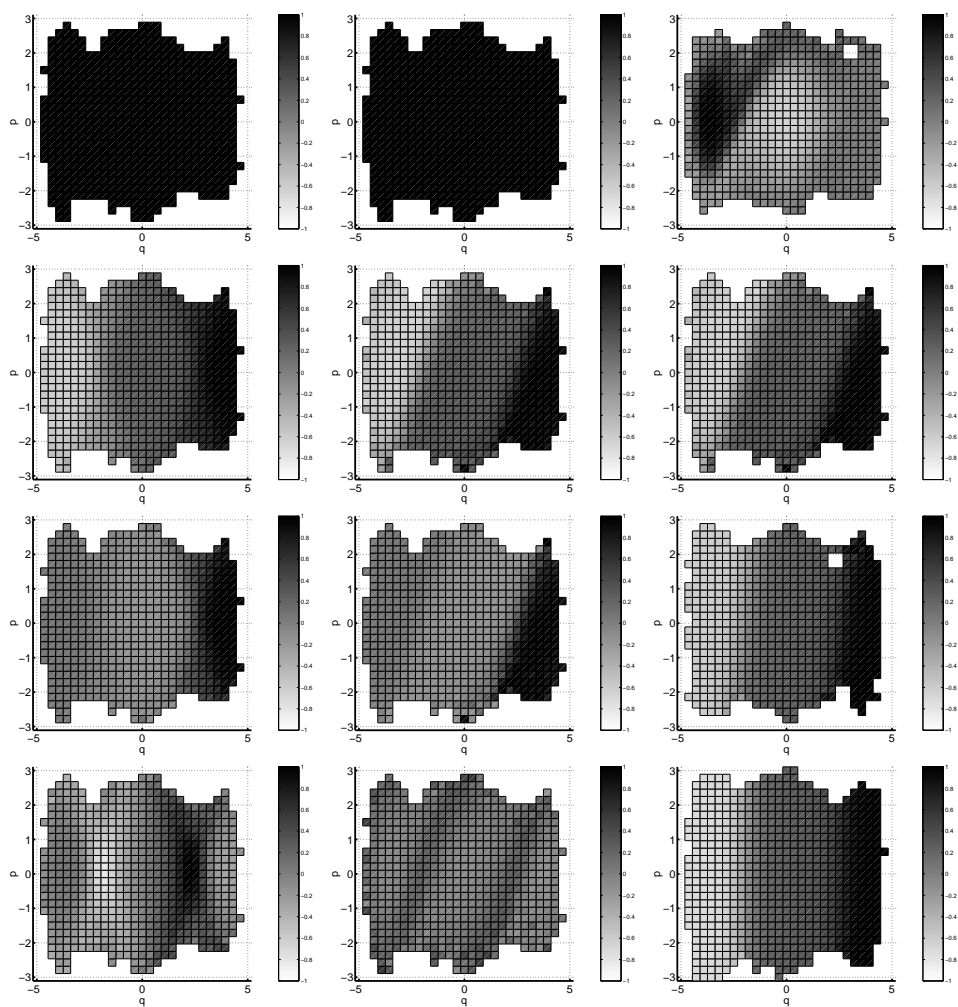


Figure 10: Left and middle: Dominant eigenfunctions for the Langevin equation for $\gamma = 1.0$. Eigenfunctions of the time-symmetrized discretization of P_τ corresponding to the eigenvalues 1.000, 0.984, 0.974, 0.701 (left) and eigenfunctions corresponding to the discretization of P_τ corresponding to 1.000, 0.969, 0.949, 0.430 (middle). Right: Second eigenfunction of P_τ for $\gamma = 0.16, 1.0, 4.0, 16.0$ (top to bottom).

Seemingly in contrast to the behavior of the eigenvalues is the decay of metastability of the corresponding clusterings, as shown in Figure 9 (bottom). This effect is due to our definition of conformations as cylindrical metastable subsets (see Sec. 1.1). Due to energy conservation in the deterministic Hamiltonian case, we would expect that subsets of the state space $\mathbf{X} = \Gamma$ corresponding to a certain energy range get more and more metastable for small γ . Since these subsets follow contour lines of H or, equivalently, of f_{can} (see Figure 5), every “cylindrical” subset will necessarily only partially intersect with these energy range subsets. This causes loss of metastability and explains the decay of metastability for small values of γ . Nevertheless, we would expect to find metastable subsets—subject to no restriction—for small γ , since the energy fluctuation of the Langevin equation decay for γ tending to zero. The mentioned structure due to nearly energy conservation is already visible in Figure 10 for the top left eigenfunction (see also [49]).

In the last part of this section we want to study the effects of time-symmetrization. Since the Langevin equation is not reversible, it is an ideal test model. In order to facilitate a comparison of the original and the time-symmetrized approach, we use the same realization of the Langevin Markov process X_n for the subsequent numerical discretization as above. The next proposition is a direct application of Theorem 4.31.

Proposition 6.4 *Fix some observation time span $\tau > 0$ and assume that the stochastic transition function r_τ corresponding to the time-symmetrized Langevin Markov process $Z_n = \{Z_n\}_{n \in \mathbf{Z}_+}$ is geometrically ergodic. Then the propagator $P_\tau : L^2(\mu_{\text{can}}) \rightarrow L^2(\mu_{\text{can}})$ corresponding to the time-symmetrized Markov process satisfies the conditions (C1) and (C2) in $L^2(\mu_{\text{can}})$.*

The assumption in Proposition 6.4 is in particular satisfied, if the stochastic transition function corresponding to the original Langevin Markov process can be proved to be uniformly ergodic. We now discretize the propagator P_τ corresponding to the time-symmetrized Langevin Markov process according to Example 5.2. Proceeding as in the first part, we end up with a *self-adjoint* 900×900 stochastic transition matrix S . Solving the eigenvalue problem for S yields:

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	...
1.000	0.984	0.974	0.701	0.649	0.539	...

In view of our algorithmic strategy, we are looking for a decomposition into three metastable subsets. The eigenfunctions corresponding to the largest eigenvalues are shown in Figure 10 (left). Application of our identification strategy yields a clustering $\{C_1, C_2, C_3\}$ with $C_1 = [q \leq -1.9] \times \mathbf{R}$, $C_2 =$

$[-1.9 < q \leq 1.8] \times \mathbf{R}$ and $C_3 = [1.8 < q] \times \mathbf{R}$. The statistical weights $\mu(C_k)$ within the canonical ensemble μ_{can} and the metastabilities $p(\tau, C_k, C_k)$ are given by the following table:

metastable subset	C_1	C_2	C_3	
statistical weight	0.358	0.585	0.057	(66)
metastability	0.975	0.977	0.921	

The coupling matrix based on the realization of the *original* Langevin Markov process X_n w.r.t. the clusters based on the time-symmetrized Markov process is given by

$$C = \begin{pmatrix} 0.975 & 0.025 & 0 \\ 0.015 & 0.977 & 0.008 \\ 0 & 0.079 & 0.921 \end{pmatrix}. \quad (67)$$

As predicted by Theorem 3.2 the metastability w.r.t. the time-symmetrized Markov process Z_n (see table (66)) and the metastability w.r.t. the original Markov process X_n (see table (64)) are identical. However, comparing the two coupling matrices (65) and (67), we see that a clustering based on the time-symmetrized process might differ from results based on the original process. This a consequence of the different resulting propagators and corresponding (projected) eigenfunctions.

6.4 Smoluchowski Equation

Denote by P_τ the propagator corresponding to the Smoluchowski Markov process. We first state under which conditions on the potential function the two requirements (C1) and (C2) on P_τ hold. Then, we numerically analyze the induced essential statistical behavior w.r.t. the positional canonical ensemble $\mu_{\mathcal{Q}}$.

Proposition 6.5 *The following holds for the Smoluchowski equation:*

- (i) *For periodic systems with position space $\Omega \subset \mathbf{R}^d$, some fixed observation time space $\tau > 0$ and smooth potential function $V : \Omega \rightarrow \mathbf{R}$, the propagator $P_\tau : L^1(\mu_{\mathcal{Q}}) \rightarrow L^1(\mu_{\mathcal{Q}})$ satisfies $r_{\text{ess}}(P_\tau) = 0$ and is asymptotically stable, hence conditions (C1) and (C2) are fulfilled in $L^1(\mu_{\mathcal{Q}})$.*
- (ii) *For bounded systems with position space $\Omega = \mathbf{R}^d$, some fixed observation time space $\tau > 0$ and smooth potential function $V : \Omega \rightarrow \mathbf{R}$ satisfying for some integer $\alpha > 0$ the growth conditions $V(q) \sim \|q\|^{2\alpha}$, $\nabla V(q) \sim \|q\|^{2\alpha-1}$ and $\partial^2 V(q) \sim \|q\|^{2\alpha-2}$ as $\|q\| \rightarrow \infty$, the propagator $P_\tau : L^2(\mu_{\mathcal{Q}}) \rightarrow L^2(\mu_{\mathcal{Q}})$ satisfies the conditions (C1) and (C2) in $L^2(\mu_{\mathcal{Q}})$.*

Proof: For periodic systems, application of Theorem 3 of Ichihara and Kunita [36] yields that the stochastic transition function $p(t, q, dy)$ is absolutely continuous w.r.t. μ_Q admitting a C^∞ density. Hence, the density is bounded, since Ω is compact, and therefore $r_{\text{ess}}(P_\tau) = 0$ by Theorem 4.9. Furthermore, we get asymptotic stability of P_τ according to [46], hence conditions (C1) and (C2) hold in $L^1(\mu_Q)$. For bounded systems, the statement follows by Theorem 4.31, if we can prove that the stochastic transition function is V -uniformly ergodic. But this is an immediate result of Theorem 5.3 by Mattingly et al. [51]. \square

Due to Mattingly et al. [51], the condition on the growth rate of the potential function in the case of bounded systems can be weakened (see cond. 5.1 in [51]). For an analysis of the Smoluchowski equation from a symmetric Markov semigroup point of view see Davies [8].

By Proposition 6.5 the application of the algorithmic approach to our test systems is theoretically justified. In order to discretize the propagator, we proceed according to Example 5.1 using the Trapezoid rule with $N = 300000$, $M = 1$ and the forward Euler or Euler–Maruyama scheme [41] with internal step size $\Delta t = 0.02$. A realization looks comparable to Figure 6. Discretizing the state space $\Omega = [-5, 5]$ with periodic boundary conditions into 30 equal-sized intervals, we obtain a 30×30 stochastic transition matrix S . Solving the eigenvalue problem for S yields:

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	...
1.000	0.950	0.915	0.387	0.227	0.125	...

Evaluating the indicator for $r_{\text{ess}}(P_\tau)$ we get $[\Delta(P_\tau)] = 0.361$ —for a further analysis of the indicator within a hierarchical context see below. As for the preceding model systems, we look for a decomposition into three metastable subsets. The eigenfunctions corresponding to the largest eigenvalues are shown in Figure 7 (right). Applying the identification algorithm, we end up with a clustering $\{C_1, C_2, C_3\}$ with $C_1 = \{q \leq -2.1\}$, $C_2 = \{-2.1 < q \leq 1.7\}$ and $C_3 = \{1.7 < q\}$. The statistical weights $\mu(C_k)$ within the positional canonical ensemble μ_Q and the metastabilities $p(\tau, C_k, C_k)$ are given by the following table:

metastable subset	C_1	C_2	C_3
statistical weight	0.353	0.589	0.058
metastability	0.948	0.956	0.867

The essential statistical behavior is given by the coupling matrix

$$C = \begin{pmatrix} 0.950 & 0.050 & 0 \\ 0.030 & 0.957 & 0.013 \\ 0 & 0.138 & 0.862 \end{pmatrix}.$$

From a theoretical point of view we know that $r_{\text{ess}}(P_\tau) = \Delta(P_\tau) = 0$ holds for periodic systems due to Prop. 6.5. Does the indicator $[\Delta(P_\tau)]$ reproduce this result? The following table shows the indicator based on different decompositions of the state space into m equal-sized intervals:

m	30	50	100	200
$[\Delta(P_\tau)]$	0.361	0.224	0.116	0.070

(68)

The values are in good agreement with the theoretical value $r_{\text{ess}}(P_\tau) = 0$. The previous results are based on a discretization of the propagator according to Example 5.1. This was possible, since the state space is very low-dimensional. In higher dimensions, we will have to use a discretization procedure according to Example 5.2 that is based on a realization of the Markov process $X_n = \{X_{n\tau}\}$ for some fixed time $\tau > 0$. Exploiting a realization of X_n we see that the dominant eigenvalues, the resulting clusterings, their statistical weights and their metastabilities are almost indistinguishable from the results previously obtained, while the indicator $[\Delta(P_\tau)]$ behaves quite differently:

m	30	50	100	200
$[\Delta(P_\tau)]$	0.409	0.389	0.500	0.500

The values for $[\Delta(P_\tau)]$ decrease first and then stay constant for a decomposition into more than 100 subsets. This seems to contradict the theoretical results, but can be understood in the context mentioned at the end of Section 5.1: Since the Smoluchowski process is reversible, we used the original as well as the reversed sampling to discretize the propagator. Analyzing the stochastic transition matrix S for, e.g., $m = 100$ we see that $[\Delta(P)] = S_{2,13} = S_{2,16} = 0.5$. Since the statistical weight of the 2nd subset is $3e - 6$, it implies that only one sampling point hit the second subset. Taking into account also the reversed sampling, we see that the second subset was exactly hit two times, which after normalization results in the value $S_{2,13} = S_{2,16} = 0.5$. In this case, $[\Delta(P_\tau)]$ indicates an insufficient sampling of the 2nd subset rather than an upper bound of the essential spectral radius, as was already discussed at the end of Section 5.1. A refinement of the sampling would decrease the value of $[\Delta(P_\tau)]$, as we already know from table (68).

As in the Langevin case, for a fixed inverse temperature β there is a one-parameter family of Smoluchowski equations parameterized by γ (or σ). The corresponding family of infinitesimal generators $\{\mathcal{L}_\gamma\}_{\gamma>0}$ has a very

simple form, since for fixed β and arbitrary $\gamma > 0$

$$\mathcal{L}_\gamma = \frac{1}{\gamma} \left(\underbrace{\frac{1}{\beta} \Delta_q - \nabla_q V(q) \cdot \nabla_q}_{\mathcal{L}_\beta} \right). \quad (69)$$

Hence, each generator \mathcal{L}_γ is simply a multiple of \mathcal{L}_β and we get the following relationship for the eigenvalues and -functions:

$$\mathcal{L}_\beta v = \lambda v \Leftrightarrow \mathcal{L}_\gamma v = \frac{\lambda}{\gamma} v.$$

We deduce that the eigenfunctions of \mathcal{L}_γ are *independent* of γ . For the corresponding propagator $P_t v = \exp(t\mathcal{L}_\gamma)v$ we deduce

$$\mathcal{L}_\gamma v = \frac{\lambda}{\gamma} v \Rightarrow P_t v = \exp\left(\frac{\lambda}{\gamma} t\right) v. \quad (70)$$

Figure 9 (top right) shows the eigenvalues of $P_\tau = \exp(\tau\mathcal{L}_\gamma)$ for different values of γ . From (70) we would expect a logarithmic decay with $1/\gamma$, which can be observed in a semi-logarithmic plot (not shown). Since the eigenvalues decay with decreasing γ , we expect that the metastabilities of the resulting decompositions also decay with decreasing γ due to the upper bound in Thm. 3.1. This phenomenon can be observed in Figure 9 (bottom right).

6.5 Comparison of Model Systems

In broad terms the Hamiltonian system with randomized momenta, the Langevin equation and the Smoluchowski equation behave quite similar (for the chosen model parameters). The numerical investigations show comparable results for the clusterings into metastable subsets, their statistical weights and the corresponding coupling matrices reflecting the essential statistical behavior. Qualitative different behavior is displayed by the fourth eigenfunctions (bottom line of Figure 7). But since for each model dynamics the fourth eigenvalue is well separated from the three dominant ones by a spectral gap, the influence on the dynamical behavior is expected to be less important.

The Hamiltonian system with randomized momenta is uniquely determined by specifying a potential function and some inverse temperature, hence specifying the canonical ensemble. In contrast to that, for fixed β , the Langevin and the Smoluchowski equation still depend on the friction constant γ , which can be related to the viscosity of the surrounding. As a consequence, the results the dynamical behavior depends on γ . Comparing the Langevin and the Smoluchowski model, we observe increasing

agreement of eigenvalues and metastabilities for increasing values of γ (see Figure 9). This is what we should expect, since the Smoluchowski equation was derived as a high-friction approximation of the Langevin equation (Section 2.4). This can also be understood from a dominant eigenfunction point of view: Figure 10 (right) shows the second dominant eigenfunctions for different values of γ ; the not depicted third eigenfunctions behaves similarly. For small friction the eigenfunctions still display features of the deterministic Hamiltonian system, while for larger values of γ the eigenfunctions convert more and more to a special product form, namely the product of a function acting on positions times a *constant* function in the momenta. Since the “loss of information” by projecting such a function to the position space is negligible, this might explain the good approximation quality of the Langevin by the Smoluchowski equation in the case of high friction. This is also reflected by the observation that the spectrum of the Langevin propagator becomes more and more “real-valued” (see Figure 8), as this is the case for the Smoluchowski equation. However, for small γ , the Langevin and the Smoluchowski equation exhibit different behavior. While in the Langevin case some eigenvalues accumulate in $\lambda = 1$ for $\gamma \ll 1$, this is not the case for the Smoluchowski equation, where all eigenvalues but $\lambda = 1$ tend to zero due to relation (70).

So far, we have presented an analysis of metastability for some fixed observation time span $\tau > 0$. How do the results depend on τ ? For the Hamiltonian system with randomized momenta, first investigations are documented in [68]. Exemplified for a small molecule, the dependence of the eigenvalues on τ is analyzed¹⁴. It looks almost exponential. For the Langevin and Smoluchowski equation the exponential dependence of the eigenvalues on τ is theoretically deducible on the basis of the semigroup property $P_\tau = \exp(\tau\mathcal{L})$. We further conclude that in this case the corresponding eigenfunctions and hence the thereon based metastable subsets are *independent* of any observation time span.

¹⁴Actually, Schütte analyzed in [68] the dependence of the eigenvalues on the inverse temperature β , which in combination with Section 3.7.2 of [68], where an (inverse) temperature scaling is related to a rescaling of the observation time span τ , gives the stated result.

7 Application to Large Systems

In this section we want to demonstrate that the algorithmic strategy presented in Section 3.2 can be applied to identify biomolecular conformations even for large systems as, for instance, small biomolecules with hundreds of atoms. For large systems, we have to face two particular problems:

1. How to approximate the stationary distribution in a high-dimensional space?
2. How to decompose the high-dimensional state space in order to discretize the propagator?

We will address these problems in the following.

7.1 Monte Carlo methods

The typical approach to sample the canonical distribution in a high-dimensional space is via Monte Carlo techniques. There is an extremely rich and varied literature on Monte Carlo methods (see, e.g., [44, 75]) and every converging method would allow to realize the problem of sampling the invariant distribution. In addition, we may also apply molecular dynamics based techniques, e.g., constant temperature sampling of the canonical distribution [54, 2]. It is widely known, that Monte Carlo simulations may suffer from possible “trapping problems” [48]. This kind of problem occurs when the Monte Carlo Markov process gets trapped near a local potential energy minimum due to high energy barriers so that a proper sampling of the entire state space within reasonable computing times is impossible. As illustrated in [68], this phenomenon is related to the existence of metastable subsets for the Monte Carlo Markov process.

There exists various strategies addressing the trapping problem. Especially the so-called extended ensemble methods, which are based on reweighting techniques, are gaining significant popularity. Recently, Fischer presented a promising alternative approach, the uncoupling–coupling Monte Carlo (UCMC) technique [20]. It links Monte Carlo sampling methods with the algorithmic strategy to the identification of metastable subsets, as described in this thesis. Loosely speaking, it exploits a clustering of the state space, when metastability starts to become “visible” but is far from causing trapping problems. As outline in [22] this may allow to sample the canonical distribution with reasonable computational effort.

7.2 Adaptive Discretization Techniques

There are different ways of facing the second problem, the decomposition of the state space. We assume that the canonical distribution has properly

been sampled by some Monte Carlo method. Then the following possibilities arise.

Essential Degrees of Freedom. Typical biomolecular systems contain hundreds or thousands of atoms. If we would generate a decomposition of the state space by simply decomposing every degree of freedom, the number of subsets within the decomposition and thus the dimension of the stochastic transition matrix would grow exponentially with the size of the molecular system. Chemical insight into biomolecular systems allows to circumvent this “curse of dimensionality”. Conformations of biomolecules are mostly described in terms of a few *essential degrees of freedom*. In the subspace of essential degrees of freedom most of the positional fluctuations occur, while in the remaining degrees of freedom the motion can be considered as “physically constrained”. *Based on the sampling of the canonical distribution*, we may determine essential degrees of freedom either in the position space according to Amadei et al. [1] or in the space of internal degrees of freedom, e.g., dihedral angles [35], by statistical analysis of circular data. Either case is based on a principal component analysis of the sampling via analyzing a covariance matrix. As shown in [35], this procedure may result in a tremendous reduction of the number of degrees of freedom and, consequently, in a moderate number of subsets within the decomposition when discretizing the essential variables only. The principal component analysis is a linear approach to essential degrees of freedom. A characterization and identification of more general nonlinear essential degrees of freedom is subject to investigations within a current research project [70] and part of a current diploma thesis [78].

Self-Organizing Maps. An alternative approach is to discretize the propagator by means of self-organizing maps, a special kind of neural networks. Self-organizing maps allow to cluster the Monte Carlo sampling data by assigning each sampling point to the nearest neurons, each of them representing a subset of the decomposition. We have demonstrated its successful application to sampling data of biomolecular systems in [28]. More advanced extensions, such as “box-neurons” and a hierarchical embedding, have recently been designed [26, 27].

Clustering Algorithms. A third approach of decomposing the state space is based on clustering the sampling data by means of clustering algorithms (see, e.g., [37] and cited references). These methods cluster according to structural similarity: The set of sampling points is partitioned into disjoint subsets with the property that two states belonging to the same subset are in some sense structural closer to each other than two states belonging to different subsets. A crucial question is the design of appropriate measures

of structural similarity. In the biomolecular application context, these measures can either be based on the Cartesian coordinates of the molecules or on the internal degrees of freedom. In contrast to the former the latter approach is invariant under rotations and translations of the entire molecule. For an application to biomolecular systems see [35].

Solving the Eigenvalue Problem. Finally, we want to remark that although the stochastic transition matrix resulting from the discretization may be quite large, it turns out to be sparse in our application context. Furthermore, since the algorithmic strategy is based solely on the dominant eigenvalues and corresponding eigenfunctions, we can apply subspace oriented iterative techniques (see, e.g., [64]) to solve the eigenvalue problem. It is important to notice that the convergence rate of those methods depends only on the spectral gap between the cluster of dominant eigenvalues and the remaining part of the spectrum and is *independent of the size of the stochastic transition matrix* and hence of the number of discretization subsets.

However, it should be clear that any refinement process of the discretization is limited by the quality of the underlying sampling data, since the approximation quality of the stochastic transition matrix is based on the interplay between sampling data and fineness of the discretization (see Sec. 5.3).

7.3 Analyzing a Small Biomolecule

This section illustrates the performance of the algorithmic approach to the triribonucleotide adenylyl(3'-5')cytidylyl(3'-5')cytidin (r(ACC)) model system in vacuum, see Figure 11. Its physical representation is based on the GROMOS96 extended atom force field [77], resulting in $N = 70$ atoms, hence $\Omega = \mathbf{R}^{210}$ and $\Gamma = \mathbf{R}^{420}$. The internal fluctuations are modeled w.r.t. the Hamiltonian system with randomized momenta. For details see [35].

The sampling of the canonical distribution was generated using an adaptive temperature hybrid Monte Carlo¹⁵ (ATHMC) method [21] at $T = 300\text{K}$ resulting in the sampling sequence $q_1, \dots, q_{32000} \in \Omega$. The dynamical fluctuations within the canonical ensemble were approximated by integrating $M = 4$ short trajectories of length $\tau = 80\text{fs}$ starting from each sampling point. To facilitate transitions, analogous to the ATHMC sampling, the momenta were chosen according to the momenta distribution $\mathcal{P}(p)$ corresponding to four different temperatures between $300\text{K} - 400\text{K}$ and reweighted afterwards. This resulted in a total of $4 \times 32.000 = 128.000$ transitions.

The configurational space was discretized using all four essential degrees of freedom, which were identified by means of a statistical analysis of the sampling data (see Sec. 7.2), resulting in $d = 36$ discretization subsets.

¹⁵ATHMC is part of the earlier mentioned UCMC method (see Section 7.1).

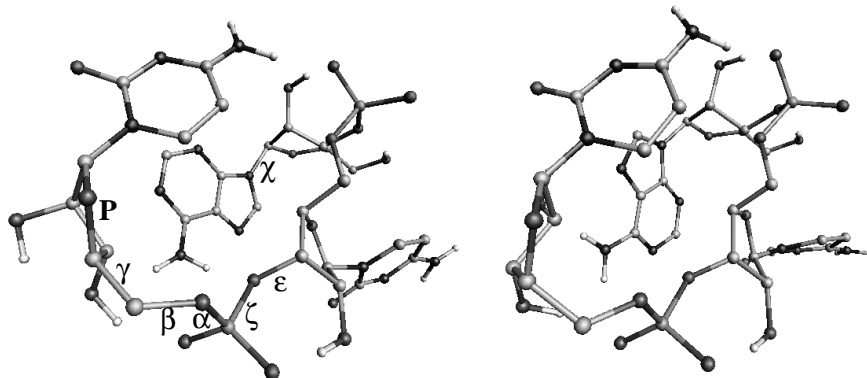


Figure 11: Two representatives of different conformations of r(ACC). Left: The χ angle around the first glycosidic bond is in *anti* position (-175 degrees) and the terminal ribose pucker P is in C(3')endo C(2')exo conformation. Right: The χ angle is in *syn* position (19 degrees) and the terminal ribose in C(2')endo C(3')exo conformation. Visualization by amira [42].

Then the 36×36 stochastic transition matrix S was computed based on the 128.000 transitions taking the different weighting factors into account. The computation of the eigenvalues of S close to 1 yielded a cluster of eight eigenvalues with a significant gap to the remaining part of the spectrum, as shown in the following table:

k	1	2	3	4	5	6	7	8	9	...
λ_k	1.00	0.99	0.98	0.97	0.96	0.95	0.93	0.90	0.81	...

Finally, we computed conformations based on the corresponding eight eigenvectors of S via the identification algorithm presented in Section 5.4. We identified eight conformations; their statistical weights and metastabilities are shown in the following table:

conformations	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
statistical weight	0.11	0.01	0.12	0.03	0.32	0.04	0.29	0.10
metastability	0.99	0.94	0.96	0.89	0.99	0.95	0.98	0.96

The transition probabilities between the different conformations are visualized schematically in Figure 12. The matrix allows to define a hierarchy between the clusters, which is inherent to the algorithm. On the top level, there are two clusters, one consisting of the conformations C_1, \dots, C_4 and the other consisting of the conformations C_5, \dots, C_8 . This structure corresponds to the two 4×4 blocks on the diagonal. On the next level, each

of these clusters splits up into two subclusters yielding four conformations $\{C_1, C_2\}$, $\{C_3, C_4\}$, $\{C_5, C_6\}$, $\{C_7, C_8\}$. On the bottom level, each cluster is further divided resulting in eight conformations.

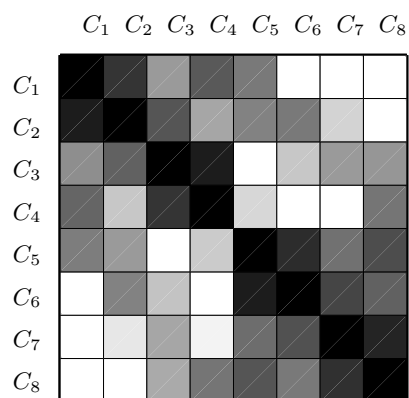


Figure 12: Schematical visualization of the transition probabilities $p(\tau, C_i, C_j)$ between the conformation C_i (row) and C_j (column). The colors are chosen according to the logarithm of the corresponding entries: from $p \approx 0$ (light) to $p \approx 1$ (dark).

Summary

Algorithmic approaches for the identification of essential statistical behavior have successfully been applied to study deterministic dynamical systems and molecular systems in a Hamiltonian context.

This thesis unifies and extends theory and algorithmic concepts from the formerly considered special classes of dynamical systems to the broader class of Markovian systems. We provide a detailed analysis of metastability and a new theoretical justification of the transfer operator based approach to metastability (Sec. 3). It is based on an instructive theorem (Theorem 3.1) specifying the relation between eigenvalues close to 1 and the existence of a decomposition into metastable subsets. This thesis contributes new links between spectral properties of transfer operators and well established Doeblin and ergodicity conditions for Markov processes and operators (Thms. 4.13, 4.24, 4.31).

We obtain a rather complete understanding in the $L^1(\mu)$ setting for general Markov processes (Secs. 4.2, 4.3), and for the $L^2(\mu)$ setting in the case of reversible Markov processes (Sec. 4.4). This allows us to successfully extend the concepts to new model systems, and we investigated for the first time the essential statistical behavior of the Langevin and the Smoluchowski equation in comparison with the Hamiltonian system with randomized momenta. We furthermore suggested an algorithmic indicator for the essential spectral radius (Sec. 5.1), which proved to be useful in application to our test system.

We outlined the strategies for studying larger molecular systems and successfully demonstrate its application to the study of the triribonucleotide $r(\text{ACC})$ (Sec. 7).

References

- [1] A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen. Essential dynamics of proteins. *Proteins*, 17:412–425, 1993.
- [2] H. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, 72:2384, 1980.
- [3] F. A. Bornemann and C. Schütte. Homogenization of Hamiltonian systems with a strong constraining potential. *Physica D*, 102:57–77, 1997.
- [4] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein. Metastability in stochastic dynamics of disordered mean-field models. *Probab. Theor. Rel. Fields*, 119:99–161, 2001.
- [5] F. E. Browder. On the spectral theory of elliptic differential operators I. *Math. Ann.*, 142:22–131, 1961.
- [6] J. B. Conway. *A Course in Functional Analysis*. Springer, New York, 1990.
- [7] F. Cordes and T. Galliat, 2001. Details to [27] (private communication).
- [8] E. B. Davies. *One-Parameter Semigroups*. Academic Press, London, 1980.
- [9] E. B. Davies. Metastable states of symmetric Markov semigroups I. *Proc. London Math. Soc.*, 45(3):133–150, 1982.
- [10] E. B. Davies. Metastable states of symmetric Markov semigroups II. *J. London Math. Soc.*, 26(2):541–556, 1982.
- [11] E. B. Davies. *Heat Kernels and Spectral Theory*. Cambridge Uni. Press, Cambridge, 1989.
- [12] M. Dellnitz and O. Junge. An adaptive subdivision technique for the approximation of attractors and invariant measures. *Comput. Visual. Sci.*, 1:63–68, 1998.
- [13] M. Dellnitz and O. Junge. On the approximation of complicated dynamical behavior. *SIAM J. Num. Anal.*, 36(2):491–515, 1999.
- [14] P. Deuffhard, M. Dellnitz, O. Junge, and C. Schütte. Computation of essential molecular dynamics by subdivision techniques. In [15], pages 98–115, 1999.
- [15] P. Deuffhard, J. Hermans, B. Leimkuhler, A. E. Mark, S. Reich, and R. D. Skeel, editors. *Computational Molecular Dynamics: Challenges, Methods, Ideas*, volume 4 of *Lecture Notes in Computational Science and Engineering*. Springer, 1999.
- [16] P. Deuffhard and A. Hohmann. *Numerische Mathematik I*. de Gruyter, Berlin, 1993.
- [17] P. Deuffhard, W. Huisinga, A. Fischer, and C. Schütte. Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains. *Lin. Alg. Appl.*, 315:39–59, 2000.
- [18] J. L. Doob. *Stochastic Processes*. John Wiley & Sons, Inc., New York, 1953.
- [19] N. Dunford and J. T. Schwartz. *Linear Operators, Part I: General Theory*, volume VII of *Pure and Applied Mathematics*. Interscience, New York, 1957.
- [20] A. Fischer. An uncoupling–coupling technique for Markov chains Monte Carlo methods. Konrad-Zuse-Zentrum, Berlin. Report 00-04, 2000.
- [21] A. Fischer, F. Cordes, and C. Schütte. Hybrid Monte Carlo with adaptive temperature in a mixed-canonical ensemble: Efficient conformational analysis of RNA. *J. Comput. Chem.*, 19:1689–1697, 1998.
- [22] A. Fischer, C. Schütte, P. Deuffhard, and F. Cordes. Hierarchical uncoupling–coupling of metastable conformations. In T. Schlick and H. H. Gan, editors, *Computational Methods for Macromolecules: Challenges and Applications—Proceedings of the 3rd International Workshop on Algorithms for Macromolecular Modelling*, Lecture Notes in Computational Science and Engineering. Springer, 2002.

- [23] M. Freidlin and A. Wentzell. *Random Perturbations of Dynamical Systems*. Springer, New York, 1984. Series in Comprehensive Studies in Mathematics.
- [24] G. Froyland. Private communication. Feb. 2001.
- [25] G. Froyland and M. Dellnitz. Statistically optimal almost-invariant sets: Efficient detection and adaptive resolution. Preliminary work, 2001.
- [26] T. Galliat and P. Deuffhard. Adaptive hierarchical cluster analysis by self-organizing box maps. Konrad-Zuse-Zentrum, Berlin. Report SC-00-13, 2000.
- [27] T. Galliat, P. Deuffhard, R. Roitzsch, and F. Cordes. Automatic identification of metastable conformations via self-organized neural networks. In T. Schlick and H. H. Gan, editors, *Computational Methods for Macromolecules: Challenges and Applications—Proceedings of the 3rd International Workshop on Algorithms for Macromolecular Modelling*, Lecture Notes in Computational Science and Engineering. Springer, 2002.
- [28] T. Galliat, W. Huisinga, and P. Deuffhard. Self-organizing maps combined with eigenmode analysis for automated cluster identification. In H. Bothe and R. Rojas, editors, *Neural Computation*, pages 227–232. ICSC Academic Press, 2000.
- [29] W. Gilks, S. Richardson, and D. Spiegelhalter, editors. *Markov chain Monte-Carlo in Practice*. Chapman and Hall, London, 1997.
- [30] R. Haberlandt, S. Fritzsche, G. Peinel, and K. Heinzinger. *Molekulardynamik - Grundlagen und Anwendungen*. Vieweg, 1995.
- [31] D. W. Heermann. *Computer Simulation Methods*. Springer, Berlin, 2nd edition, 1990.
- [32] H. Heuser. *Funktionalanalysis*. Teubner, Stuttgart, 1986.
- [33] W. Huisinga. Impression resulting from discussions with J. Mattingly, C. Schütte and A. Stuart during the Workshop on "Computational Stochastic Differential Equations", March 2001, Warwick.
- [34] W. Huisinga. The essential spectral radius and asymptotic properties of transfer operators in L^1 . *submitted to Dynam. Systems Appl.*, 2000. Available via <http://www.math.fu-berlin.de/~huisinga>.
- [35] W. Huisinga, C. Best, R. Roitzsch, C. Schütte, and F. Cordes. From simulation data to conformational ensembles: Structure and dynamic based methods. *J. Comp. Chem.*, 20(16):1760–1774, 1999.
- [36] K. Ichihara and H. Kunita. A classification of the second order degenerate elliptic operators and its probabilistic characterization. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 30:235–254, 1974.
- [37] A. K. Jain and R. C. Dubes. *Algorithms for Clustering Data*. Prentice Hall, New Jersey, Advanced Reference Series edition, 1988.
- [38] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer, New York, 1991.
- [39] T. Kato. Perturbation theory of nullity, deficiency and other quantities of linear operators. *J. d'Analyse Math.*, 6:261–322, 1958.
- [40] T. Kato. *Perturbation Theory for Linear Operators*. Springer, Berlin, 1995. Reprint of the 1980 edition.
- [41] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, New York, 1991.
- [42] Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB), Indeed-Visual Concepts GmbH and TGS Template Graphics Software Inc. *Amira—Advanced Visualization, Data Analysis and Geometry Reconstruction, User's Guide and Reference Manual*, 2000.

- [43] M. A. Krasnoselskii, P. P. Zabreiko, E. I. Pustynnik, and P. E. Sbuloevskii. *Integral Operators in Spaces of Summable Functions*. Noordhoff Int. Pub., Leyden, 1976.
- [44] D. P. Landau and K. Binder. *Monte Carlo Simulations in Statistical Physics*. Cambridge Uni. Press, Cambridge, 2000.
- [45] K. Lange. Decomposition of substochastic transition functions. *Proc. Amer. Math. Soc.*, 37:575–580, 1973.
- [46] A. Lasota and M. C. Mackey. *Chaos, Fractals and Noise*, volume 97 of *Applied Mathematical Sciences*. Springer, New York, 2nd edition, 1994.
- [47] A. Lebow and M. Schechter. Semigroups of operators and measures of noncompactness. *J. Funct. Anal.*, 7:1–26, 1971.
- [48] E. Leontidis, B. M. Forrest, A. H. Widmann, and U. W. Suter. Monte Carlo algorithms for the atomistic simulation of condensed polymer phases. *J. Chem. Soc. Faraday Trans.*, 91(16):2355–2368, 1995.
- [49] S. Lorenz. Asymptotic behavior of perturbed dynamical systems in the context of modelling biomolecules. Diploma thesis, Department of Mathematics and Computer Science, Free University Berlin, 2001.
- [50] L. Maligranda. Weakly compact operators and interpolation. *Acta Appl. Math.*, 72:79–89, 1992.
- [51] J. Mattingly, A. M. Stuart, and D. J. Higham. Ergodicity for SDEs and approximations: Locally Lipschitz vector fields and degenerated noise. Submitted, 2001.
- [52] S. Meyn and R. Tweedie. *Markov Chains and Stochastic Stability*. Springer, Berlin, 1993.
- [53] E. Nelson. *Dynamical Theories of Brownian Motion*. Mathematical Notes. Princeton Uni. Press, 1967.
- [54] S. Nosé. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81(1):511–519, 1984.
- [55] E. Nummelin. *General Irreducible Markov Chains and Non-Negative Operators*. Cambridge Uni. Press, Cambridge, London, New York, New Rochelle, 1984.
- [56] R. D. Nussbaum. The radius of the essential spectrum. *Duke Math. J.*, 37:473–478, 1970.
- [57] B. Øksendal. *Stochastic Differential Equations*. Universitext. Springer, Berlin, Heidelberg, fifth edition, 1998.
- [58] W. K. Olson and B. Victor. Modeling DNA deformations. *Current Opinion in Structural Biology*, 10:286–297, 2000.
- [59] M. Reed and B. Simon. *Methods of Modern Mathematical Physics IV, Analysis of Operators*. Academic Press, New York, 1978.
- [60] D. Revuz. *Markov Chains*. North-Holland, Amsterdam, Oxford, 1975.
- [61] H. Risken. *The Fokker-Planck Equation*. Springer, New York, 2nd edition, 1996.
- [62] G. O. Roberts and J. S. Rosenthal. Geometric ergodicity and hybrid Markov chains. *Elect. Comm. in Probab.*, 2:13–25, 1997.
- [63] G. O. Roberts and R. L. Tweedie. L^2 and L^1 convergence are equivalent for reversible Markov chains. *Accepted in J. Appl. Probab*, 2000. Preprint.
- [64] Y. Saad. *Numerical Methods for Large Eigenvalue Problems*. Manchester Uni. Press, Manchester, 1992.
- [65] J. M. Sanz-Serna and M. P. Calvo. *Numerical Hamiltonian Problems*, volume 7 of *Applied Mathematics and Mathematical Computation*. Chapman & Hall, London, 1994.

- [66] M. Schechter. On the essential spectrum of an arbitrary operator I. *J. Math. Anal. Appl.*, 13:205–215, 1966.
- [67] M. Schechter. Quantities related to strictly singular operators. *Indiana Math. J.*, 21:1061–1071, 1972.
- [68] C. Schütte. *Conformational Dynamics: Modelling, Theory, Algorithm, and Application to Biomolecules*. Habilitation Thesis, Fachbereich Mathematik und Informatik, Freie Universität Berlin, 1998.
- [69] C. Schütte, A. Fischer, W. Huisinga, and P. Deuffhard. A direct approach to conformational dynamics based on hybrid Monte Carlo. *J. Comput. Phys., Special Issue on Computational Biophysics*, 151:146–168, 1999.
- [70] C. Schütte and W. Huisinga. Von Mehrskaligkeit zur effektiven Dynamik: Ein Transfer-Operator-Zugang über essentielle Freiheitsgrade. Schwerpunktprogramm der Deutschen Forschungsgesellschaft: "Analysis, Modellierung und Simulation von Mehrskalensproblemen".
- [71] C. Schütte and W. Huisinga. On conformational dynamics induced by Langevin processes. In B. Fiedler, K. Gröger, and J. Sprekels, editors, *EQUADIFF 99 - International Conference on Differential Equations*, volume 2, pages 1247–1262, Singapore, 2000. World Scientific.
- [72] C. Schütte, W. Huisinga, and P. Deuffhard. Transfer operator approach to conformational dynamics in biomolecular systems. In B. Fiedler, editor, *Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems*, pages 191–223. Springer, 2001.
- [73] G. Singleton. Asymptotically exact estimates for metastable Markov semigroups. *Quart. J. Math. Oxford*, 35(2):321–329, 1984.
- [74] G. Singleton. *Some Properties of Metastable Semigroups*. PhD thesis, King's College London, 1986.
- [75] A. D. Sokal. Monte Carlo methods in statistical mechanics: Foundations and new algorithms. Lecture Notes, Cours de Troisième Cycle de la Physique en Suisse Romande, Lausanne, 1989.
- [76] A. M. Stuart. Geometric ergodicity for SDEs arising in statistical physics. Warwick University preprint, April 2001.
- [77] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*. vdf Hochschulverlag AG, ETH Zürich, 1996.
- [78] J. Walter. Preparatory studies. Diploma thesis, Department of Mathematics and Computer Science, Free University Berlin, 2001.
- [79] L. Weis. On perturbations of Fredholm operators in $L_p(\mu)$ -spaces. *Proc. Amer. Math. Soc.*, 67:287–292, 1977.
- [80] L. Weis. Integral operators and changes of density. *Indiana Univ. Math. J.*, 31(1):83–96, 1982.
- [81] L. Weis. Decompositions of positive operators and some of their applications. In K. D. Bierstedt and B. Fuchssteiner, editors, *Functional Analysis: Surveys and Recent Results III*, pages 95–115. North-Holland, 1983.
- [82] L. Weis. Approximation by weakly compact operators in L_1 . *Math. Nachr.*, 118:321–326, 1984.
- [83] L. Weis. On the computation of some quantities in the theory of Fredholm operators. *Rend. Circ. Mat. Palermo, II (Supplemento)*, 5:109–133, 1984.
- [84] D. Werner. *Funktionalanalysis*. Springer, Berlin, 2nd edition, 1997.

- [85] K. Yosida and S. Kakutani. Operator–theoretical treatment of Markoff’s process and mean ergodic theorem. *Ann. of Math.*, 42(1):188–228, 1941.
- [86] H. X. Zhou, S. T. Wlodec, and J. A. McCammon. Conformation gating as a mechanism for enzyme specificity. *Proc. Nat. Acad. Sci. USA*, 95(9280–9283), 1998.
- [87] R. Zwanzig. Problems in nonlinear transport theory. In J. Ehlers et al., editor, *Systems Far From Equilibrium*, Lecture Notes in Physics, pages 198–225. Springer, 1980.