Freie Universität Berlin
Institut für Mathematik
Arnimallee 6
14195 Berlin

Master's thesis:

A variational approach for conformation dynamics

Feliks Nüske

July 6, 2012

Dr. Frank Noé, FU Berlin

Ich versichere hiermit an Eides statt, dass ich die vorliegende Masterarbeit selbstständig und ohne unzulässige fremde Hilfe erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Table of contents

I. Introduction

Stochastic dynamical systems play an important role in many branches of mathematics and applications. Loosely speaking, such a system is a superposition of a deterministic dynamical system with random fluctuations, often called "noise". A typical example is the simulation of molecules (molecular dynamics, MD). In principle, one could model a molecule as a classical physical system by assigning to each atom its three spatial and three momentum coordinates and propagate these coordinates using some dynamical model. Since the molecule can typically exchange heat with its surroundings, energy is not conserved, which means that classical Hamiltonian dynamics is not a suitable dynamical model. Instead, the exchange of heat is usually modelled by some random influence. As an example, one can use Smoluchowski dynamics: If we denote the number of atoms by N and the $3N$-dimensional position vector by $\mathbf{x}$, the time evolution is given by the stochastic differential equation

$$\dot{\mathbf{x}}(t) = \frac{1}{m\gamma}\nabla V(\mathbf{x}(t)) + \sqrt{2D}dW_t. \tag{I.1}$$

Here, V is the potential energy function of the system, D is the diffusion constant, γ denotes friction and W_t is $3N$-dimensional Brownian motion. We will come back to this model in more detail at a later point, but note that this equation defines a stochastic process instead of a deterministic dynamical system. Instead of asking for a deterministic position in the $3N$-dimensional state space, we ask for the probability to find the system in a certain region of the state space.
Molecular systems very often display so called **conformations**, or metastable states. This means that while the system still oscillates and fluctuates, the overall geometry remains the same for long times, [Schütte, Huisinga, Deuflhard, Fischer, 1999]. Only occasionally, transitions from one conformation to another can be observed. An example of this phenomenon is shown in Figure I.1. A quantity of interest would be the average waiting time until such a transition occurs, because it would help to understand the overall behaviour of such a molecule.

It has also been shown by [Schütte, Huisinga, Deuflhard, Fischer, 1999] that the time evolution of probability densities described above can be computed by the action of a linear integral operator, called the propagator. Moreover, the spectral properties of this operator

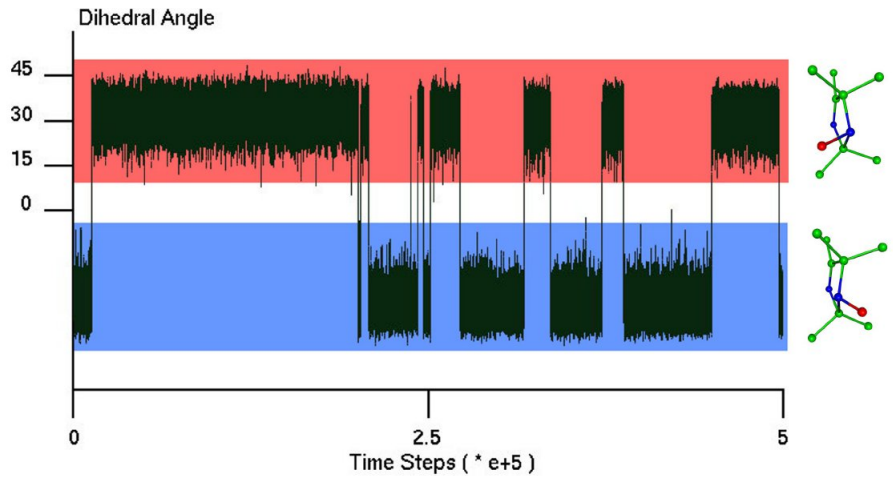


Figure I.1.: An internal coordinate of a small molecule transitioning between two different conformations, [Weber, 2010, ch. 1.1].

can be related to average waiting times, the quantities of interest. Therefore, estimating eigenvalues and eigenfunctions of the propagator is an interesting task. Markov state models, [Prinz et al, 2011], have been frequently used to this end. The main idea is to discretize the state space into a finite number of sets, and then approximate the true propagator by a finite-dimensional matrix operator. The eigenvalues of this matrix can be used as an approximation of the real eigenvalues. However, the quality of such an approximation largely depends on the choice of discretization. Finding a good discretization requires the use of clustering techniques to group the data. In the following, we present a different approach to approximating the dominant eigenvalues and eigenfunctions, not based on partition of unity membership functions, but on the use of smooth functions in combination with a variational principle. The main idea is to use the shape of the molecular energy function to choose the basis functions needed for the approximation. We hope that this can be done without the application of a clustering method to sampled data and thus help to avoid many numerical instabilities. We will develop the theory in chapter II, then apply the method to one dimensional examples of a diffusion process in chapter III, and finally tackle a higher dimensional problem in chapter IV. In the end, we hope that this might lead to a robust and computationally affordable method to compute eigenvalues and relevant time scales of stochastic processes stemming from real world examples.

II. The variational principle

II.1. The transfer operator

We consider a time-dependent stochastic process X_t, $t \geq 0$, on a usually high-dimensional and continuous state space Ω. Let the process satisfy the **Markov property**, that is, for all $x_0, x_1, \ldots, x_n$ and $y \in \Omega$, $t_0 > t_1 > \ldots > t_n \geq 0$ and $\tau > 0$, we have

$$\mathbb{P}\left(X_{t_0+\tau} = y | X_{t_0} = x_0, X_{t_1} = x_1, \ldots, X_{t_n} = x_n\right) = \mathbb{P}\left(X_{t_0+\tau} = y | X_{t_0} = x_0\right). \tag{II.1}$$

In other words, the future behaviour of the process only depends on the current state and not on the past. A more detailed and more precise formulation of the Markov property can be found in [Behrends, 2011, ch. 2]. Let us consider processes which are time-homogeneous, meaning that the conditional probability $\mathbb{P}(X_{t_0+\tau} = y | X_{t_0} = x)$ in the above equation does not depend on t_0, but only on the time difference τ and on the positions x, y. In this case, the process allows us to define a function p:

$$p(x, y; \tau) := \mathbb{P}(X_{t+\tau} = y | X_t = x) = \mathbb{P}(X_\tau = y | X_0 = x). \tag{II.2}$$

We shall call this function the **transition kernel**. It enables us to understand the evolution of probability densities in time. Suppose that, at time $t = 0$, the system is distributed according to a distribution ρ_0. The probability that, after time $\tau > 0$, the system transitioned from a point x to another point y, is given by the product of the probability of being at x at time $t = 0$ and the conditional probability to go from x to y in time τ. Consequently, the density ρ_τ belonging to time τ, evaluated at y, is obtained from integrating this product over all possible states x:

$$\rho_\tau(y) = \int_\Omega \mathrm{d}x\, p(x, y; \tau)\rho_0(x). \tag{II.3}$$

Furthermore, we assume the process to be sufficiently **ergodic**, which means in essence that the system cannot be decomposed into two dynamically independent components, and every state of the system will be visited infinitely often over an infinite run of the process. In this case, there is a unique probability distribution $\mu : \Omega \mapsto \mathbb{R}$, which is invariant in time, i.e.

$$\int_\Omega \mathrm{d}x\, p(x, y; \tau)\mu(x) = \mu(y), \tag{II.4}$$

regardless of τ, [Sarich, Noé, Schütte, 2010, ch. 2.2] The function μ is called the **invariant measure**, the **invariant density** or the **stationary distribution** of the process.
Also, we assume the transition kernel to satisfy the **detailed balance condition**, i.e.

$$\mu(x)p(x,y;\tau) = \mu(y)p(y,x;\tau) \quad \forall x,y \in \Omega. \tag{II.5}$$

Detailed balance is an important assumption that is justified in many applications from physics. In a system not satisfying Equation II.5, there would be a pathway where one direction is preferred compared to the other. This would allow the process to produce work, that is, to convert thermal energy into work. This is impossible in physical systems due to the second law of thermodynamics, [Prinz et al, 2011, sec. 2.A].

We see that Equation II.3 defines the action of a lag time dependent integral operator $\mathcal{P}(\tau)$ on suitable probability densities ρ. Let us call it the **propagator** of the system, which has the following important property:

Lemma II.1 (Chapman-Kolmogorov equation): *For any $\tau_1, \tau_2 > 0$, we have*

$$\mathcal{P}(\tau_1 + \tau_2) = \mathcal{P}(\tau_1)\mathcal{P}(\tau_2). \tag{II.6}$$

—

Proof. First, we observe that the Markov property and time-homogeneity imply for $x, z \in \Omega$,

$$p(x,z;\tau_1+\tau_2) = \mathbb{P}(X_{\tau_1+\tau_2} = z | X_0 = x) = \frac{1}{\mathbb{P}(X_0 = x)}\mathbb{P}(X_{\tau_1+\tau_2} = z, X_0 = x) \tag{II.7}$$

$$= \frac{1}{\mathbb{P}(X_0 = x)} \int_\Omega \mathrm{dy}\, \mathbb{P}(X_{\tau_1+\tau_2} = z | X_{\tau_2} = y)\mathbb{P}(X_{\tau_2} = y, X_0 = x) \tag{II.8}$$

$$= \int_\Omega \mathrm{dy}\, p(y,z;\tau_1)\frac{\mathbb{P}(X_{\tau_2} = y, X_0 = x)}{\mathbb{P}(X_0 = x)} = \int_\Omega \mathrm{dy}\, p(y,z;\tau_1)p(x,y;\tau_2). \tag{II.9}$$

Using this, we show that for a function ρ,

$$\mathcal{P}(\tau_1+\tau_2)\rho(z) = \int_\Omega \mathrm{dx}\, p(x,z;\tau_1+\tau_2)\rho(x) = \int_\Omega \mathrm{dx} \int \mathrm{dy}\, p(y,z;\tau_1)p(x,y;\tau_2)\rho(x) \tag{II.10}$$

$$= \int_\Omega \mathrm{dy}\, p(y,z;\tau_1)\mathcal{P}(\tau_2)\rho(y) = \mathcal{P}(\tau_1)\mathcal{P}(\tau_2)\rho(z). \tag{II.11}$$

■

Let us additionally assume that $\mathcal{P}(\tau)\rho \to \rho$ for all suitable functions ρ, and let us define $\mathcal{P}(0) := \mathrm{id}$. This means that Lemma II.1 is also valid for all $\tau_1, \tau_2 \geq 0$. Before we go on to study the propagator, let us introduce a slight modification to it. Denoting by $L^2_\mu(\Omega)$ the space of all functions $v : \Omega \mapsto \mathbb{R}$ which are square-integrable with respect to the weight function μ, i.e. $\int_\Omega v^2(x)\mu(x)dx$ is finite, with the weighted scalar-product $\langle v \mid w \rangle_\mu = \int_\Omega \mathrm{dx}\, v(x)w(x)\mu(x)$, we make the following definition:

Definition II.2 (Transfer operator): *The* **transfer operator** $\mathcal{T}(\tau)$ *acting on a function* $u \in L^2_\mu(\Omega)$ *is defined by:*

$$\mathcal{T}(\tau)u(y) := \frac{1}{\mu(y)} \int_\Omega \mathrm{dx}\, p(x, y; \tau)\mu(x)u(x). \tag{II.12}$$

—

The domain of definition is chosen to have this operator act on a Hilbert space. We note the most important properties of $\mathcal{T}(\tau)$:

Lemma II.3: *If the transition kernel* $p(x, y; \tau)$ *is given by a smooth and bounded probability density, the transfer operator* $\mathcal{T}(\tau)$ *is linear, bounded, compact and self-adjoint.*

—

Proof. Linearity follows directly from the definition. In order to prove boundedness, let us first use detailed balance to find:

$$\langle \mathcal{T}(\tau)v \mid \mathcal{T}(\tau)v \rangle_\mu = \int_\Omega \mathrm{dy}\, \left[\mathcal{T}(\tau)v(y)\right]^2 \mu(y) \tag{II.13}$$

$$= \int_\Omega \mathrm{dy}\, \frac{1}{\mu(y)^2} \left[\int_\Omega \mathrm{dx}\, p(x, y; \tau)\mu(x)v(x)\right]^2 \mu(y) \tag{II.14}$$

$$= \int_\Omega \mathrm{dy}\, \frac{1}{\mu(y)} \mu(y)^2 \left[\int_\Omega \mathrm{dx}\, p(y, x; \tau)v(x)\right]^2. \tag{II.15}$$

Before we proceed, we make the following general observation. If π is some probability density on Ω and $f : \Omega \mapsto \mathbb{R}$ a function, we have

$$0 \leq \omega^2 \left[f\right]_\pi = \mathbb{E}\left[f^2\right]_\pi - \left(\mathbb{E}\left[f\right]_\pi\right)^2, \tag{II.16}$$

where we use $\omega^2 \left[\,\cdot\,\right]_\pi$ to denote the variance of f with respect to the distribution π. Hence,

$$\left(\int_\Omega \mathrm{dx}\, f(x)\pi(x)\right)^2 = \left(\mathbb{E}\left[f\right]_\pi\right)^2 \leq \mathbb{E}\left[f^2\right]_\pi = \int_\Omega \mathrm{dx}\, f^2(x)\pi(x). \tag{II.17}$$

In Equation II.15, $p(y, x; \tau)$ is, by assumption, a smooth probability density in x. We can consequently apply the above estimate to the squared integral and obtain

$$\langle \mathcal{T}(\tau)v \mid \mathcal{T}(\tau)v \rangle_\mu \leq \int_\Omega \mathrm{d}y\, \mu(y) \int_\Omega \mathrm{d}x\, p(y, x; \tau) v^2(x) \tag{II.18}$$

$$= \int_\Omega \mathrm{d}x\, v^2(x) \int_\Omega \mathrm{d}y\, \mu(y) p(y, x; \tau) = \int_\Omega \mathrm{d}x\, v^2(x) \mu(x) = \|v\|_\mu^2. \tag{II.19}$$

We have thus found that $\mathcal{T}(\tau)$ is a well-defined and bounded operator on $L^2_\mu(\Omega)$ with operator norm less or equal to one. By inserting the constant function $\mathbb{1}$ which satisfies $\mathbb{1}(x) = 1 \quad \forall x \in \Omega$, Equation II.4 immediately yields that $\mathcal{T}(\tau)\mathbb{1} = \mathbb{1}$ and consequently, $\|\mathcal{T}(\tau)\| = 1$. As $p(x, y; \tau)$ is bounded for all x, y, we also have $p(x, y; \tau) \in L^2_{\mu\times\mu}(\Omega \times \Omega)$, therefore $\mathcal{T}(\tau)$ is a Hilbert-Schmidt operator and thus compact, [Werner, 2002, ch. 6, p. 272]. Finally, self-adjointness also follows from the detailed balance condition:

$$\langle \mathcal{T}(\tau)u \mid v \rangle_\mu = \int_\Omega \mathrm{d}y\, \frac{1}{\mu(y)} \left[\int_\Omega \mathrm{d}x\, p(x, y; \tau) \mu(x) u(x) \right] \mu(y) v(y) \tag{II.20}$$

$$= \int_\Omega \int_\Omega \mathrm{d}x\, \mathrm{d}y\, p(y, x; \tau) \mu(y) u(x) v(y) \tag{II.21}$$

$$= \int_\Omega \mathrm{d}x\, \frac{1}{\mu(x)} \left[\int_\Omega \mathrm{d}y\, p(y, x; \tau) \mu(y) v(y) \right] \mu(x) u(x) = \langle u \mid \mathcal{T}(\tau)v \rangle_\mu . \tag{II.22}$$

■

The transfer operator defines a compact and self-adjoint map on the Hilbert space $L^2_\mu(\Omega)$. From functional analysis, we know that it has rich spectral properties, see [Werner, 2002, ch. 6, p. 241] : The spectrum is real-valued and, because of $\|\mathcal{T}(\tau)\| = 1$, contained within the interval $[-1, 1]$. All non-zero spectral values are discrete and are indeed eigenvalues with corresponding eigenfunctions. The eigenvalues can be ordered into a series λ_i with $\lambda_i \to 0$ as $i \to 0$. The Hilbert space $L^2_\mu(\Omega)$ can be decomposed into the direct sum of the eigenfunctions' linear span and the operator's kernel. If one extends the linear span by an orthonormal basis of the kernel, we obtain an orthonormal basis ψ_i of the full Hilbert space such that the action of $\mathcal{T}(\tau)$ completely decomposes into the action on each of these functions. If we write $v \in L^2_\mu(\Omega)$ as $v = \sum_{i=1}^\infty \langle v \mid \psi_i \rangle_\mu \psi_i$, we then find that

$$\mathcal{T}(\tau)v = \sum_{i=1}^\infty \langle v \mid \psi_i \rangle_\mu \mathcal{T}(\tau)\psi_i = \sum_{i=1}^\infty \lambda_i \langle v \mid \psi_i \rangle_\mu \psi_i, \tag{II.23}$$

where possibly $\lambda_i = 0$ holds from some index i on. As we have already noticed in the course of proving Lemma II.3, $\lambda_1 = 1$ is an eigenvalue with corresponding eigenfunction $\psi_1 = \mathbb{1}$. In many applications, it turns out that this eigenvalue is simple and dominant, meaning that its multiplicity is one and that -1 is not an eigenvalue of $\mathcal{T}(\tau)$. Additionally, there usually is a number of further eigenvalues $1 > \lambda_2 > \ldots > \lambda_m$, which are strictly smaller but "close" to one, whereas the remaining spectrum is contained in a ball around zero which is essentially bounded away from these eigenvalues, see again [Sarich, Noé, Schütte, 2010, ch 2.2]. Let

us from now on assume this to be the case, and let us call these eigenvalues the **dominant eigenvalues** and their corresponding functions $\psi_2, \ldots, \psi_m$ the **dominant eigenfunctions**.

Consider now the map $\mathcal{M} : L^2_\mu(\Omega) \mapsto L^2_{\mu^{-1}}(\Omega)$ defined by $\mathcal{M}\psi(x) := \mu(x)\psi(x)$ for any function $\psi \in L^2_\mu(\Omega)$. Here, $L^2_{\mu^{-1}}(\Omega)$ is the L^2-space with weight function $\mu^{-1}(x) = \frac{1}{\mu(x)}$, where we assumed the stationary density to be non-zero, having in mind that in our later contexts, it will always be given by the Boltzmann distribution. $\mathcal{M}$ is linear and well-defined, indeed, if $\psi \in L^2_\mu(\Omega)$, then $\langle \mathcal{M}\psi \mid \mathcal{M}\psi \rangle_{\mu^{-1}} = \int_\Omega \mathrm{d}x\, \mu(x)\psi(x)\mu(x)\psi(x)\mu^{-1}(x) = \|\psi\|^2_\mu$. Moreover, $\mathcal{M}$ is a bijective map, which can rapidly be checked, and because of

$$\langle \mathcal{M}\psi \mid \phi \rangle_{\mu^{-1}} = \int_\Omega \mathrm{d}x\, \mu(x)\psi(x)\phi(x)\mu^{-1}(x) = \left\langle \psi \mid \mathcal{M}^{-1}\phi \right\rangle_\mu, \tag{II.24}$$

$\mathcal{M}$ is also unitary. Looking back at Equation II.3 and Definition II.2, we find that on $L^2_{\mu^{-1}}(\Omega)$, $\mathcal{P}(\tau) = \mathcal{M}\mathcal{T}(\tau)\mathcal{M}^{-1}$. The propagator differs from the transfer operator by no more than a unitary transformation, and it thus inherits the self-adjointness and compactness properties we have just derived for $\mathcal{T}(\tau)$. In particular, if $\psi_i \in L^2_\mu(\Omega)$ is a transfer operator eigenfunction, then $\phi_i := \mathcal{M}\psi = \mu\psi$ is a propagator eigenfunction corresponding to the same eigenvalue, and vice versa. The Hilbert space $L^2_{\mu^{-1}}(\Omega)$ decomposes in the same way as $L^2_\mu(\Omega)$ does, only replacing the orthonormal basis $\{\psi_i\}$ by $\{\phi_i\}$.

Example 1: If $\Omega = \{x_1, \ldots, x_N\}$ is finite, it is enough to specify the conditional transition probabilities $p_{ij}(\tau) = \mathbb{P}(X_\tau = x_j | X_0 = x_i)$. A probability density simply becomes a vector $\mathbf{p}_\tau = (p_1(\tau), \ldots, p_N(\tau))$, assigning a probability $p_i(\tau)$ to each state x_i, and satisfying $\sum_{i=1}^N p_i(\tau) = 1$. Probability vectors are propagated in time by multiplication with the matrix $\mathrm{P}(\tau) = (p_{ij}(\tau))$ from the left, since

$$p_i(\tau) = \mathbb{P}(X_\tau = x_i) = \sum_{j=1}^N \mathbb{P}(X_\tau = x_i | X_0 = x_j)\mathbb{P}(X_0 = x_j) = \sum_{j=1}^N p_{ji}(\tau)p_j(0). \tag{II.25}$$

Consequently, $\mathbf{p}_\tau = \mathbf{p}_0\,\mathrm{P}(\tau)$. The linear propagator simply becomes the matrix $\mathrm{P}(\tau)^T$. The eigenfunctions ϕ_i are the left-hand eigenvectors of the matrix $\mathrm{P}(\tau)$, in particular, the stationary distribution is a vector $\pi := \phi_1$, which satisfies $\pi\,\mathrm{P}(\tau) = \pi$. Since the transformation $\mathcal{M}$ acts on a function by point wise multiplication with the stationary density, $\mathcal{M}$ can be written as a matrix Π, which is diagonal and carries the vector π on its diagonal. Since, with this notation, reversibility can be written as $\Pi\,\mathrm{P}(\tau) = \mathrm{P}(\tau)^T\Pi$, we find that

$$\mathcal{T}(\tau) = \mathcal{M}^{-1}\mathcal{P}(\tau)\mathcal{M} = \Pi^{-1}\,\mathrm{P}(\tau)^T\Pi = \Pi^{-1}\Pi\,\mathrm{P}(\tau) = \mathrm{P}(\tau). \tag{II.26}$$

We see that propagator eigenvectors, which are left-hand eigenvectors of the P-matrix, become transfer operator eigenvectors upon transposing and multiplication with Π^{-1}, and thus become right-hand eigenvectors of the P-matrix. In the finite-dimensional case, the difference between the propagator and the transfer operator is just a matter of left-hand and right-hand eigenvectors of the same matrix.

—

The spectral decomposition, the Chapman-Kolmogorov equation Lemma II.1 and the assumption $\mathcal{P}(\tau)\phi \to \phi$ for all $\phi \in L^2_{\mu^{-1}}(\Omega)$ show us that the i-th eigenvalue $\lambda_i(\tau)$ is a continuous function of τ and satisfies $\lambda_i(\tau_1+\tau_2) = \lambda_i(\tau_1)\lambda_i(\tau_2)$ as well as $\lambda_i(0) = 1$. Therefore, it can be written as $\lambda_i(\tau) = e^{-\kappa_i \tau}$ for some $\kappa_i > 0$, that is, it has an exponential decay rate. Clearly, $\kappa_1 = 0$ and $\kappa_2, \ldots, \kappa_m$ are close to zero. Looking at the action of $\mathcal{P}(\tau)$ on some function ρ once more shows that

$$\mathcal{P}(\tau)\rho = \sum_{i=1}^{\infty} \langle \rho \mid \phi_i \rangle_{\mu^{-1}} \mathcal{P}(\tau)\phi_i = \sum_{i=1}^{\infty} e^{-\kappa_i \tau} \langle \rho \mid \phi_i \rangle_{\mu^{-1}} \phi_i. \tag{II.27}$$

If the time lag τ is very close to zero, almost all terms in the above sum will contribute. If τ becomes larger, however, there is a range of time lags τ where, for $i > m$, all exponential decay terms $e^{-\kappa_i \tau}$ have essentially vanished, whereas all terms with $i \leq m$ still contribute. If τ increases even further, namely if $\tau \gg \frac{1}{\kappa_2}$, there are no contributions except the very first term. For a probability distribution ρ, we have $\langle \rho \mid \phi_1 \rangle_{\mu^{-1}} = \langle \rho \mid \mu \rangle_{\mu^{-1}} = 1$, so $\mathcal{P}(\tau)\rho \approx \mu$. The system is then distributed according to its equilibrium distribution and has basically "forgotten" the initial deviation from equilibrium expressed by ρ. Each dominant eigenvalue defines a time scale $\frac{1}{\kappa_i}$ and a corresponding slow process which equilibrates only if one waits much longer than this time scale. The slow processes are the link between eigenvalues and metastable states. A slow process usually corresponds to an equilibration process between two metastable regions, and the implied time scale $\frac{1}{\kappa_i} = -\frac{\tau}{\log \lambda_i(\tau)}$ defines an average transition time. For this reason, the approximation of dominant eigenvalues is of such importance.

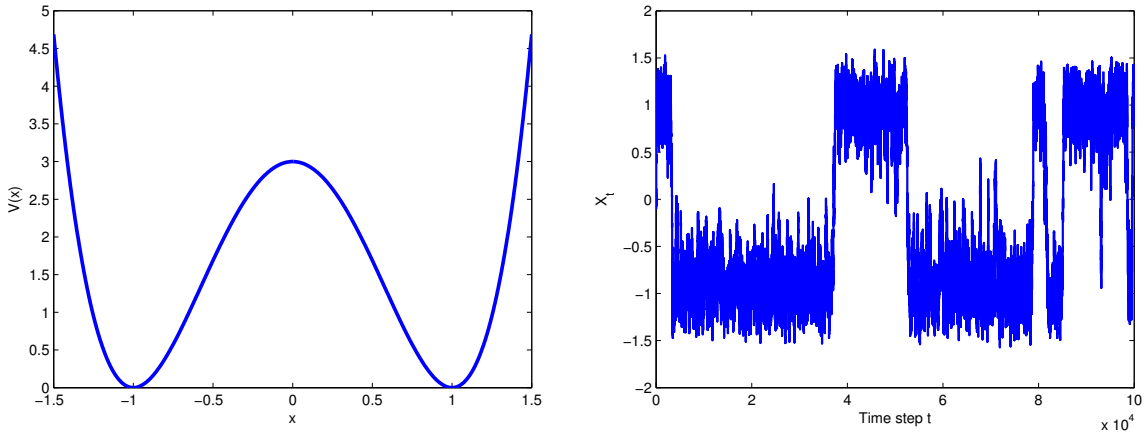
Example 2: Consider a four state system with stochastic transition matrix

$$\mathrm{P} = \begin{pmatrix} 0.7571 & 0.2429 & 0 & 0 \\ 0.1609 & 0.8306 & 0.0085 & 0 \\ 0 & 0.0084 & 0.8294 & 0.1622 \\ 0 & 0 & 0.2413 & 0.7587 \end{pmatrix}. \tag{II.28}$$

Clealy, most of the transitions will occur either between states x_1 and x_2 or between x_3 and x_4, whereas a transition from x_2 to x_3 will be a very rare event. Grouped together, x_1, x_2 and x_3, x_4 form two metastable states. The eigenvalues are $\lambda_1 = 1$, $\lambda_2 = 0.9900$, $\lambda_3 = 0.5964$ and $\lambda_4 = 0.5894$. Clearly, there is one dominant eigenvalue related to the slow transition process between the two metastable states.

—

Example 3: In chapter 2, we will discuss the diffusion process of a one-dimensional particle in an energy landscape V. The particle is subject to a force F, which equals the derivative of the energy function, but is additionally perturbed by random fluctuations. For the potential function V shown in Figure II.1a, the particle will spend most of the time around one of the two minima of V, and will just occasionally cross the barrier separating them. Figure II.1b shows a trajectory of the process, where the metastable behaviour is clearly visible. This is another typical example of metastable states, we will investigate it in more detail in the next



(a) (b)

Figure II.1.: The potential energy function V displaying two metastable states around $x = -1$ and $x = 1$, and a sample trajectory of the corresponding diffusion process.

chapter.

II.2. Markov state models

Markov state models (MSMs), [Prinz et al, 2011], are a widely used technique to approximate the eigenvalues of stochastic processes. The idea is to partition the state space into finitely many disjoint sets $A_1, \ldots, A_s$ with $\bigcup_{i=1}^{s} A_i = \Omega$, and only consider the transitions between any two of these sets. This means that instead of noticing every transition between any two points of the full continuous state space, we only take care of transitions between the sets A_i. Let us compute the conditional transition probability $p_{ij}(\tau)$ between two sets A_i and A_j, over time τ:

$$p_{ij}(\tau) = \mathbb{P}(X_\tau \in A_j | X_0 \in A_i) = \frac{\mathbb{P}(X_\tau \in A_j, X_0 \in A_i)}{\mathbb{P}(X_0 \in A_i)} \tag{II.29}$$

$$= \frac{\int_{A_i}\int_{A_j} \mathrm{dx}\,\mathrm{dy}\,\mu(x)p(x,y;\tau)}{\int_{A_i}\mathrm{dx}\,\mu(x)}. \tag{II.30}$$

The matrix $\mathrm{P}(\tau)$ with entries $p_{ij}(\tau)$ is a row-stochastic matrix which defines a Markov process on the finite state space $A_1, \ldots, A_s$, and it is called the MSM transfer matrix. Probability densities on this space, i.e. probability vectors $\mathbf{p} = (p_1, \ldots, p_s)$ with $\sum_{i=1}^{s} p_i = 1$, are transported by multiplication with $\mathrm{P}(\tau)$ from the left, i.e. $\mathbf{p}_\tau = \mathbf{p}_0\,\mathrm{P}(\tau)$. As an approximation to the true eigenvalues of the continuous propagator, one simply computes the eigenvalues of $\mathrm{P}(\tau)$. To this end, one has to find the matrix entries $p_{ij}(\tau)$. These are usually estimated from

simulation trajectories: Let $X_0, \ldots, X_N$ be a trajectory of length N, with a simulation time step $\Delta t = \tau$. One estimates $\mu_i := \int_{A_i} \mathrm{dx}\, \mu(x)$ as the number of simulation steps X_k with $X_k \in A_i$, divided by N. The transition probability $\int_{A_i} \int_{A_j} \mathrm{dx}\, \mathrm{dy}\, \mu(x) p(x, y; \tau)$ is estimated by the number of transitions from A_i to A_j, divided by the total number of transitions, in this case $N - 1$. In practice, it is often useful to choose τ as some general integer multiple of the simulation window Δt, i.e. $\tau = n\Delta t$, with n not necessarily equal to one. In this case, only transitions over n simulation time steps are taken into account, which usually leads to a higher approximation quality, as we shall see later.
It is very important to note that the Markov process defined by the MSM transfer matrix is an approximation of the continuous dynamics: The definition of the transfer matrix in Equation II.30 overlooks all transitions within one of the sets A_i. But these transitions are important, because the probability to cross over to some other set A_j very much depends on the current position within set A_i. Making use of the MSM therefore automatically results in a systematic error that effects the accuracy of the estimated eigenvalues. It is quite easy to show that this error decays exponentially proportional to the second eigenvalue. In their 2010 paper, Sarich, Noé and Schütte show that the pre-factor can be split up into two independent components. One of them depends on the lag time only, whereas the other one comes from the choice of discretization, [Sarich, Noé, Schütte, 2010, theorem 3.1]. The discretization is therefore of great importance, and a good choice can increase the approximation quality a lot.

Example 4: The four state system Equation II.28 shows the MSM approximation of the diffusion process shown in Figure II.1. We chose the four sets to be the intervals $[-2, -1]$, $[-1, 0]$, $[0, 1]$ and $[1, 2]$. The transition matrix $\mathrm{P} = \mathrm{P}(\tau)$ was estimated from a sample trajectory of 20 million steps. The lag time dependent improvement of the approximation quality can be visualized by looking at the second implied time scale $t_2 = -\frac{\tau}{\log(\lambda_2(\tau))}$. The faster the estimated time scale converges, the higher the quality of the approximation. Figure II.2 shows how the estimated time scale improves upon increasing the time lag. The four state discretization is a very simple one. We will achieve much faster convergence towards the correct time scale in the next chapter.

—

We would like to make use of two more aspects emphasized in this work. First, the construction of the MSM can be seen as the projection of the transfer operator $\mathcal{T}(\tau)$ onto a finite subspace of the Hilbert space $L^2_\mu(\Omega)$, namely the linear span of indicator functions corresponding to the sets A_i. Second, as pointed out in chapter 3.4 of that article, it is not necessary to restrict this projection to indicator functions, but the error estimate is still true if $\mathcal{T}(\tau)$ is projected onto a much more general subspace of $L^2_\mu(\Omega)$. This is, in a sense, the starting point of our work. Knowing that the discretization is crucial to the approximation quality, we will try to use the linear span of smooth functions as the subspace referred to above, and hope that this might work equally well or even better in some cases. In order to formulate our method, we will need a variational principle, which will be derived in the next section.

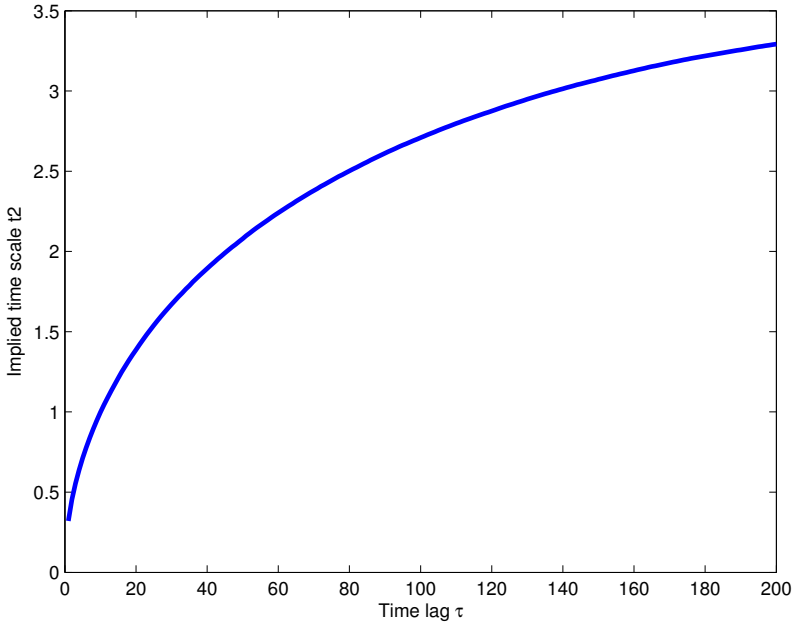


Figure II.2.: Improvement of estimated time scale t_2 upon increasing the lag time τ, displayed in multiples of the simulation length.

II.3. Variational principle

We start with the following definition:

Definition II.4: *Let f and g be observables of the state space, i.e. functions $f, g : \Omega \to \mathbb{R}$. The* **autocorrelation function** $\mathrm{acf}(f, g; \tau)$ *is defined by:*

$$\mathrm{acf}(f, g; \tau) := \int_\Omega \int_\Omega \mathrm{dx}\,\mathrm{dy}\, f(x) p(x, y; \tau) \mu(x) g(y). \tag{II.31}$$

—

The autocorrelation function of f and g is the expectation value of the product $f(x)g(y)$ with respect to the transition density $\mu(x)p(x, y; \tau)$. Furthermore, we see from Equation II.31 that it also equals the **Rayleigh coefficient** $\langle \mathcal{T}(\tau) f \mid g \rangle_\mu = \langle f \mid \mathcal{T}(\tau) g \rangle_\mu =: \langle f \mid \mathcal{T}(\tau) \mid g \rangle$, because of

$$\langle f \mid \mathcal{T}(\tau) \mid g \rangle = \langle \mathcal{T}(\tau) f \mid g \rangle_\mu = \int_\Omega \mathrm{dy}\, \frac{1}{\mu(y)} \left[\int_\Omega \mathrm{dx}\, p(x, y; \tau) \mu(x) f(x) \right] \mu(y) g(y) \tag{II.32}$$

$$= \int_\Omega \int_\Omega \mathrm{dx}\,\mathrm{dy}\, f(x) p(x, y; \tau) \mu(x) g(y) = \mathrm{acf}(f, g; \tau). \tag{II.33}$$

For a symmetric operator like $\mathcal{T}(\tau)$, the Rayleigh coefficient is also called *matrix element* or simply *expectation value*. The next two results will reveal to us the significance of autocorrelation functions when it comes to approximating eigenvalues and -functions, [Noé, 2011, sec. 2.3]:

Lemma II.5: *The transfer operator's eigenfunctions ψ_i satisfy the relation:*

$$\mathrm{acf}(\psi_i, \psi_i; \tau) = \lambda_i. \tag{II.34}$$

—

Proof. All we have to do is recognize the action of $\mathcal{T}(\tau)$ on its eigenfunctions and use the orthogonality relation between those:

$$\mathrm{acf}(\psi_i, \psi_i; \tau) = \langle \mathcal{T}(\tau)\psi_i \mid \psi_i \rangle_\mu = \lambda_i \langle \psi_i \mid \psi_i \rangle_\mu = \lambda_i. \tag{II.35}$$

■

Theorem II.6 (Variational principle): *For any function $\psi \in L^2_\mu(\Omega)$ which is orthogonal to ψ_1, we have:*

$$\mathrm{acf}(\psi, \psi; \tau) \leq \lambda_2. \tag{II.36}$$

—

Proof. First, expand ψ in terms of the orthonormal basis $\{\psi_i\}$. Note that there is no overlap with ψ_1:

$$\mathrm{acf}(\psi, \psi; \tau) = \langle \mathcal{T}(\tau)\psi \mid \psi \rangle_\mu = \sum_{i,j=2}^{\infty} c_i c_j \langle \mathcal{T}(\tau)\psi_i \mid \psi_j \rangle_\mu , \tag{II.37}$$

where $c_i = \langle \psi \mid \psi_i \rangle_\mu$. Like in the previous proof, we use the action of the transfer operator on its eigenfunctions:

$$\mathrm{acf}(\psi, \psi; \tau) = \sum_{i,j=2}^{\infty} c_i c_j \lambda_i \langle \psi_i \mid \psi_j \rangle_\mu = \sum_{i=2}^{\infty} c_i^2 \lambda_i. \tag{II.38}$$

Finally, we recall that the eigenvalues are sorted in decreasing order, and that in an orthonormal basis expansion, the squares of the coefficients sum up to unity:

$$\mathrm{acf}(\psi, \psi; \tau) \leq \lambda_2 \sum_{i=2}^{\infty} c_i^2 = \lambda_2. \tag{II.39}$$

■

Remark II.7: Likewise, we can prove that for any function ψ which is orthogonal to the first k eigenfunctions, the expression $\mathrm{acf}(\psi, \psi; \tau)$ is bounded from above by λ_{k+1}.

—

Among all candidate functions, namely those which are orthogonal to the first eigenfunction ψ_1, the true second eigenfunction maximizes the Rayleigh coefficient. For some function ψ, $\langle \mathcal{T}(\tau)\psi \mid \psi \rangle_\mu$ can be viewed as a measure of how well it approximates the second eigenfunction, and that is what we will be trying to do: Maximize the Rayleigh coefficient among a number of ansatz functions, and compare different choices of candidate functions. Variational methods of that sort are used in a number of different fields as well, e.g. for the approximation of electronic wavefunctions in quantum mechanics. In the next section, we show how the Rayleigh coefficient can be maximized within the linear span of some given test functions.

II.4. The Ritz method and the Roothan-Hall method

For all of the upcoming results, we follow [Szabo, Ostlund, 1989, ch. 1.3].

Theorem II.8 (Ritz method): *Let $\chi_1, \ldots, \chi_m \in L^2_\mu(\Omega)$ be mutually orthonormal. The Rayleigh coefficient is maximized among these function by the eigenvector $\mathbf{b}_1$ corresponding to the greatest eigenvalue ξ_1 of the matrix eigenvalue problem*

$$\mathrm{H}\,\mathbf{b}_1 = \xi_1 \mathbf{b}_1, \tag{II.40}$$

where H is the density matrix with entries

$$h_{ij} = \mathrm{acf}(\chi_i, \chi_j; \tau) = \int_\Omega \int_\Omega \mathrm{dx}\,\mathrm{dy}\; \chi_i(x) p(x, y; \tau) \mu(x) \chi_j(y). \tag{II.41}$$

More precisely, the maximal Rayleigh coefficient is found by taking the linear combination of the functions χ_i with coefficients taken from $\mathbf{b}_1$.

—

Proof. Let us first, for a function $\hat{\psi} = \sum_{i=1}^m b_i \chi_i$, express the Rayleigh coefficient in terms of the b_i:

$$\mathrm{acf}(\hat{\psi}, \hat{\psi}; \tau) = \left\langle \hat{\psi} \mid \mathcal{T}(\tau) \mid \hat{\psi} \right\rangle = \sum_{i,j=1}^m b_i b_j \langle \chi_i \mid \mathcal{T}(\tau) \mid \chi_j \rangle = \sum_{i,j=1}^m b_i b_j h_{ij}. \tag{II.42}$$

Since the matrix-elements h_{ij} are fixed, $\mathrm{acf}(\hat{\psi}, \hat{\psi}; \tau)$ can be seen as a differentiable function of the coefficients b_i. We now need to maximize this function with respect to a constraint, namely, that the solution has unit norm:

$$1 = \|\hat{\psi}\|^2 = \left\langle \hat{\psi} \mid \hat{\psi} \right\rangle_\mu = \sum_{i,j=1}^{m} b_i b_j \left\langle \chi_i \mid \chi_j \right\rangle_\mu = \sum_{i=1}^{m} b_i^2, \tag{II.43}$$

as the basis functions χ_i were assumed to be orthonormal. In summary, we need to maximize the function

$$F(b_1, \ldots, b_m) = \sum_{i,j=1}^{m} b_i b_j h_{ij} - \xi(\sum_{i=1}^{m} b_i^2 - 1), \tag{II.44}$$

where ξ is a Lagrange multiplier, with respect to the coefficients b_i. Since the H-matrix is symmetric, this results in the equations:

$$0 = \frac{\partial F}{\partial b_i} = 2\sum_{j=1}^{m} h_{ij} b_j - 2\xi b_i, \quad i = 1, \ldots, m, \tag{II.45}$$

$$\Rightarrow \xi b_i = \sum_{j=1}^{m} h_{ij} b_j, \quad i = 1, \ldots, m. \tag{II.46}$$

We have thus arrived at the eigenvalue equation $\mathrm{H}\,\mathbf{b} = \xi \mathbf{b}$. Since H is a symmetric matrix, we can find m solutions $\mathbf{b}_i$ corresponding to real eigenvalues ξ_i. If we now compute the Rayleigh coefficient of the function $\hat{\psi}_i = \sum_{j=1}^{m} b_{i,j} \chi_j$ generated from such a solution, we find:

$$\left\langle \hat{\psi}_i \mid \mathcal{T}(\tau) \mid \hat{\psi}_i \right\rangle = \sum_{j,k=1}^{m} b_{i,j} b_{i,k} \left\langle \chi_j \mid \mathcal{T}(\tau) \mid \chi_k \right\rangle = \sum_{j,k=1}^{m} b_{i,j} b_{i,k} h_{jk} \tag{II.47}$$

$$= \sum_{j=1}^{m} \xi_i b_{i,j}^2 = \xi_i, \tag{II.48}$$

where we used the eigenvalue relation $\sum_{k=1}^{m} b_{i,k} h_{jk} = \xi_i b_{i,j}$. Thus, the solution corresponding to the largest eigenvalue ξ_1 maximizes the Rayleigh coefficient among the basis functions χ_i, as we claimed.

■

Lemma II.9: *The second largest eigenvalue ξ_2 of Equation II.40 satisfies $\xi_2 \leq \lambda_2$.*

—

Proof. Let $\hat{\psi}_1$ and $\hat{\psi}_2$ be the functions generated from the eigenvectors $\mathbf{b}_1$, $\mathbf{b}_2$ and their eigenvalues ξ_1, ξ_2, respectively. Consider a linear combination $\hat{\psi} := x\hat{\psi}_1 + y\hat{\psi}_2$, for $x, y \in \mathbb{R}$. If $\hat{\psi}$ is normalized, we find, by orthonormality:

$$1 = \left\langle \hat{\psi} \mid \hat{\psi} \right\rangle_\mu = x^2 \left\langle \hat{\psi}_1 \mid \hat{\psi}_1 \right\rangle_\mu + y^2 \left\langle \hat{\psi}_2 \mid \hat{\psi}_2 \right\rangle_\mu + 2xy \left\langle \hat{\psi}_1 \mid \hat{\psi}_2 \right\rangle_\mu = x^2 + y^2. \tag{II.49}$$

Moreover, as $\hat{\psi}_1$ and $\hat{\psi}_2$ diagonalize the H-matrix:

$$\left\langle \hat{\psi}_1 \mid \mathcal{T}(\tau) \mid \hat{\psi}_2 \right\rangle = \sum_{i,j=1}^{m} b_{1,i} b_{2,j} \left\langle \chi_i \mid \mathcal{T}(\tau) \mid \chi_j \right\rangle = \sum_{i,j=1}^{m} b_{1,i} b_{2,j} h_{ij} \tag{II.50}$$

$$= \sum_{i=1}^{m} \xi_2 b_{1,i} b_{2,i} = 0. \tag{II.51}$$

Therefore, computing the Rayleigh coefficient for $\hat{\psi}$ yields:

$$\left\langle \hat{\psi} \mid \mathcal{T}(\tau) \mid \hat{\psi} \right\rangle = x^2 \left\langle \hat{\psi}_1 \mid \mathcal{T}(\tau) \mid \hat{\psi}_1 \right\rangle + y^2 \left\langle \hat{\psi}_2 \mid \mathcal{T}(\tau) \mid \hat{\psi}_2 \right\rangle + 2xy \left\langle \hat{\psi}_1 \mid \mathcal{T}(\tau) \mid \hat{\psi}_2 \right\rangle \tag{II.52}$$

$$= x^2 \xi_1 + y^2 \xi_2 = (1 - y^2)\xi_1 + y^2 \xi_2 = \xi_1 - y^2(\xi_1 - \xi_2). \tag{II.53}$$

Depending on the choice of y, this value is between ξ_2 and ξ_1. If we now had $\lambda_2 < \xi_2$, the Rayleigh coefficient of $\hat{\psi}$ would always be greater than λ_2. But this means that no normalized linear combination of $\hat{\psi}_1$ and $\hat{\psi}_2$ could be orthogonal to ψ_1, as, by Theorem II.6, we would then find the Rayleigh coefficient to be bounded by λ_2. But this is impossible, one can always find a linear combination of two given vectors which is orthogonal to a given third vector.

■

Clearly, this Lemma generalizes to the third, fourth, ... eigenvalue.

Once we are starting out with a basis set $\{\chi_i\}$, the linear combination stemming from the eigenvector $\mathbf{b}_1$ is an approximation of the first eigenfunction $\psi_1 = \mathbb{1}$, and the eigenvalue ξ_1 should be close to $\lambda_1 = 1$. Moreover, because of the last result, we can use the function built from $\mathbf{b}_2$ as an approximation of ψ_2, and ξ_2 becomes an estimate of the real eigenvalue λ_2. This procedure can be repeated for the following eigenvectors. From this, we learn that maximization of the Rayleigh coefficient really means to maximize the approximation quality, because if ξ_2 is an estimate of λ_2, then $-\frac{\tau}{\log \xi_2}$ is also an estimate of the implicit time scale t_2. We know that the larger the estimated time scale and the faster it converges upon increasing τ, the better the approximation. In this sense, the estimated eigenvalues $\xi_2, \xi_3, \ldots$ really measure how good the approximation from this basis set works.
The advantage of the preceding results is that the entries of the density matrix can be estimated from finite sampling trajectories. Since these are the Rayleigh coefficients of the basis functions, and since Rayleigh coefficients are expectation values with respect to the transition density $\mu(x)p(x, y; \tau)$, they can be estimated from a trajectory $X_1, \ldots, X_N$ by

$$\mathrm{acf}(\chi_i, \chi_j; \tau) \sim \frac{1}{N - \tau} \sum_{i=1}^{N-\tau} \chi_i(X_i)\chi_j(X_{i+\tau}). \tag{II.54}$$

Here, τ has to be chosen as an integer, and as a lag time thus becomes an integer multiple of the simulation window Δt.

The only problem left is the requirement that the basis functions were so far assumed to orthonormal. However, this can easily be circumvented. Again, let $\chi_1, \ldots, \chi_m \in L^2_\mu(\Omega)$ be normalized basis functions as before, except that they are no longer assumed to be orthogonal. We can still derive a similar result:

Theorem II.10 (Roothan-Hall method): *For basis functions as above, the linear combination that maximizes the Rayleigh coefficient is given by the solution* $\mathbf{b}_1$ *corresponding to the greatest eigenvalue* ξ_1 *of the generalized eigenvalue problem*

$$\mathrm{H}\,\mathbf{b}_1 = \xi_1\,\mathrm{S}\,\mathbf{b}_1, \tag{II.55}$$

where H *is as in Theorem II.8 and* S *is the overlap matrix with entries*

$$s_{ij} = \langle \chi_i \mid \chi_j \rangle_\mu . \tag{II.56}$$

—

Proof. Starting out the same way as in Theorem II.8, we first have to reformulate the normalization constraint. Equation II.43 then becomes:

$$1 = \|\hat{\psi}\|^2 = \sum_{i,j=1}^{m} b_i b_j \langle \chi_i \mid \chi_j \rangle_\mu = \sum_{i,j=1}^{m} b_i b_j s_{ij}. \tag{II.57}$$

This results in an effective functional of the form:

$$F(b_1, \ldots, b_m) = \sum_{i,j=1}^{m} b_i b_j h_{ij} - \xi(\sum_{i,j=1}^{m} b_i b_j s_{ij} - 1). \tag{II.58}$$

Maximization requires the solution of:

$$0 = 2\sum_{j=1}^{m} b_j h_{ij} - 2\xi \sum_{j=1}^{m} b_j s_{ij}. \tag{II.59}$$

$$\Rightarrow \xi \sum_{j=1}^{m} b_j s_{ij} = \sum_{j=1}^{m} b_j h_{ij}, \tag{II.60}$$

for every $i \in \{1, \ldots, m\}$. This is the generalized eigenvalue equation stated above. Since H is symmetric and S is positive definite, this problem has m real eigenvalues ξ_i and corresponding eigenvectors $\mathbf{b}_i$ which are orthonormal with respect to the S-weighted scalar product, i.e. we have $\mathbf{b}_i^T\,\mathrm{S}\,\mathbf{b}_j = \delta_{ij}$. Therefore, similar to Equation II.47 - Equation II.48, we find for the corresponding functions $\hat{\psi}_i := \sum_{j=1}^{m} b_{i,j}\chi_j$:

$$\left\langle \hat{\psi}_i \mid \mathcal{T}(\tau) \mid \hat{\psi}_i \right\rangle = \sum_{j,k=1}^{m} b_{i,j} b_{i,k} \langle \chi_j \mid \mathcal{T}(\tau) \mid \chi_k \rangle = \sum_{j,k=1}^{m} b_{i,j} b_{i,k} h_{jk} \tag{II.61}$$

$$= \xi_i \sum_{j,k=1}^{m} b_{i,j} b_{i,k} s_{j,k} = \xi_i \mathbf{b}_i^T\,\mathrm{S}\,\mathbf{b}_i = \xi_i. \tag{II.62}$$

So again, by taking the solution $\mathbf{b}_1$ with maximal eigenvalue ξ_1, we have found the linear combination that maximizes the Rayleigh coefficient.

■

Clearly, the Ritz method is just a special case of the Roothan-Hall method, where S is given by the $m \times m$-identity matrix. With the Roothan-Hall method, we can use any basis set, orthonormal or not. Like the density matrix H, the overlap matrix S can also be estimated from finite sampling. Since we are mostly interested in the propagator eigenfunctions, we will weight the solution with the stationary distribution afterwards. In most cases, we will have to estimate μ from a sample trajectory as well. Our method consequently consists of the following steps:

(a) Compute a sufficiently long or sufficiently many short sampling trajectories of the process and check convergence.

(b) Estimate the stationary distribution from the sample.

(c) Estimate the H- and S-matrix from the sample.

(d) Solve the generalized eigenvalue problem from Theorem II.10.

(e) Weight the solution with the estimated stationary density.

Lastly, we see that the Markov state model approximation is also obtained from using the Ritz method with indicator functions. So, like in section II.2, choose a partition of the state space into s mutually disjoint sets $A_1, \ldots, A_s$, and let the candidate basis functions $\chi_1, \ldots, \chi_s$ be given by the indicator functions of these sets, normalized by a pre-factor $\frac{1}{\sqrt{\pi_i}}$, with $\pi_i = \int_{A_i} \mathrm{dx}\, \mu(x)$. Computing the density matrix H yields:

$$h_{ij} = \frac{1}{\sqrt{\pi_i \pi_j}} \int_{A_i} \mathrm{dx} \int_{A_j} \mathrm{dy}\, \mu(x) p(x, y; \tau) \tag{II.63}$$

$$= \frac{\sqrt{\pi_i}}{\sqrt{\pi_j}} \frac{\mathbb{P}(X_\tau \in A_j, X_0 \in A_i)}{\mathbb{P}(X_0 \in A_i)} = \frac{\sqrt{\pi_i}}{\sqrt{\pi_j}} p_{ij}(\tau). \tag{II.64}$$

The density matrix is the same as the MSM transfer matrix up to a similarity transformation with transformation matrix $\Pi^{\frac{1}{2}}$, which is the diagonal matrix with the square roots of the local stationary probabilities as diagonal entries. The Ritz method therefore generates the right-hand eigenvalues of the MSM transfer matrix and is equivalent to the construction of a Markov state model.

III. Diffusion processes and application in one dimension

We are now prepared to apply the variational methods we have just derived to a number of example systems. Our guiding example will be a diffusion process. This is a system where a number of classical particles move under the influence of a deterministic force field, generated by some potential energy function. If there was no further influence, this system could be modelled by classical equations of motion. However, the particles are also subject to random fluctuations, which interfere with the deterministic motion. The fundamental concepts of such a system will be introduced below. As a result, the system trajectory can no longer be uniquely predicted. Instead, we have a stochastic process to which our theoretical results can be applied.

III.1. Brownian motion and stochastic differential equations

The fundamental model for the description of random fluctuations is **Brownian motion** or the **Wiener process**. In 1826-27, R. Brown studied the irregular behaviour of pollen grains in water. He and many others observed that the pathways taken by such a particle were highly irregular. Moreover, whereas the average distance from the starting point seemed to vanish, the mean fluctuation from the average seemed to be growing linearly. These observations motivated the following mathematical model, [Evans, ch. 3].

Definition III.1: *A one-dimensional stochastic process W_t on a probability space Ω, defined for $t \geq 0$, is called a Brownian motion or a Wiener process if it satisfies:*

(a) $W_0 = 0$ *a.e.*

(b) For any times $t_1, \dots, t_k$, the increments $W_{t_k} - W_{t_{k-1}}, \dots, W_{t_2} - W_{t_1}$ are independent random variables.

(c) The increments $W_t - W_s$ are normally distributed according to $W_t - W_s \sim \mathcal{N}(0, t-s)$, $\forall t > s \geq 0$.

—

An $\mathbb{R}^n$-valued stochastic process where each component is a one-dimensional Wiener process is called n-dimensional Wiener process. In the 1920s, N. Wiener constructed a Brownian motion and thereby mathematically proved the existence of such a process. In fact, there is a number of different ways how this can be achieved, but the explicit realization is of little importance. All we need to know are the three conditions from the above definition and the most important properties. It can be shown that a Brownian motion defines a Markov process, that the sample paths, i.e. the functions $t \mapsto W_t$, are continuous, but almost everywhere in Ω nowhere differentiable with respect to time, see Figure III.1 and [Behrends, 2011, ch. 5.2]. This model has turned out to be a very suitable for many different phenomena, reaching way beyond the simulation of particles in a fluid. For instance, Brownian motion has frequently been used to model financial markets.

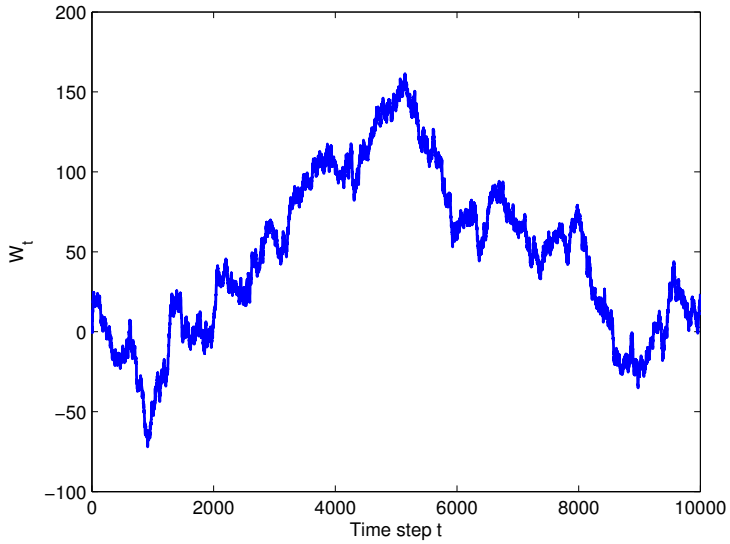


Figure III.1.: A sample path of the Wiener process. Clearly, the path is continuous but not differentiable as a function of time.

Despite the irregularity of its sample paths, Brownian motion can be used to model phenomena involving differential calculus. Suppose that the change of a quantity $X(t)$ over a short

time Δt is given by

$$\Delta X(t) = \Delta t F(X(t), t) + \sigma W_t, \tag{III.1}$$

meaning that the change is composed of a deterministic function F depending on the quantity $X(t)$ and the time plus a random fluctuation, modelled by Brownian motion. Here, $\sigma > 0$ is simply a control parameter governing the strength of the perturbation. Upon dividing by Δt and passing to the limit $\Delta t \to 0$, one is tempted to write down a differential equation

$$\frac{dX(t)}{dt} = F(X(t), t) + \sigma dW_t, \tag{III.2}$$

where we would like to have dW_t denote the time derivative of Brownian motion. We have just seen that such a derivative does not exist. However, Equation III.2 does make sense if it is transformed into an integral form. This involves the definition of a stochastic integral and the Ito formula, [Behrends, 2011, ch. 6, 7]. In particular, Brownian motion itself can be shown to solve Equation III.2 in this sense, if $F = 0$.

III.2. Diffusion process

Consider a classical particle of mass m moving under the influence of friction $\gamma > 0$ and of a force field $\mathbf{F}$, generated by a potential energy function V, i.e. $\mathbf{F} = -\nabla V$. By Newton's equation, the position vector $\mathbf{x}(t) \in \mathbb{R}^3$ satisfies the relation

$$m\frac{d^2\mathbf{x}(t)}{dt^2} = -m\gamma\frac{d\mathbf{x}(t)}{dt} - \nabla V(\mathbf{x}(t)). \tag{III.3}$$

Let us now suppose that there are random perturbations, caused, for instance, by collisions with small particles, which continuously alter the particle's motion. We therefore make Equation III.3 a stochastic differential equation by adding σdW_t,

$$m\frac{d^2\mathbf{x}(t)}{dt^2} = -m\gamma\frac{d\mathbf{x}(t)}{dt} - \nabla V(\mathbf{x}(t)) + \sigma dW_t, \tag{III.4}$$

keeping in mind that it is meant to be solved in an integral sense. It is known that $\sigma = \sqrt{2m\gamma k_B T}$, where $k_B = 1.3806488\mathrm{J/K}$ is Boltzmann's constant and T is the system temperature. This equation is typically called the **Langevin equation** of our system, the solution $\mathbf{x}(t) =: \mathbf{X}_t$ is a Markov process, called a **diffusion process**. Equation III.4 holds in the same way if we study a system of n particles, in this case, $\mathbf{X}_t$ is a stochastic process in $3n$ dimensions. If the friction parameter γ is large enough, it is justified to drop the second derivative on the left-hand side of Equation III.4, [Zwanzig, 2001, sec. 2.2]. Rearranging the terms modifies the equation to

$$\frac{d\mathbf{x}(t)}{dt} = -\frac{1}{m\gamma}\nabla V(\mathbf{x}(t)) + \sqrt{2D}dW_t, \tag{III.5}$$

where we have defined $D := \frac{k_B T}{m\gamma}$ This equation is the Langevin equation in the high friction limit, the underlying dynamics is called **Smoluchowski dynamics** or also **Brownian dynamics**. It will be the interest of our studies in what follows.
The Langevin equation itself can hardly be solved, except in a few rare cases. However, it is helpful to restrict oneself to the study of probability densities. Instead of trying to find an expression for the stochastic process $\mathbf{X}_t$ itself, we are only interested in computing the average probability $\rho(\mathbf{x}, t)$ to find the particle in a position $\mathbf{x}$ at time t. This probability density also satisfies a differential equation, called the Smoluchowski equation:

Theorem III.2 (Smoluchowski equation): *The probability $\rho(\mathbf{x}, t)$ to find a system obeying the Langevin Equation III.5 at position $\mathbf{x}$ at time t satisfies the differential equation:*

$$\frac{\partial \rho(\mathbf{x}, t)}{\partial t} = \frac{1}{m\gamma} \nabla \cdot (\nabla V(\mathbf{x}) \rho) + D \Delta \rho(\mathbf{x}, t). \tag{III.6}$$

Here, $\nabla \cdot$ denotes the divergence of a vector field and Δ the Laplace operator of a scalar field.

—

Proof. A derivation can be found in [Zwanzig, 2001, ch 2.2].

■

Equation III.6 is a special case of a **Fokker-Planck equation**. This result enables us to find the stationary density for the diffusion process:

Theorem III.3: *The stationary solution of the Smoluchowski Equation III.6 is the* **Boltzmann distribution**

$$\mu(\mathbf{x}) = \frac{1}{Z} \exp\left(-\beta V(\mathbf{x})\right), \tag{III.7}$$

where $\beta := \frac{1}{k_B T}$ is the inverse Boltzmann temperature and $Z := \int_\Omega dx \exp(-\beta V(\mathbf{x}))$ is a normalization constant, called the **partition function**.

—

Proof. If $\frac{\partial \rho}{\partial t} = 0$ holds, we have to check that

$$-\frac{1}{m\gamma} \nabla \cdot (\nabla V(\mathbf{x}) \mu(\mathbf{x})) = D \Delta \mu(\mathbf{x}). \tag{III.8}$$

We find:

$$D\Delta\mu(\mathbf{x}) = \frac{k_B T}{m\gamma Z} \sum_i \frac{\partial}{\partial x_i} \left[-\beta \frac{\partial V(\mathbf{x})}{\partial x_i} \exp\left(-\beta V(\mathbf{x})\right) \right] \tag{III.9}$$

$$= \frac{k_B T}{m\gamma Z} \left[-\beta \Delta V(\mathbf{x}) \exp\left(-\beta V(\mathbf{x})\right) + \sum_i \beta^2 \left(\frac{\partial V(\mathbf{x})}{\partial x_i} \right)^2 \exp\left(-\beta V(\mathbf{x})\right) \right] \tag{III.10}$$

$$= -\frac{1}{m\gamma} \left[\Delta V(\mathbf{x}) \mu(\mathbf{x}) + \nabla V(\mathbf{x}) \cdot \nabla \mu(\mathbf{x}) \right] \tag{III.11}$$

$$= -\frac{1}{m\gamma} \nabla \cdot \left(\nabla V(\mathbf{x}) \mu(\mathbf{x}) \right). \tag{III.12}$$

■

We note that μ has to be well-defined for this result to make sense, i.e. the integral defining Z has to exist. This mainly requires that $V(\mathbf{x}) \to \infty$ fast enough as $|\mathbf{x}| \to \infty$. If the state space is dynamically connected, which we will always assume, this can also be shown to be sufficient for the process to be ergodic and meet the requirements from chapter II, [Sarich, Noé, Schütte, 2010, ch. 2.2]. Therefore, we have just shown that μ is the stationary distribution of the diffusion process, and we can now apply the foregoing theory.

III.3. Numerical considerations

Knowing the invariant distribution of a diffusion process helps us a lot, because our computational result for the first eigenfunction can be compared to the function $\exp(-\beta \nabla V(\mathbf{x}))$, and should be almost equal to it up to a pre-factor. However, little more can be computed in advance, except for a few simple systems. In particular, the transition kernel $p(x, y; \tau)$ is unknown to us. But in order to apply the variational methods, we need to perform a numerical simulation of the process, which means that if the current simulation step is $\mathbf{x} \in \Omega$, we need to draw the next step with time lag Δt from the distribution $p(\mathbf{x}, \cdot\,; \Delta t)$. We therefore approximate this distribution by a very common procedure, the so-called Euler-Maruyama method. It consists of combining a deterministic explicit Euler step with a normally distributed random step. If the current simulation position is $\mathbf{x}$, the deterministic Euler step is

$$\mathbf{x} \mapsto \mathbf{x} - \Delta t \nabla V(\mathbf{x}). \tag{III.13}$$

This approximates the position that the process would attain if it was unperturbed. But since it is subject to normally distributed noise which, after time Δt, has variance $2D\Delta t$, we add a normally distributed random number with standard deviation $\sqrt{2D\Delta t}$. Altogether,

the simulation step becomes

$$\mathbf{x} \mapsto \mathbf{x} - \Delta t \nabla V(\mathbf{x}) + \sqrt{2D\Delta t}\eta, \tag{III.14}$$

where $\eta \sim \mathcal{N}(0,1)$. Clearly, this is just an approximation of the true dynamics, but a good one as long as Δt is chosen small enough. However, having in mind the difficulties arising from the use of the explicit Euler method to unperturbed dynamical systems, it can also lead to serious trouble, a problem we will encounter in the next chapter.
We are left with the question of how to choose an appropriate basis set. In the previous chapter, we have seen that the eigenfunctions ϕ_i of the propagator and the eigenfunctions ψ_i of the transfer operator differ only by a factor of the stationary density. An approximation scheme using the Roothan-Hall method results in an estimate of the functions ψ_i. But since it is our goal to determine the propagator eigenfunctions ϕ_i, we can either start with a set of functions suitable to approximate the functions ψ_i, and weight the result with a factor μ afterwards, or we can go the other way round. That means to start with a number of functions suitable to estimate the ϕ_i-functions and divide them by μ prior to the computation. After the computation, we simply take the resulting linear combination of our original basis functions to estimate the functions ϕ_i. There is also a third way: If $f \in L^2(\Omega)$ is an element of the unweighted L^2-space, then $\frac{f}{\sqrt{\mu}}$ is in $L^2_\mu(\Omega)$ on the one hand, and $\sqrt{\mu}f$ is in $L^2_{\mu^{-1}}(\Omega)$ on the other hand. These functions are somehow "in between" the two domains, and they can also be used with the Roothan-Hall method. To this end, they have to be divided by the square root of μ before the computation, and the resulting linear combination has to be re-weighted by $\sqrt{\mu}$ afterwards.
Our approach to all of the following problems will be to make use of the shape of the energy function. The function V will always display a number of minima, which correspond to metastable regions of the state space. We will try to use local functions, like Gaussian hats, which are essentially non-zero only in a neighbourhood of the metastable regions. Since functions in $L^2_{\mu^{-1}}(\Omega)$ and $L^2(\Omega)$ at least have to decay to zero if $|x|$ becomes large, whereas functions in $L^2_\mu(\Omega)$ can be constant or even unbounded, we will prefer one of the latter two ways. Additionally, using unbounded functions can lead to instabilities if rare states outside the metastable regions are visited by a sampling trajectory, because these functions can become very large for such states. In the experiments, both of the two preferred ways seemed to perform well. All of the computations shown below were ran using Gaussian functions weighted with the square root of the stationary density.

III.4. Diffusion in a quadratic potential

A simple but important example is diffusion in a quadratic potential $V(x) = \frac{1}{2}bx^2$, for $x \in \mathbb{R}$ and $b > 0$. Inserting this into Equation III.7 shows that $\mu(x) = \frac{1}{\sqrt{2\pi\alpha}} \exp\left(-\frac{x^2}{2\alpha}\right)$, the stationary distribution becomes a Gaussian distribution with zero mean and variance $\alpha = \frac{k_B T}{b}$. The potential and its invariant measure are depicted in Figure III.2. It turns out

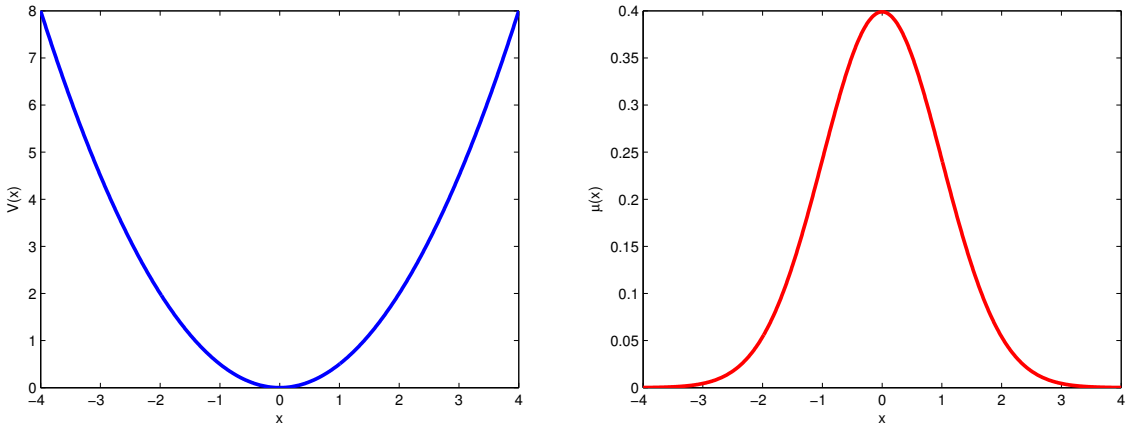


Figure III.2.: Quadratic potential $V(x) = 0.5x^2$ and its invariant distribution.

that this is one of the rare cases where all important quantities can be determined analytically. Using stochastic integrals and Ito's formula, we derive an explicit expression for this process in Lemma A.1. Using this expression, we then show that if the process starts with a distribution ρ_0, the expectation value at time t is given by $\mathbb{E}[\rho_t] = e^{-\theta t}\mathbb{E}[\rho_0]$, where we have defined $\theta := \frac{b}{m\gamma}$. Moreover, the variance evolves according to $\omega^2[\rho_t] = e^{-2\theta t}\omega^2[\rho_0] + (1 - e^{-2\theta t})\alpha$. Regardless of the initial distribution, the process quickly approaches the stationary Gaussian distribution. Even the complete set of eigenfunctions can be found analytically:

Lemma III.4: *The propagator eigenfunctions are given by the appropriately scaled Hermite polynomials, multiplied with the invariant measure:*

$$\phi_i(x) = \sqrt{\frac{\alpha^{i-1}}{(i-1)!}} H_{i-1}\left(\frac{x}{\sqrt{\alpha}}\right)\mu(x). \tag{III.15}$$

The corresponding eigenvalues are $\lambda_i(\tau) = e^{-\theta(i-1)\tau}$.

Proof. We also show the proof of this Lemma in the appendix. ■

Since there is only one potential minimum, there are no metastable states. Unless b is very small or mass and friction are very large, the global relaxation time $-\frac{\tau}{\log\lambda_2(\tau)} = \frac{\tau}{\theta\tau} = \frac{m\gamma}{b}$ is short, the process quickly equilibrates to its invariant distribution.
Since all these analytic quantities are at hand, we can compare them to the results obtained from the Roothan-Hall method. As shown in Figure III.3, a relatively small numerical effort leads to a good approximation of the eigenfunctions as well as the corresponding eigenvalues and implied time scales. Though this example may be a simple one, it is still important. In

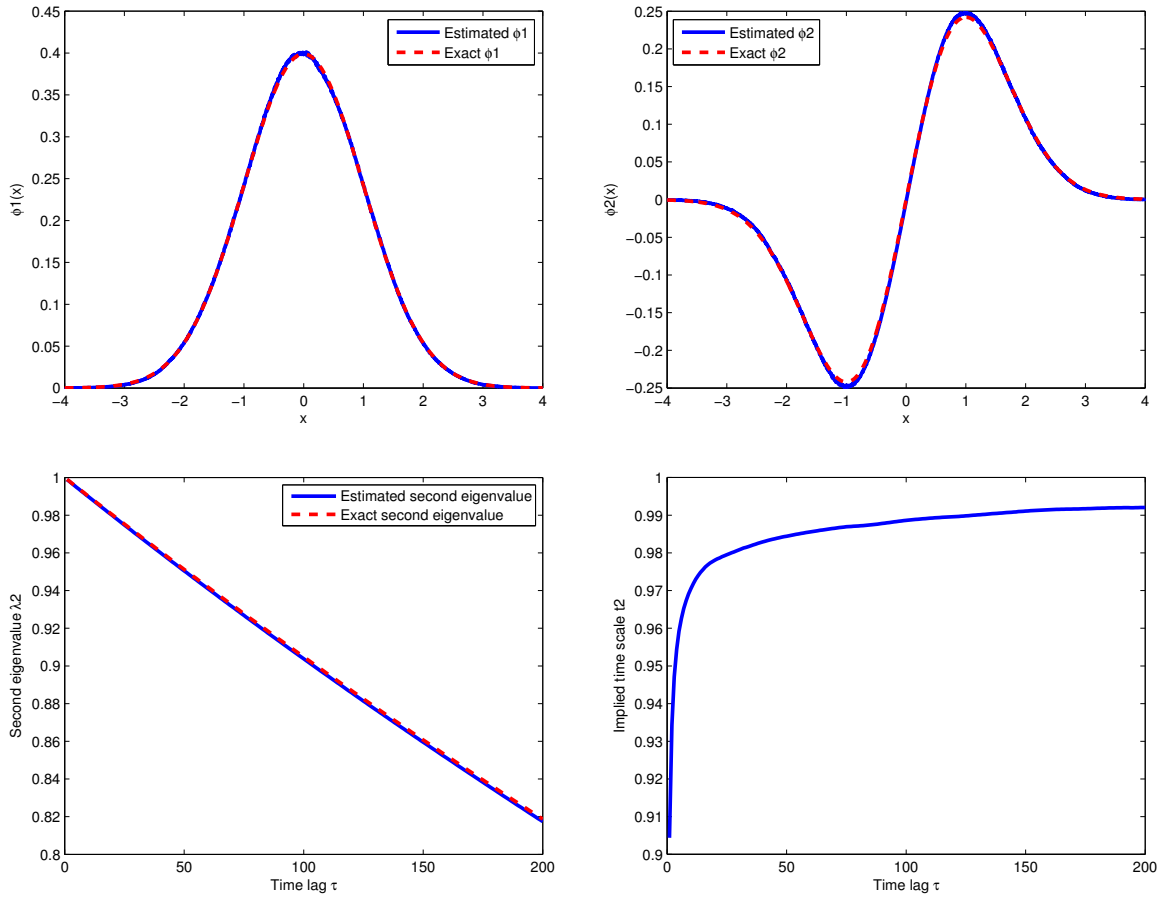


Figure III.3.: Results obtained from applying the Roothan-Hall method to diffusion in a harmonic potential. A simulation of 20 million time steps of length $\Delta t = 10^{-3}$ was used with thirteen Gaussian test functions centred at -3, -2, -1, -0.8, -0.4, -0.2, 0, 0.2, 0.4, 0.8, 1, 2, 3. The variances were set to 0.5 for all centres between -0.8 and 0.8, and to 1 for the remaining functions. (a) First eigenfunction ϕ_1. (b) Second eigenfunction ϕ_2. (c) Second eigenvalue λ_2. (d) Estimated second implied time scale t_2. Exact value is $t_2 = 1$.

molecular simulations, a harmonic potential is often used to model bonds between atoms, as well as bond angles. The harmonic potential will therefore frequently show up in applications.

III.5. Two-well potential

Another important example is a two-well potential. In general, this is some potential function displaying two minimum positions which are separated by a significant energy barrier, and which rises to infinity both to the left and to the right of the minima. For our calculations, we choose $V(x) = k(x^4 - 2x^2 + 1)$ for some $k > 0$, which has two minima at $x = -1$ and $x = 1$. The energy function and the invariant distribution are displayed in Figure III.4a and Figure III.4b. The two-well potential is a good model for certain molecular interactions where a certain degree of freedom favours two characteristic positions. The regions around the potential minima correspond to metastable states, we therefore expect one dominant slow process apart from the stationary process. We apply the Roothan-Hall method using thirteen Gaussian functions centred around the two minima, with uniform variance equal to 0.5. In order to evaluate the results, we also computed the eigenvalues of an MSM transition matrix. Here, we used a fine discretization of the state space into 100 sets, most of them situated close to the potential minima. A comparison of the results can be seen in Figure III.4. Most importantly, we see that the implied time scales are estimated very well. Both methods converge quickly to about the same value. The stationary distribution is very well approximated, and we obtain a convincing result for the second eigenfunction, clearly displaying a characteristic sign change between the two metastable regions. The computational effort required by the Roothan-Hall method is much smaller, since we only need to compute an eleven by eleven matrix instead of a 100×100 matrix.

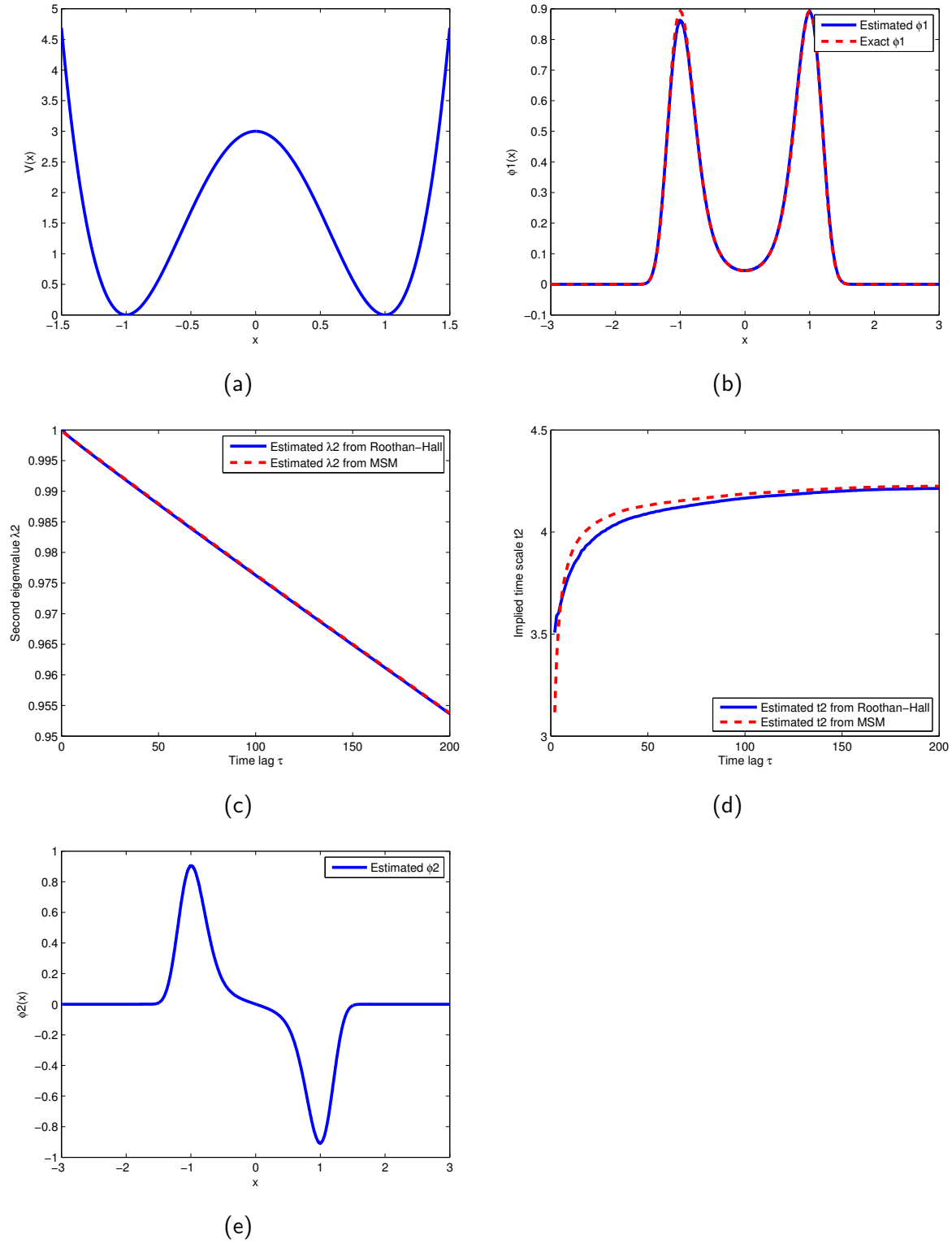


Figure III.4.: Application of the Roothan-Hall method to the one-dimensional two-well potential. A simulation of 20 million time steps with $\Delta t = 10^{-3}$ was used with thirteen Gaussian hats centred at -2, -1.5, -1.2, -1, -0.8, -0.5, 0, 0.5, 0.8, 1, 1.2, 1.5, 2. The variances were set to 1 for the centres at $x = -2, -1.5, 0, 1.5, 2$ and to 0.5 for all others. We compare the results with a 100 set MSM discretization. (a) Potential function $V(x)$. (b) First eigenfunction ϕ_1. (c) Second eigenvalue λ_2. (d) Second implied time scale t_2. (e) Second eigenfunction ϕ_2.

IV. Application to molecules

In this chapter, we are going to apply the variational method to diffusion processes in a higher dimensional state space. For systems with many degrees of freedom, the potential energy function often possesses multiple minima, and contains at times highly complicated interactions between several of the coordinates. Evaluation of the partition function becomes very difficult, and requires specialized algorithms like Markov chain Monte Carlo methods. Computing eigenvalues and characteristic time scales can be done using Markov state models, but the choice of adequate sets requires suitable clustering methods. This also leads to a huge computational effort if the state space is very high-dimensional. We have the hope that the use of variational methods can help to reduce this effort.

IV.1. The example system

Let us consider a system which is like a very much simplified small molecule. Let it consist of N atoms, with N being small, either equal to 4 or to 5 in what follows. Denote their position vectors by $\mathbf{r}_i \in \mathbb{R}^3$, $i \in \{1, \ldots, N\}$. Neighbouring atoms are connected by a bond. Denote the distance vectors between those atoms by $\mathbf{r}_{ij} := \mathbf{r}_j - \mathbf{r}_i$ and the distances by $r_{ij} := \|\mathbf{r}_{ij}\|$. For three neighbouring atoms, we define the bond angle θ_{ijk} by

$$\theta_{ijk} := \cos^{-1}\left(-\frac{\langle \mathbf{r}_{ij} \mid \mathbf{r}_{jk} \rangle}{r_{ij} r_{jk}}\right), \tag{IV.1}$$

which is the angle between $\mathbf{r}_{ij}$ and $\mathbf{r}_{jk}$. Additionally, for four atoms, we can define the **dihedral angle** ψ_{ijkl} as the angle between the plane spanned by $\mathbf{r}_{ij}$, $\mathbf{r}_{jk}$ and the one spanned by $\mathbf{r}_{jk}$, $\mathbf{r}_{kl}$. Using the normal vectors $\mathbf{n}_{ijk}$ and $\mathbf{n}_{jkl}$, given by

$$\mathbf{n}_{ijk} := \mathbf{r}_{ij} \times \mathbf{r}_{jk}, \tag{IV.2}$$

which are perpendicular to the respective planes, we find that the dihedral is the angle between the two normal vectors:

$$\psi_{ijkl} = \cos^{-1}\left(\frac{\langle \mathbf{n}_{ijk} \mid \mathbf{n}_{jkl} \rangle}{\|\mathbf{n}_{ijk}\| \|\mathbf{n}_{jkl}\|}\right). \tag{IV.3}$$

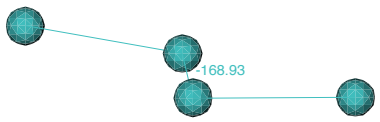


Figure IV.1.: Schematic drawing of the four atom system, corresponding either to system A or B. It shows the dihedral angle in blue. Created with [VMD, 1996].

A sketch of such a system is shown in Figure IV.1.

Let us now define a potential function V, which generates a force field acting on the molecule. A bond between two atoms is typically modelled like a spring. Therefore, we define harmonic potentials V_{ij}, depending on the distance r_{ij}, for each pair of connected atoms:

$$V_{ij} := \frac{1}{2} k_{ij} \left(d_{ij} - r_{ij}\right)^2 . \tag{IV.4}$$

Here, k_{ij} is a constant defining the strength of the potential and d_{ij} is the distance of minimal energy. The system will drive the bond lengths towards the minimum distances d_{ij}. Similarly, we set up harmonic potentials V_{ijk} for each bond angle:

$$V_{ijk} := \frac{1}{2} k_{ijk} \left(d_{ijk} - \theta_{ijk}\right)^2 . \tag{IV.5}$$

The dihedral angles usually have a number of favoured positions. In order to model this, we define the dihedral potential as

$$V_{ijkl} := k_{ijkl} \left(1 - \cos(n\psi_{ijkl})\right) , \tag{IV.6}$$

for $\psi_{ijkl} \in [-\pi, \pi)$. The position $\psi_{ijkl} = \pi$ is identical to $\psi_{ijkl} = -\pi$, consequently, it is enough to have V_{ijkl} defined on the above interval. This potential has n minimum positions,

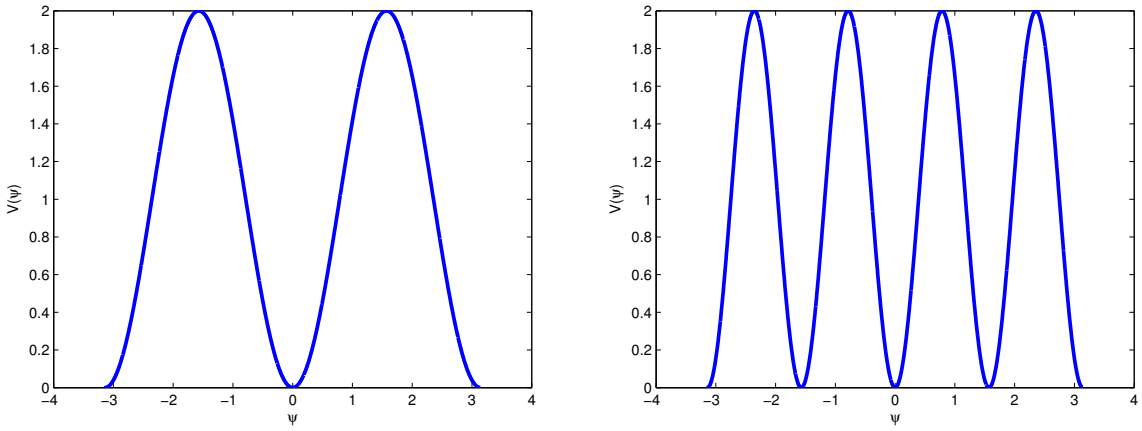


Figure IV.2.: The potential energy function $V(\psi_{ijkl}) = k_{ijkl}(1-\cos(n\psi_{ijkl}))$ for the dihedral angle, with $k_{ijkl} = 1$ and $n = 2$, $n = 4$, respectively.

as displayed in Figure IV.2. If $n = 2$, these two minima are situated at $\psi_{ijkl} = 0; -\pi$. Lastly, we would also like to include Coulomb potentials V_C defined by:

$$V_C(r_{ij}) = \frac{1}{4\pi\epsilon_0}\frac{e_1 e_2}{r_{ij}}, \tag{IV.7}$$

which describe the electrostatic repulsion or attraction between two atoms. Here, ϵ_0 is the electric constant, and e_1, e_2 are the charges of the two atoms. In our simulation, the charges will be equal to one elementary charge $e = 1.602 \cdot 10^{-19}\mathrm{C}$ in absolute value, but can have opposite signs. The total potential energy function V is then the sum of all individual energies.
For our computations, we will consider three small systems, which we call **system A**, **system B** and **system C**. The first is a four atom system which only includes the three harmonic bond interactions, two bond angle interactions, and a dihedral potential with $n = 2$. The second one also consists of four atoms, but we set $n = 4$ for the dihedral and add an attractive Coulomb interaction between atoms one and four. In the last case, we have $N = 5$, four bonds, three bond angles and two dihedrals, one with $n = 4$ and the other with $n = 2$.

If we now add random perturbations and Smoluchowski dynamics, we obtain a diffusion process with invariant distribution

$$\mu(\mathbf{x}) = \frac{1}{Z}\exp(-\beta V(\mathbf{x})), \tag{IV.8}$$

where Z is the partition function, as before, and $\mathbf{x} \in \mathbb{R}^{3N}$ is the full position vector containing all coordinates of the atoms.
It is important to point out once more that the stationary distribution and all the other eigenfunctions are defined on the $3N$-dimensional Euclidean space. As such, they probably take a highly complicated shape. At least for our examples however, the state of the system

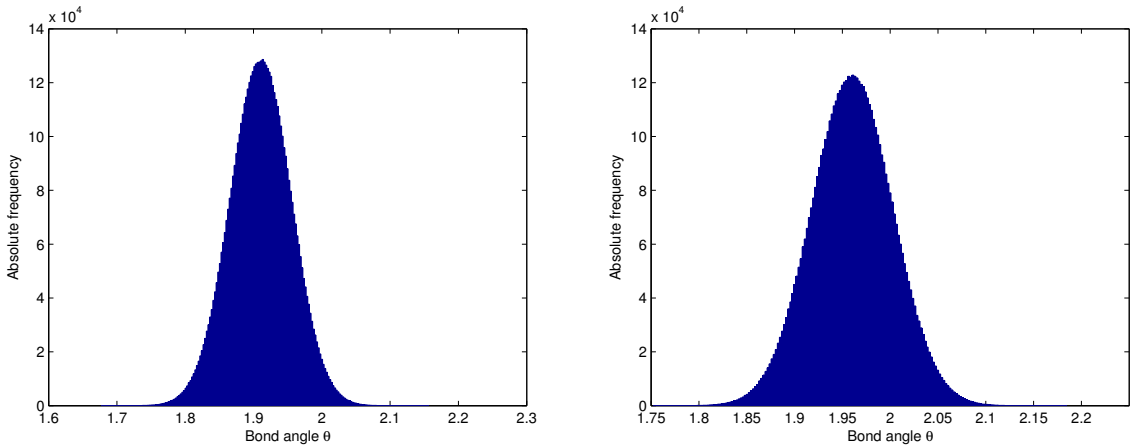


Figure IV.3.: Statistical distribution of the two bond angles of a four atom system with Coulomb 1-4 attraction. The position of minimum energy is $\theta = 1.9373 = 111.00\,\mathrm{deg}$. We see that the interaction causes a slight shift of both distributions, but because of the strong bonds, the Gaussian type distribution is well conserved. Clearly, the changes of the distributions caused by the interactions can be much more drastic, but still we hope that they will not altogether destroy the original shape.

as well as its potential energy is completely determined by the bond lengths, the bond angles and the dihedral angles. These coordinates are called the **internal coordinates**. Projected onto these internal coordinates, the eigenfunctions will probably have a much simpler form. Clearly, these projections will in general not be equal to the eigenfunctions we would find if a one-dimensional particle was propagated under the influence of the energy depending on that coordinate only, since the internal coordinates are not independent of each other (see Figure IV.3). But still, we have the hope that the projections somehow resemble these functions. This is the central idea of the following calculations: If the internal coordinates are not too heavily dependent on one another, we can try to apply the variational method with local functions, dependent on just one internal coordinate, and placed near the minimum positions of the individual potential energies. Choosing these functions does not require the use of clustering techniques, but can be based on knowledge of the molecular energy function, which is known as an expression of the internal coordinates. Since the number of internal coordinates grows at most like $\mathcal{O}(N^2)$, the same would be true for the number of basis functions needed. As before, we will stick to Gaussian functions in the square root weighted space throughout the following examples, and we will only use functions which depend on one of the dihedral coordinates.

Quantity	Symbol	Unit	Value
Boltzmann constant	k_B	$\mathrm{kJ/mol \cdot K}$	$8.3144621 \cdot 10^{-3}$
Avogadro constant	N_A	mol^{-1}	$6.02214129 \cdot 10^{23}$
Electric constant	ϵ_0	Fm^{-1}	$8.854187817 \cdot 10^{-12}$
Elementary charge	e	C	$1.602176565 \cdot 10^{-19}$
Atomic mass unit	u	kg	$1.660538912 \cdot 10^{-27}$
Mass of one carbon atom	m	kg	$12.001u$
Reference distance between atoms	d_b	nm	0.153
Reference bond angle	d_a	deg	111.00
Temperature	T	K	400 (A,B) or 600 (C)
Friction	γ	ps^{-1}	200
Bond force constant	k_b	$\mathrm{kJ/mol \cdot nm}^2$	2×10^5
Angle force constant	k_a	$\mathrm{kJ/mol \cdot rad}^2$	2×10^3
Dihedral force constant	k_d	$\mathrm{kJ/mol}$	5.92

Table IV.1.: Overview of constants and parameters used in the simulation.

IV.2. Simulation

In order to produce simulation trajectories of the above system, we implement a small Matlab-based molecular dynamics program. It allows us to define all necessary parameters, such as temperature, friction, mass of the atoms, potentials to be used, simulation time step and number of time steps. After initializing all these, it propagates the system in its Euclidean coordinates, that is, for each time step, it computes the gradient of the potential in Euclidean coordinates and adds a random perturbation. After each step, it computes the internal coordinates and saves them to a binary file. Optionally, one can also store the Euclidean coordinates as a binary file or as an .xyz-file to be used with an MD-viewer.

As for the settings, we try to make a realistic simulation and use physical units, in particular nm for distances, ps for time and $\mathrm{kJ/mol}$ for energies. We also try to use realistic values for the reference distances and angles, as well as for the force constants. However, we also wish for a certain type of behaviour to show up. First, we choose the bonds and the angles to be very strong, meaning that they should be distributed closely around the reference positions, which requires k_{ij} and k_{ijk} to be chosen sufficiently large for all bonds and all bond angles. We encounter a number of difficulties here. If k_{ij} is large, the Euler discretization can quickly lead to a blow-up. Consequently, we either have to make the simulation time step very small or friction large enough. Additionally, we want the dihedral angles to display a metastable behaviour, like the type of behaviour we have studied in the previous examples. This requirement limits the temperature, because a high temperature will flatten the energy landscape, and also keeps friction from being chosen too large, since this would lead to very long waiting times until a transition occurs. Furthermore, it keeps the time step from being made too small, since this would make simulations covering sufficiently many transitions very

expensive. It turns out to be quite difficult to find settings matching these requirements. In particular, we have to turn away from values used in real world examples a number of times. A complete list of the parameters we used is shown in Table IV.1.

IV.3. Application of the Roothan-Hall method

With a sufficiently long simulation trajectory at hand, we can turn to the approximation of eigenvalues and time scales. For system A, we expect one dominant slow process to show up. We start with a small basis set, consisting of no more than two Gaussian functions centred at the two minimum positions, in order to get a rough approximation of the eigenvalues. Furthermore, we apply the method to a large basis set of twenty-seven functions, most of them centred closely around the two minima, with variance equal to 0.1. Details about the basis sets used in this chapter are given in Table IV.2. As a reference, we also compute the eigenvalues and -vectors of two Markov state model discretizations. One of them just splits up the state space into two sets along the dihedral coordinate. The other one is a fine discretization involving 100 sets along the same coordinate. Results showing the estimated implicit time scales, the dominant eigenvalues, as well as the first two eigenfunctions, are depicted in Figure IV.4. We see that in both cases, the slow process can be resolved, and the implicit time scales converge as the lag time τ increases. As expected, the fine discretizations converge much faster than the coarse ones. Most importantly, we see that the large basis set Roothan-Hall approximation is as good as the fine MSM for most lag times. Also, both methods yield a very good approximation of the stationary density, compared to the statistical estimate obtained from sampling, and a convincing estimate of the second eigenfunction, displaying the characteristic sign change between the minima. This is a good result, it shows us that the variational method yields an approximation which is comparable to a fine MSM. We also learn from our attempts that the choice of basis functions is very influential to the approximation quality. If the functions are too strongly peaked, they are not sufficiently correlated and the eigenvalues go down. However, if they are too widely spread, they correlate too much and the H-matrix is just poorly estimated. As a consequence, the computation returns eigenvalues which are greater than one or have non-zero imaginary parts. But as we said in the beginning of the chapter, the shape of the energy function can be used to choose the right functions and move in the right direction.

For system B, the process will now display four metastable regions, but they will be populated with different frequencies. Whereas the minimum which is closest to atom number one will be visited much more often because of the electrostatic attraction, the others will be less populated. We therefore expect a number of dominant slow processes to appear, describing the transitions between different minima. Again, we apply the Roothan-Hall method with two different basis sets, a small one consisting of four functions, and a large one consisting of 35 functions, which are distributed closely around the four minima. We find that three dominant slow processes are present, and compare the estimates for implicit time scales and

eigenvalues with a four set and a 100 set MSM discretization. The results are shown in Figure IV.5 and Figure IV.6. They are similar to the previous example, both eigenvalues and time scales as well as the eigenfunctions are approximated comparably well by both the large set Roothan-Hall method and the fine MSM.

These examples had in common that the metastable behaviour was only dependent on one single coordinate. As a last step, we take a look at system C, where more than one coordinate is relevant for the description of slow processes. For the Roothan-Hall method, we combine the two basis sets used in the previous examples, that means we have 35 basis functions which depend on the first dihedral coordinate and 27 which cover the second. We compare the results to those stemming from a Markov state model discretization, which uses all possible combinations of 15 sets along each coordinate, i.e. 225 sets altogether. The second, third and fourth eigenvalue and time scale, estimated with both methods, are shown in Figure IV.7. We see that the time scales estimated with the Roothan-Hall method converge earlier. This is a good sign, because we are able to estimate the eigenvalues and time scales of the larger system by simply adding up the basis functions used for the smaller system, without having to use all possible combinations of these functions.

System	Potentials	Basis set
A	Three bonds, two bond angles, one dihedral with two minima.	Small basis: Two Gaussians, centred at $-\frac{\pi}{2}$, $\frac{\pi}{2}$, both with variance 0.3. To simplify the computation, we shifted the minima to those two positions.
		Large basis: 27 Gaussians, centred at -3.1, -3.0, -2.8, -2.6, -2.4, -2.2, -2.0, -1.5, -1.0, -0.8, -0.6, -0.4, -0.2, 0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.5, 2.0, 2.2, 2.4, 2.6, 2.8, 3.0, 3.1. All of them with variance 0.1.
B	Three bonds, two bond angles, one dihedral with four minima, electrostatic attraction between atoms one and four.	Small basis: Four Gaussians, centred at $-\frac{3\pi}{4}$, $-\frac{\pi}{4}$, $\frac{\pi}{4}$, $\frac{3\pi}{4}$, all of them with variance 0.1. To simplify the computation, we shifted the minima to those four positions.
		Large basis: 35 Gaussians centred at -3.0, -2.9, -2.8, -2.6, -2.2, -2.0, -1.8, -1.7, -1.6, -1.5, -1.4, -1.2, -1.0, -0.8, -0.4, -0.2, -0.1, 0, 0.1, 0.2, 0.4, 0.8, 1.0, 1.2, 1.4, 1.5, 1.6, 1.7, 1.8, 2.0, 2.2, 2.6, 2.8, 2.9, 3.0. All of them with variance 0.1.
C	Four bonds, three bond angles, two dihedrals with $n = 4$ for the first one and $n = 2$ for the second one.	The large set from B for the first dihedral and the large set from A for the second.

Table IV.2.: Description of the example systems and the basis sets used for them.

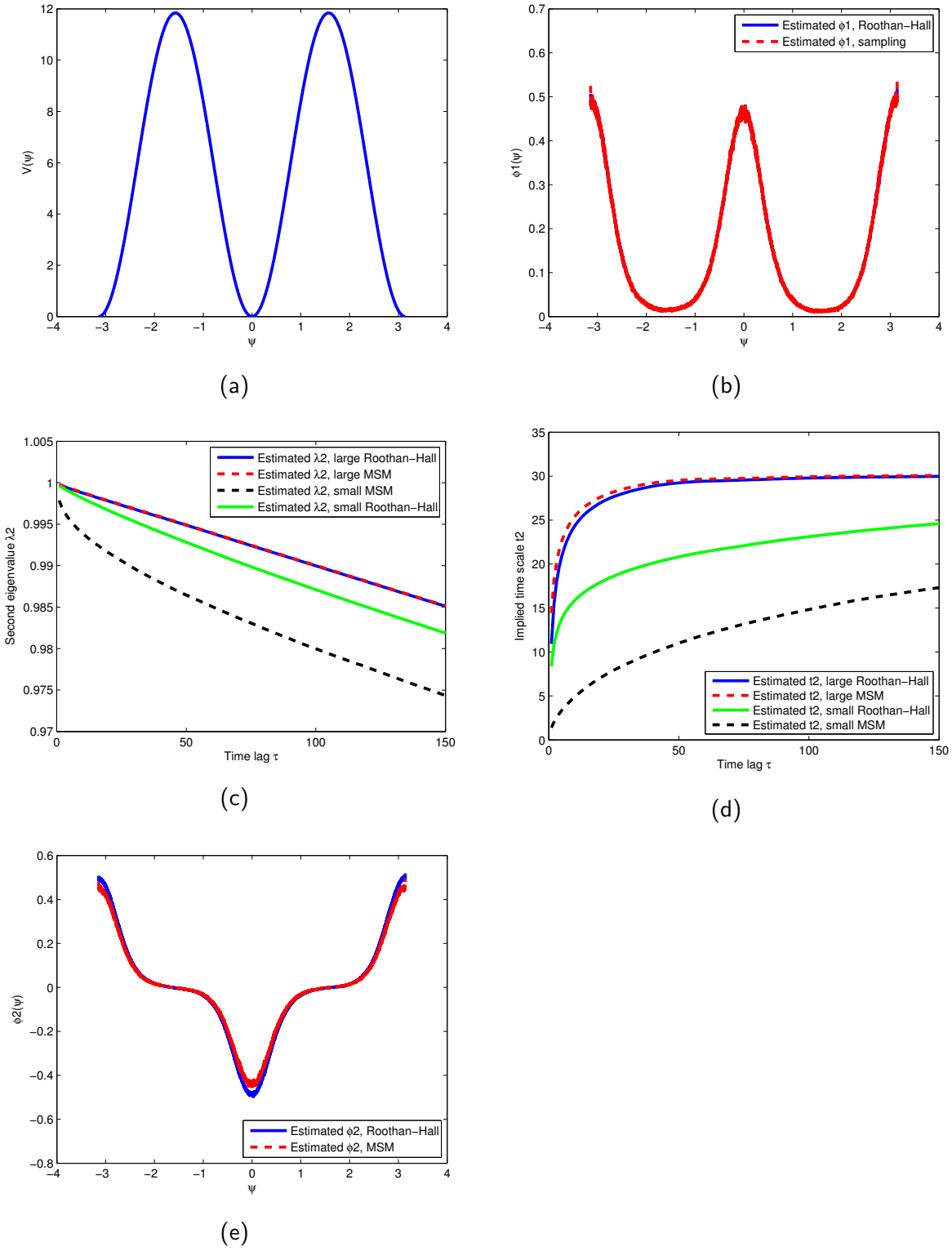


Figure IV.4.: Approximation results for system A. We used every third step of a 30 million trajectory corresponding to a sampling time step $\Delta t = 10^{-3}\mathrm{ps}$. We compare the results obtained with the small and the large basis set with a 2 set and a 100 set MSM. The functions displayed were computed from the large basis sets. (a) Potential energy for the dihedral coordinate. (b) Projection of first eigenfunction ϕ_1, compared to direct estimate from the sample. (c) Second eigenvalue λ_2. (d) Second implied time scale t_2. (e) Projection of second eigenfunction ϕ_2.

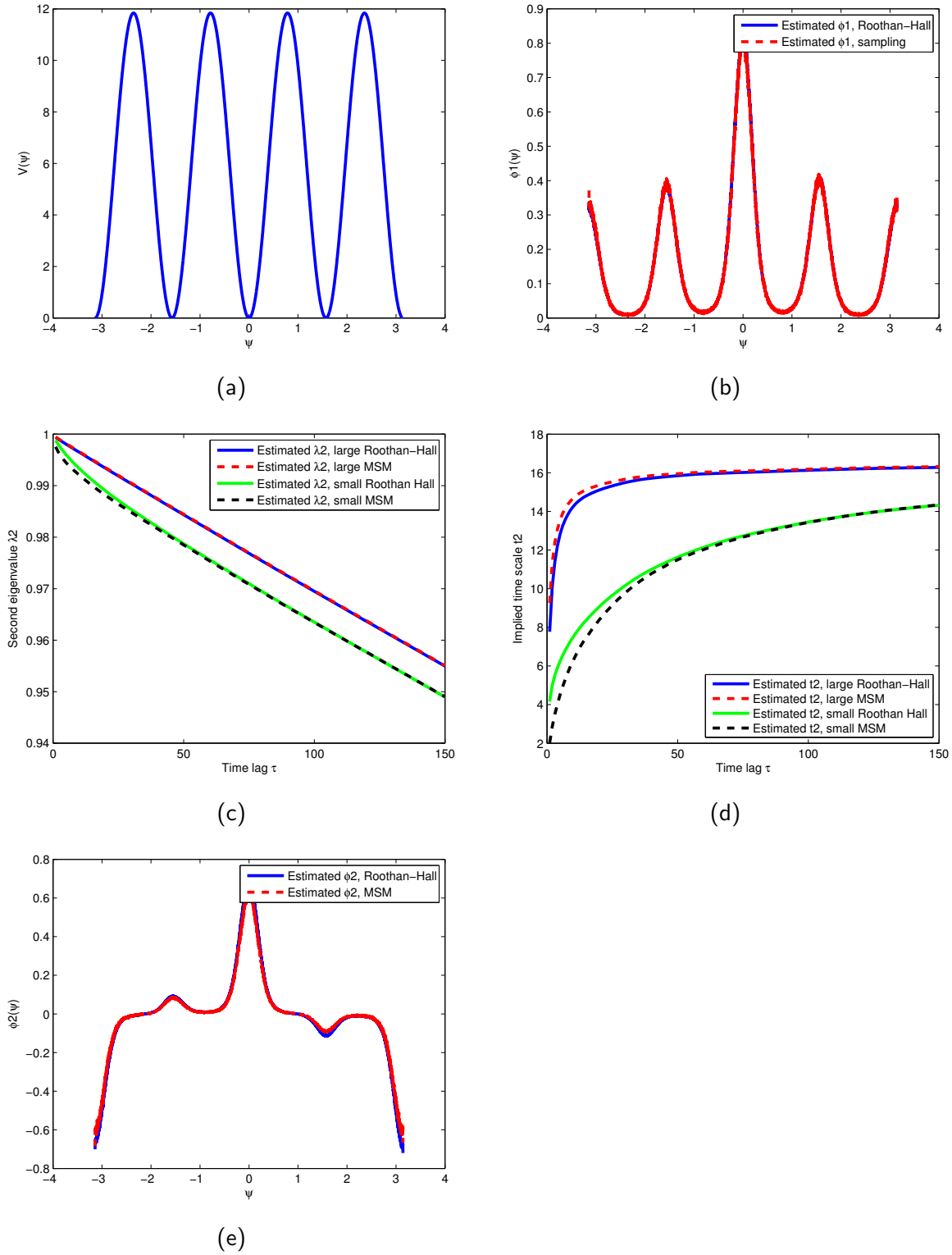


Figure IV.5.: Approximation results for system B. Every fifth step of a 30 million trajectory corresponding to a sampling time step $\Delta t = 10^{-3}\mathrm{ps}$ was used. We compare the results obtained with the small and the large basis set to a 4 set and a 100 set MSM discretization. The functions displayed were computed using the large basis sets. (a) Potential energy for the dihedral angle. (b) Projection of first eigenfunction ϕ_1, compared with direct estimate from sampling. (c) Second eigenvalue λ_2. (d) Second implied time scale t_2. (e) Projection of second eigenfunction ϕ_2.

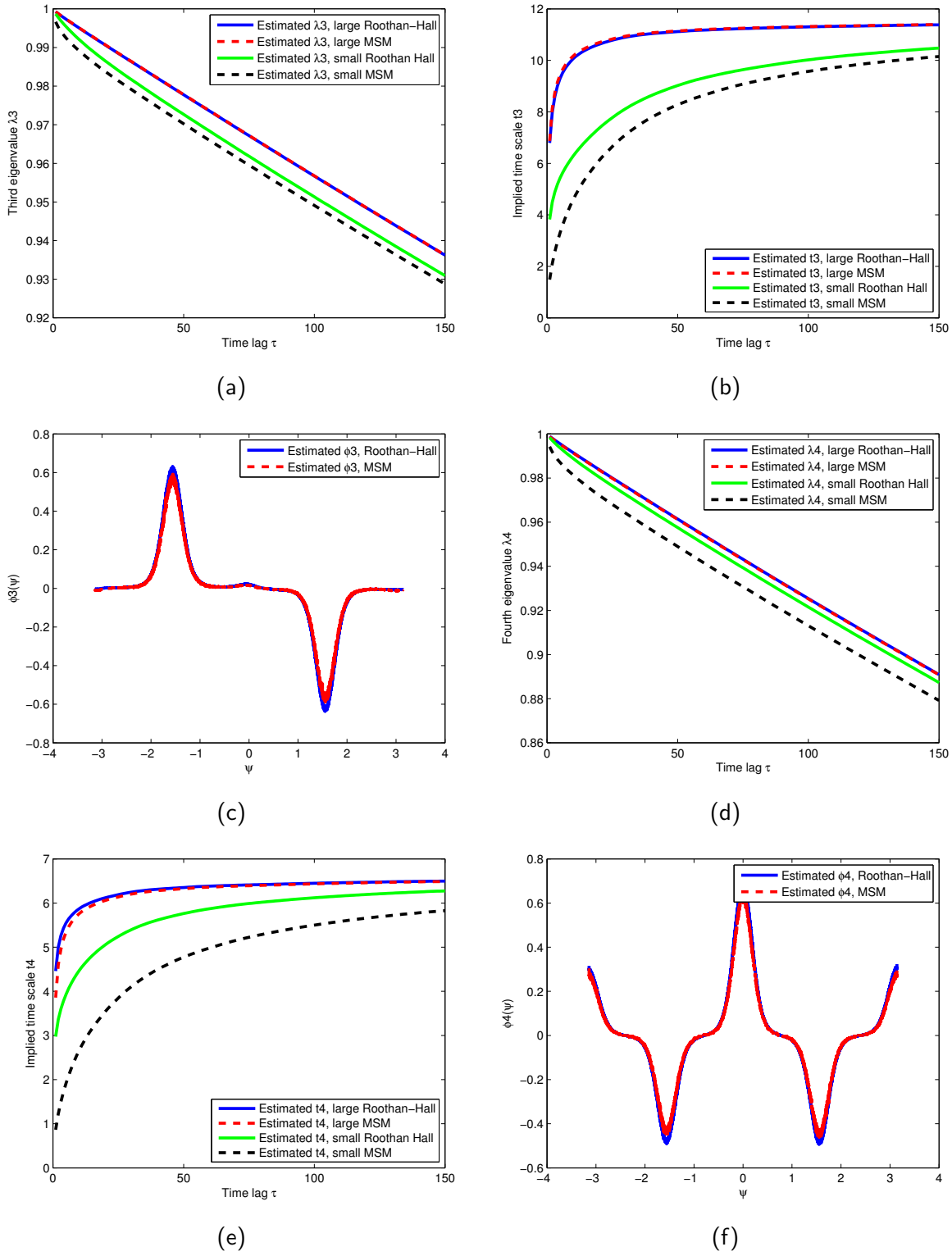


Figure IV.6.: Continuation of Figure IV.5. (a) Third eigenvalue λ_3. (b) Third implied time scale t_3. (c) Projection of third eigenfunction ϕ_3. (d) Fourth eigenvalue λ_4. (e) Fourth implied time scale t_4. (f) Projection of fourth eigenfunction ϕ_4.

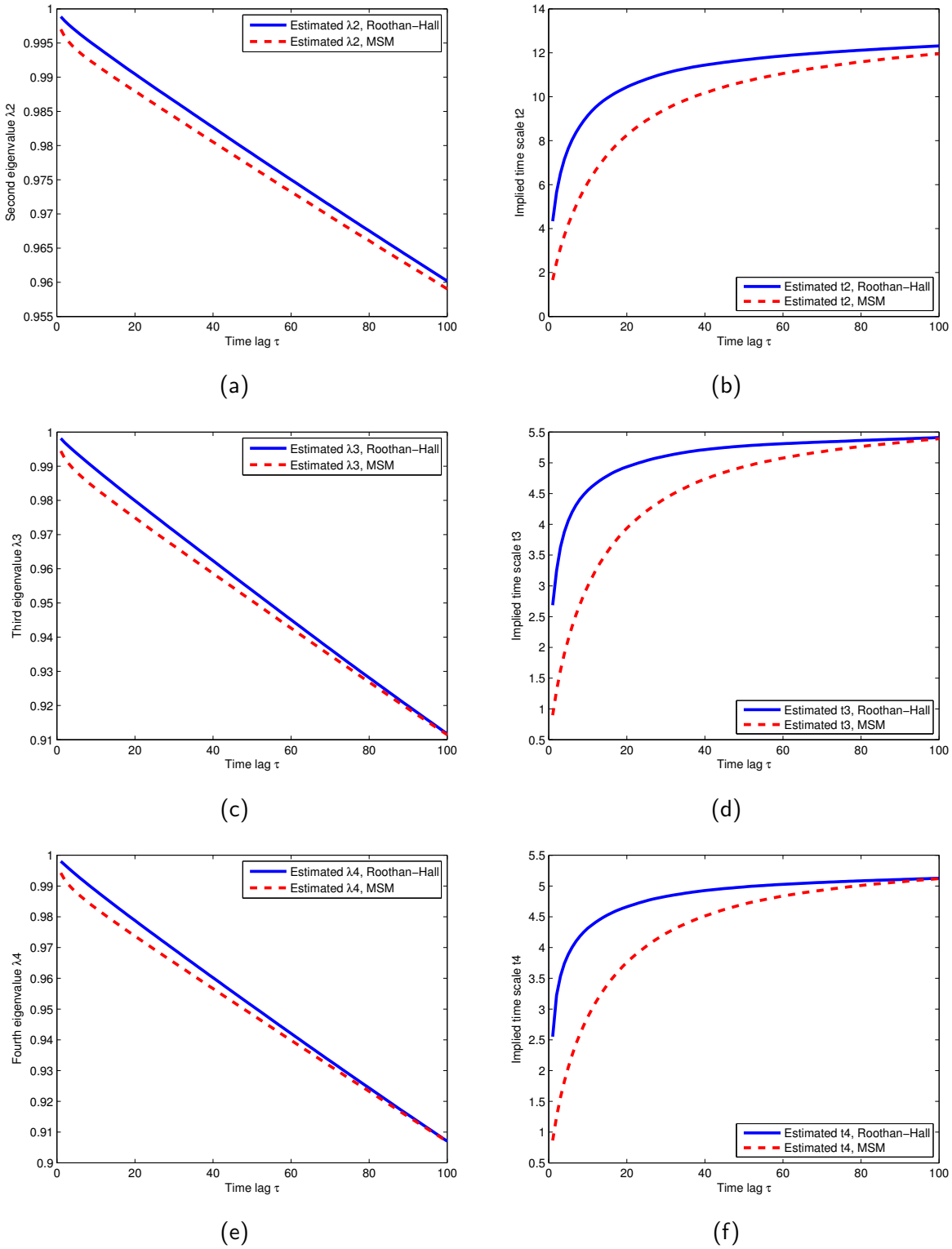


Figure IV.7.: Approximation results for system C. Every fifth step of a 10 million sampling trajectory with time step $\Delta t = 10^{-3}\mathrm{ps}$ was used. We compare the Roothan-Hall method with a combination of the large basis sets of the previous examples to a 15×15 MSM discretization. (a) Second eigenvalue λ_2. (b) Second implied time scale t_2. (c) Third eigenvalue λ_3. (d) Third implied time scale t_3. (e) Fourth eigenvalue λ_4. (f) Fourth implied time scale t_4.

V. Summary and outlook

We have presented a method to approximate the eigenvalues and time scales of slow processes in stochastic dynamical systems based on the variational principle Theorem II.6 and the Roothan-Hall method Theorem II.10. We have applied this method with Gaussian functions to one-dimensional and a few simple higher dimensional examples of diffusion processes governed by Smoluchowski dynamics. So far, we have seen that in all cases a convincing estimate of the involved time scales was achieved. Comparison with results obtained from Markov state model discretizations showed that the variational method can reproduce the quality of these.

V.1. Future work

Future work on the subject will have to deal with a number of questions:

V.1.1. Application to larger systems

Can we, in a chain with many internal coordinates, obtain a convincing result by simply adding up local basis functions, dependent on one coordinate, chosen on the basis of the molecular energy function? This assumes that the internal coordinates are sufficiently independent to allow this approximation. If it works, it would avoid the use of clustering techniques and result in a computationally affordable method to compute the dominant eigenvalues and relevant time scales. As a first test, we will have to try a five atom system with additional Coulomb interactions, and see if it works.

V.1.2. Choice of basis functions

So far, we have exclusively used Gaussian functions, and more or less guessed their shape parameters from the energy function. Can we make a more precise statement about the relation between the shape of the basis functions and the approximation quality, and what other types of functions can be used?

V.1.3. Numerical stability

We will have to address sources of numerical instabilities involved in our computations. As mentioned before, choosing basis functions which are either too uncorrelated or correlated too much can lead to meaningless results. Can we find a way to quantify these instabilities and determine how they are related to the choice of functions?

V.1.4. Importance of sampling

The method does not work without the computation of simulation trajectories. In the example we studied, it was easy to compute trajectories which visit all the metastable regions frequently often. But what does this mean for greater systems? How can we determine the minimum length of a sufficiently long trajectory, and how much does the use of finite simulations affect all the quantities involved in the method, like the H- and S-matrix?

A. Appendix

A.1. Diffusion in a quadratic potential

Let us prove the properties of the diffusion in a quadratic potential as introduced in section III.4. First, let us derive an analytic expression for the problem's solution. On stochastic integrals and Ito's formula, see [Evans, ch. 4].

Lemma A.1: *If the process is initially distributed according to X_0, then the distribution at time t is given by the stochastic integral*

$$X_t = e^{-\theta t} X_0 + \sigma \int_0^t \mathrm{dB_s}\, e^{-\theta(t-s)}. \tag{A.1}$$

—

Proof. The proof can also be found on [Wikipedia]. Let X_t be the solution of the corresponding stochastic differential equation, i.e.

$$dX_t = -\theta X_t dt + \sigma dB_t. \tag{A.2}$$

Consider the function $f(X_t, t) := e^{\theta t} X_t$. According to Ito's formula, $f(X_t, t)$ satisfies the SDE:

$$d(f(X_t,t)) = \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial x} dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}\sigma^2 dt \tag{A.3}$$

$$= \theta e^{\theta t} X_t dt - \theta e^{\theta t} X_t dt + e^{\theta t}\sigma dB_t = e^{\theta t}\sigma dB_t. \tag{A.4}$$

The process $f(X_t, t)$ is therefore given by the stochastic integral

$$e^{\theta t} X_t = f(X_t, t) = X_0 + \sigma \int_0^t \mathrm{dB_s}\, e^{\theta s}. \tag{A.5}$$

Multiplying the last expression by $e^{-\theta t}$ gives the desired result.

■

Now we can confirm the claims about the expectation and variance. The properties of stochastic integrals we use can be found again in [Evans, ch. 4].

Lemma A.2: *The solution X_t from Lemma A.1 satisfies*

$$\mathbb{E}\left[X_t\right] = e^{-\theta t}\mathbb{E}\left[X_0\right], \tag{A.6}$$

$$\omega^2\left[X_t\right] = e^{-2\theta t}\omega^2\left[X_0\right] + (1 - e^{-2\theta t})\alpha, \tag{A.7}$$

with $\alpha = \frac{k_B T}{b} = \frac{\sigma^2}{2\theta}$.

—

Proof. Directly using Equation A.1, we find

$$\mathbb{E}\left[X_t\right] = e^{-\theta t}\mathbb{E}\left[X_0\right] + \sigma\mathbb{E}\left[\int_0^t \mathrm{dB_s}\, e^{-\theta(t-s)}\right] = e^{-\theta t}\mathbb{E}\left[X_0\right], \tag{A.8}$$

since the expectation of stochastic integrals over scalar functions is always zero. Similarly, using that the expectation of the square of a stochastic integral over a scalar function equals the normal integral over the square of that function, we compute

$$\mathbb{E}\left[X_t^2\right] = e^{-2\theta t}\mathbb{E}\left[X_0^2\right] + 2\sigma e^{-\theta t}\mathbb{E}\left[X_0\right]\mathbb{E}\left[\int_0^t \mathrm{dB_s}\, e^{-\theta(t-s)}\right] + \sigma^2\mathbb{E}\left[\left(\int_0^t \mathrm{dB_s}\, e^{-\theta(t-s)}\right)^2\right] \tag{A.9}$$

$$= e^{-2\theta t}\mathbb{E}\left[X_0^2\right] + \sigma^2\mathbb{E}\left[\int_0^t \mathrm{ds}\, e^{-2\theta(t-s)}\right] \tag{A.10}$$

$$= e^{-2\theta t}\mathbb{E}\left[X_0^2\right] + \frac{\sigma^2}{2\theta}(1 - e^{-2\theta t}). \tag{A.11}$$

Therefore,

$$\omega^2\left[X_t\right] = \mathbb{E}\left[X_t^2\right] - \left(\mathbb{E}\left[X_t\right]\right)^2 = e^{-2\theta t}\mathbb{E}\left[X_0^2\right] + \alpha(1 - e^{-2\theta t}) - e^{-2\theta t}\left(\mathbb{E}\left[X_0\right]\right)^2 \tag{A.12}$$

$$= e^{-2\theta t}\omega^2\left[X_0\right] + \alpha(1 - e^{-2\theta t}). \tag{A.13}$$

■

Next, let us determine the transition kernel $p(x, y; \tau)$. If the current position of the process is known to be $x \in \Omega$, its current distribution is a delta function centred at x. Consequently, by the preceding Lemma, the distribution at time $\tau > 0$ has expectation value $e^{-\theta\tau}x$ and variance $\alpha(1 - e^{-2\theta\tau})$. We can verify that this distribution is a Gaussian function with precisely these shape parameters:

Lemma A.3: *With $\nu := (1-e^{-2\theta\tau})$, the transition function $p(x,y;\tau)$ is given by a Gaussian*

$$p(x,y;\tau) = \frac{1}{\sqrt{2\pi\alpha\nu}} \exp\left(-\frac{(y-e^{-\theta\tau}x)^2}{2\alpha\nu}\right). \tag{A.14}$$

—

Proof. We show that p satisfies the Smoluchowski Equation III.6 with initial condition $p(x,\,\cdot\,;0) = \delta_x$. Since the initial condition is clearly fulfilled, we check that the differential equation holds for all $\tau > 0$. In order to do so, we note that $\frac{\theta}{\alpha} = \frac{b^2}{m\gamma k_B T} = \frac{D}{\alpha^2}$, whereas $\frac{D}{\alpha} = \frac{b}{m\gamma} = \theta$. On the one hand we find:

$$\frac{\partial p}{\partial \tau} = -\left[\frac{2(y-e^{-\theta\tau}x)\theta e^{-\theta\tau}x2\alpha\nu - (y-e^{-\theta\tau}x)^2 4\alpha\theta e^{-2\theta\tau}}{4\alpha^2\nu^2}\right]p - \frac{1}{2}p\frac{2\pi\alpha 2\theta e^{-2\theta\tau}}{2\pi\alpha\nu} \tag{A.15}$$

$$= \frac{b^2}{m\gamma k_B T}p\left[-\frac{yxe^{-\theta\tau} - e^{-2\theta\tau}x^2 - yxe^{-3\theta\tau} + x^2e^{-4\theta\tau} - y^2e^{-2\theta\tau} + 2yxe^{-3\theta\tau} - x^2e^{-4\theta\tau}}{\nu^2}\right] \tag{A.16}$$

$$- p\frac{\theta e^{-2\theta\tau}}{\nu} \tag{A.17}$$

$$= \frac{b^2}{m\gamma k_B T}p\left[-\frac{yxe^{-\theta\tau} - e^{-2\theta\tau}x^2 + yxe^{-3\theta\tau} - y^2e^{-2\theta\tau}}{\nu^2}\right] - p\frac{\theta e^{-2\theta\tau}}{\nu}. \tag{A.18}$$

On the other hand we have:

$$\frac{1}{m\gamma}\frac{\partial}{\partial y}(byp) + D\frac{\partial^2}{\partial y^2}p = \frac{\theta p\nu}{\nu} - \frac{b^2}{m\gamma k_B T}p\frac{(y^2 - yxe^{-\theta\tau})\nu}{\nu^2} - \frac{p\theta}{\nu} \tag{A.19}$$

$$+ \frac{b^2}{m\gamma k_B T}p\frac{y^2 - 2yxe^{-\theta\tau} + x^2e^{-2\theta\tau}}{\nu^2} \tag{A.20}$$

$$= -p\frac{\theta e^{-2\theta\tau}}{\nu} + \frac{b^2}{m\gamma k_B T}p\frac{-y^2 + yxe^{-\theta\tau} + y^2e^{-2\theta\tau} - yxe^{-3\theta\tau}}{\nu} \tag{A.21}$$

$$+ \frac{b^2}{m\gamma k_B T}p\frac{y^2 - 2yxe^{-\theta\tau} + x^2e^{-2\theta\tau}}{\nu^2} \tag{A.22}$$

$$= -p\frac{\theta e^{-2\theta\tau}}{\nu} + \frac{b^2}{m\gamma k_B T}p\frac{-yxe^{-\theta\tau} + y^2e^{-2\theta\tau} - yxe^{-3\theta\tau} + x^2e^{-2\theta\tau}}{\nu^2}. \tag{A.23}$$

Since all terms cancel, the equation holds.

■

Knowing this, we can prove the statement about the eigenfunctions of the propagator:

Proof of Lemma III.4. For convenience, we restrict the proof to the case that $\alpha = 1$ and leave out all pre-factors. The case $i = 1$ was already treated when we discussed the stationary distribution, so let us check the statement for $i = 2$. By partial integration we find that:

$$\int \mathrm{dx}\, p(x,y;\tau) x \exp\left(-\frac{x^2}{2}\right) = -\int \mathrm{dx}\, p(x,y;\tau) \frac{d}{dx} \exp\left(-\frac{x^2}{2}\right) \tag{A.24}$$

$$= \int \mathrm{dx}\, e^{-\theta\tau} \frac{(y - e^{-\theta\tau}x)}{\nu} p(x,y;\tau) \exp\left(-\frac{x^2}{2}\right) \tag{A.25}$$

$$= \frac{e^{-\theta\tau}}{\nu} y \int \mathrm{dx}\, p(x,y;\tau) \exp\left(-\frac{x^2}{2}\right) \tag{A.26}$$

$$- \frac{e^{-2\theta\tau}}{\nu} \int \mathrm{dx}\, p(x,y;\tau) x \exp\left(-\frac{x^2}{2}\right). \tag{A.27}$$

Upon multiplication by ν, we see that the last term cancels, and we are left with:

$$\int \mathrm{dx}\, p(x,y;\tau) x \exp\left(-\frac{x^2}{2}\right) = e^{-\theta\tau} y \int \mathrm{dx}\, p(x,y;\tau) \exp\left(-\frac{x^2}{2}\right) = e^{-\theta\tau} y \exp\left(-\frac{y^2}{2}\right). \tag{A.28}$$

Having established the first two cases, we can prove the assertion for general i by induction. We will need the recursion relation for Hermite polynomials, which reads

$$H_{i+1}(x) = xH_i(x) - iH_{i-1}(x), \tag{A.29}$$

and also holds for the eigenfunctions ϕ_i. Furthermore, similarly to the first case, we can use that $\phi_i(x) = (-1)^i \frac{d^i}{dx^i} \exp(-\frac{x^2}{2})$, which is one way to define the Hermite functions. The course of the proof is then quite the same as for $i = 2$:

$$\int \mathrm{dx}\, p(x,y;\tau) \phi_{i+1}(x) = (-1)^{i+1} \int \mathrm{dx}\, p(x,y;\tau) \frac{d^{i+1}}{dx^{i+1}} \exp\left(-\frac{x^2}{2}\right) \tag{A.30}$$

$$= (-1)^i \int \mathrm{dx}\, e^{-\theta\tau} \frac{(y - e^{-\theta\tau}x)}{\nu} p(x,y;\tau) \frac{d^i}{dx^i} \exp\left(-\frac{x^2}{2}\right) \tag{A.31}$$

$$= \frac{e^{-\theta\tau}}{\nu} y \int \mathrm{dx}\, p(x,y,\tau)\phi_i(x) - \frac{e^{-2\theta\tau}}{\nu} \int \mathrm{dx}\, p(x,y;\tau) x \phi_i(x) \tag{A.32}$$

$$= \frac{e^{-\theta\tau}}{\nu} \lambda_i y \phi_i(y) - \frac{e^{-2\theta\tau}}{\nu} \int \mathrm{dx}\, p(x,y;\tau) \left[\phi_{i+1}(x) + i\phi_{i-1}(x)\right]. \tag{A.33}$$

Again, multiplying the equation by ν eliminates the terms containing ϕ_{i+1} and the factor $e^{-2\theta\tau}$. We are left with:

$$\int \mathrm{dx}\, p(x,y;\tau) \phi_{i+1}(x) = e^{-\theta\tau} \lambda_i y \phi_i(y) - i e^{-2\theta\tau} \int dx p(x,y;\tau) \phi_{i-1}(x) \tag{A.34}$$

$$= e^{-\theta\tau} \lambda_i y \phi_i(y) - i e^{-2\theta\tau} \lambda_{i-1} \phi_{i-1}(y). \tag{A.35}$$

By assumption, we have that $e^{-\theta\tau}\lambda_{i-1} = \lambda_i$. With one more application of the recursion relation, we arrive at the final result:

$$\int \mathrm{d}x\, p(x, y; \tau)\phi_{i+1}(x) = e^{-\theta\tau}\lambda_i \left[y\phi_i(y) - i\phi_{i-1}(y)\right] \tag{A.36}$$

$$= e^{-\theta\tau}\lambda_i \phi_{i+1}(y), \tag{A.37}$$

which proves both the assertions about the eigenfunctions and the corresponding eigenvalues. The pre-factors are then chosen to assure that the eigenfunctions are normalized.

■

Acknowledgements

I would like to thank all those who have supported me while I was working on this thesis. In particular, thank you to Frank Noé, for offering me to work on the subject and for all the time you spent helping me and giving me advice throughout the last year. I am also very grateful to Guillermo Pérez Hernández, who spent long hours helping me with the programming and many other problems. Thank you very much, Phillip Berndt, for all of the great conversations we had in the past years, and for answering all my questions at any time. Thank you, Tamara Wallenhauer, for all the support and encouragement you had for me, as well as for your patience with me. I would also like to say thanks to my family and many other of my friends for helping me through this time.

Bibliography

[Behrends, 2011] Ehrhard Behrends,
Markovprozesse und stochastische Differentialgleichungen,
FU Berlin, 2011,
`http://page.mi.fu-berlin.de/bhrnds/stochdg2011/`

[Evans] Lawrence C. Evans,
An Introduction to Stochastic Differential Equations,
University of California,
`http://math.berkeley.edu/~evans/SDE.course.pdf`

[Noé, 2011] Frank Noé,
A Variational Approach to Modelling Slow Processes in Stochastic Dynamical Systems,
FU Berlin, 2011,
`http://compmolbio.biocomputing-berlin.de/index.php/publications`

[Prinz et al, 2011] Jan-Hendrik Prinz, Hao Wu, Marco Sarich, Bettina Keller, Martin Senne, Martin Held, John D. Chodera, Christof Schütte and Frank Noé,
Markov models of molecular kinetics: Generation and validation,
The Journal of Chemical Physics 134, 174105, 2011

[Sarich, Noé, Schütte, 2010] Marko Sarich, Frank Noé and Christof Schütte,
On the Approximation Quality of Markov State Models,
SIAM Multiscale Model. Simul., 8 : 1154 - 1177, 2010.

[Schütte, Huisinga, Deuflhard, Fischer, 1999] Christof Schütte, Alexander Fischer, Wilhelm Huisinga and Peter Deuflhard
A Direct Approach to Conformational Dynamics based on Hybrid Monte Carlo,
J. Comput. Phys., 151 : 146 - 168, 1999

[Szabo, Ostlund, 1989] Attila Szabo and Neil S. Ostlund,
Modern Quantum Chemistry,
1st edition, revised, McGraw-Hill Publishing, New York, 1989

[VMD, 1996] W. Humphrey, A. Dalke and K. Schulten,
VMD - Visual Molecular Dynamics,
J. Molec. Graphics, 14 : 33 - 38.

[Weber, 2010] Markus Weber,

A Subspace Approach to Molecular Markov State Models via an Infinitesimal Generator,
ZIB report 09-27, 2010,
`http://www.zib.de/weber/`

[Werner, 2002] Dirk Werner,
Funktionalanalysis,
4th edition, Springer-Verlag, Berlin, 2002

[Wikipedia] *Ornstein-Uhlenbeck process*,
`http://en.wikipedia.org/wiki/Ornstein-Uhlenbeck_process`

[Zwanzig, 2001] Robert Zwanzig,
Nonequilibrium Statistical Mechanics,
1st edition, Oxford University Press, Oxford, 2001