# Reduced Stochastic Models for Complex Molecular Systems [*]

Illia Horenko[†], Evelyn Dittmer and Christof Schütte

*Institute of Mathematics II, Free University Berlin, Arnimallee 2–6, 14195 Berlin, Germany*

March 21, 2005

*Dedicated to the 60th birthday of Peter Deuflhard*

**Abstract**

We present a new numerical method for the identification of the most important metastable states of a system with complicated dynamical behavior from time series information. The approach is based on the representation of the effective dynamics of the full system by a Markov jump process between metastable states, and the dynamics within each of these metastable states by rather simple stochastic differential equations (SDEs). Its algorithmic realization exploits the concept of Hidden Markov Models (HMMs) with output behavior given by SDEs. A first complete algorithm including an explicit Euler-Murayama-based likelihood estimator has already been presented in [14]. Herein, we present a semi-implicit exponential estimator that, in contrast to the Euler-Murayama-based estimator, also allows for reliable parameter optimization for time series where the time steps between single observations are *large*. The performance of the resulting method is demonstrated for some generic examples, in detail compared to the Euler-Murayama-based estimator, and finally applied to time series originating from a 100 ns B-DNA molecular dynamics simulation.

*Keywords:* HMM, Ohrnstein-Uhlenbeck process, maximum likelihood principle,metastability

## 1 Introduction

The macroscopic dynamics of many complex systems is mainly characterized by the existence of metastable large scale structures, i.e., configurations which are

[*] Supported by the DFG research center "Mathematics for key technologies" (FZT 86) in Berlin.

[†] *Correspondence to:* I. Horenko; E-mail: horenko@math.fu-berlin.de

persistent for long periods of time. On the longest time scales the *effective* or *essential dynamics* is a kind of flipping process between these structures [8, 12], while on closer inspection it exhibits a rich temporal multiscale structure [17]. In other words, the effective or macroscopic dynamics of such systems is given by a jump process that hops between the metastable sets while the dynamics within these sets might be mixing on time scales that are smaller than the typical waiting time between the hops. In many applications the Markovian picture is an appropriate description of the dynamics since typical correlations times inside of the metastable sets are sufficiently smaller than the waiting times between hops (and thus much smaller than the timescale the effective description is intended to cover).

There are several recently proposed *set-oriented* approaches to the algorithmic identification of metastable sets of a complex system, and to the computation of the transition probabilities between them [18, 5, 3, 4]. These approaches are based on the construction of a transition matrix that describes transition probabilities between sets in the state space of the system. The identification of metastable sets then is based on analysis of this transition matrix [19, 6, 4]. For higher dimensional systems this always requires a set coarse graining of state space (a partition of state space in disjoint sets that avoids the curse of dimensionality) that has to be designed carefully since the resulting metastable sets are unions of the set from the partition.

Recently, the authors introduced an alternative concept [14] that is no longer purely set-oriented and is not based on (traditional) coarse graining concepts. The proposed approach aims at computing the *optimal representation* of the observed stochastic dynamics as a combination of (few) rather simple stochastic processes. In case of metastability the overall dynamical behavior of the observed complex system can be understood as a Markov chain switching between such (hopefully simple) stochastic processes. The associated algorithmic problem in such a case is: (1) to find an optimal decomposition of a given observation sequence into subsequences produced by simple stochastic processes, and (2) to determine the optimal parameters of the processes. In [14] these steps are realized based on the *expectation maximization* (EM) techniques combined with the *Euler-Murayama* discretization of the underlying stochastic dynamics. However, there are advantages but also serious disadvantages: On the one hand, application of the Euler-Murayama discretization allows to construct an *explicit* likelihood estimator and thus a fast and numerically efficient algorithm. On the other hand, with respect to the estimation quality of the algorithm, this also implies severe limitations on the observation stepsize (distance between the observations in the time series). In the present article we will discuss an extension of the techniques presented in [14] avoiding the Euler-Murayama discretization but still gaining an efficient algorithm, and investigate the possibilities of overcoming the stepsize limitations.

We will proceed as follows: First, we will present the general concept. Then, we will present the construction of the identification algorithm with special emphasis on the distinction between the Euler-Marayama-based algorithm and the extension presented herein (which will allow to avoid any discretization in time). Next, we will discuss the algorithmic realization of the two algorithmic approaches. In the last section, we will illustrate the application of this two techniques to suitable metastable time series, compare the two approaches in detail, and demonstrate that the algorithm suggested herein in fact allows to tackle rather coarsely spaced time series.

## 2 Metastable States with Internal Stochastic Dynamics

We suggest to approximate the effective dynamics by stochastic dynamical equations (SDEs) of the following type for the state $x \in \mathbf{R}^n$ of the system [14]:

$$\begin{aligned} dx(t) &= -D_x V^{(q(t))}(x(t)) + \sigma^{(q(t))}\, dW(t) && (1) \\ q(t) &= \text{Markov jump process with states } 1, \ldots, M, && (2) \end{aligned}$$

where $W(t)$ denotes standard Brownian motion, $D_x$ means differentiation wrt. $x$, $\bar{\sigma} = (\sigma^{(1)}, \ldots, \sigma^{(M)})$ contains noise intensities, and $\mathcal{V} = (V^{(1)}, \ldots, V^{(M)})$ interaction potentials. The jump process $q(t)$ is intended to mimic the hopping of the effective dynamics from one metastable set to another metastable set such that its hopping rates have to be related to the transition rates between the sets. The jump process thus can be represented by an $M \times M$ rate matrix $R$. The SDEs (1) then have to approximate the (more rapidly mixing) dynamics within the metastable states, and thus have to have correlation times that are significantly smaller than the typical waiting times between hops of the jump process. Concluding, the model is completely characterized by the tuple $(v, R, \mathcal{V}, \bar{\sigma})$, where $v$ denotes the initial statistical distribution of $x(0)$ .

In the following we assume that the potentials $V^{(q)}$ are of harmonic form:

$$V^{(q)}(x) = \frac{1}{2} D^{(q)} (x - \mu^{(q)})^2 + V_0^{(q)}. \tag{3}$$

This assumption simplifies the derivation of the parametrization algorithms significantly. We could allow for a larger class of potentials. For example, the entire derivation presented herein analogously goes through if the potential is a linear functional of its parameters. This, for example, is true for polynomial potentials; for this case one can even find parameter estimation procedures in the literature [20]. However, it is important to emphasize that the algorithmic concept advocated herein tries to realize the so-to-say *simplest reduced dynamical model* for metastable complex systems: Markov jump processes and stochastic diffusion governed by harmonic potentials within each metastable state. As is illustrated in [14], the potential dynamical substructure within each metastable state can hierarchically be represented in this setting such that the use of nonharmonic potentials may be obsolete.

Now, the whole stochastic dynamical process (1) is completely determined by the parameters $\Theta = (D^{(q)}, \mu^{(q)}, \bar{\sigma})$.

Consequently, we have to find a procedure that can determine the *optimal* model $\lambda = (v, R, \Theta)$ for the *given* time series resulting from long-term simulation of the complex system under consideration. The information about which and how many metastable sets are present in the time series is understood as being *hidden* within the data. Then, metastability is identified in terms of metastable states that are hidden states in the sense of Hidden Markov models (HMM), i.e., we try to assign to any state from the given time series the number of the hidden metastable state to which it belongs. The metastable states then are represented by aggregates containing those states that are assigned to the same metastable state. We will present a procedure that solves this assignment problem *and* the estimation problem for the parameters $(v, R, \Theta)$ simultaneously and iteratively. This procedure will result from the maximum likelihood principle. We herein present a semi-implicit likelihood estimator based on the *exact* solution of the SDEs. This has to be contrasted to the algorithmic strategy based on the Euler-Murayama discretization of the SDEs in (1) that has been presented in [14]. As it will be demonstrated in Section 3 below, and in contrast to the Euler-Murayama-based algorithm, the new strategy allows for reliable parameter optimization for time series with large time steps between single observations.

## 2.1 HmmSde and EM algorithm

Our problem is parameter estimation for system (1)+(2), that is, we want to fit the parameters to the *given* observation data $(O_{t_j})_{j=0,\ldots,T}$. Fitting means to identify a parameter configuration that maximizes the probability that the observed data is produced by the model specified by the parameter set. Parameter estimation will be realized by means of the *maximum likelihood principle*, i.e., the functional to be maximized will be given by a likelihood function $\mathcal{L}$ that will be constructed in the following way: For given parameters $\lambda$, the likelihood $\mathcal{L}(\lambda | O_t, q_t)$ has to be the probability of output $x(t_j) = O_{t_j}$, $j = 0, \ldots, T$, as well as of the associated sequence of hidden states $(q_t)$ (the state sequence of the Markov jump process at times $t_j$, $j = 0, \ldots, T$). Thus, in order to construct $\mathcal{L}$ appropriately, we have to know the probability of output of $x(t_j)$ under the condition of being in metastable state $q_{t_j}$ and of the past observations $O_{t_0}, \ldots, O_{t_{j-1}}$ for given parameters $\lambda$. We will see that we can determine this probability by considering the propagation of probability densities by the SDE associated with metastable state $q_{t_j}$.

**Construction of the likelihood.** The statistical model is composed of two stochastic processes, from which one is assumed to be hidden. The hidden process, i.e., the Markov jump process (2), is completely determined by the rate matrix R and initial distribution $v$. In contrast to an ordinary HMM, the observed process (1) herein does not consist of i.i.d. random variables, but of few SDEs (1).

The observed process is assumed to be a continuous process observed at discrete time points. Suppose that the observed data $(O_t)$ is given with constant time stepping $\tau$, i.e., $t_k = t_{k-1} + \tau$ for all $k = 1, \ldots, T$, we simply write $t =$

$1, ..., T$. The time-discrete transition matrix $A$ of the Markov chain is obtained from rate matrix R by

$$A = \exp(\tau R).$$

$A = (a_{ij})$ contains the transition probability between hidden states within two consecutive steps of the observations, i.e., $a_{ij}$ is the transition probability from hidden state $i$ to hidden state $j$ after time $t + \tau$ under the condition to be in $i$ at time $t$.

Therefore the likelihood function that has to be maximized over the parameters $\lambda = (v, A, \Theta)$ is just the joint probability distribution for the observation and hidden state sequences

$$\mathcal{L}(\lambda|O, q) = p(O, q|\lambda) = v(q_0)\rho(O_0|q_0) \prod_{t=1}^{T} a_{q_{t-1}q_t} \rho(O_t|q_t, O_{t-1}, \ldots, O_0), \qquad (4)$$

where $\rho(O_t|q_t, O_{t-1}, \ldots, O_0)$ denotes the probability of output $O_t$ under the condition of being in hidden state $q_t$, and past observation of $O_0, \ldots, O_{t-1}$.

**Algorithmic Realization.** The next task now will be to construct algorithms that

(1) determine the optimal parameters $(v, A, \Theta)$ by maximizing the likelihood $\mathcal{L}(\lambda|O, q)$; this is a nonlinear global optimization problem,

(2) determine the optimal sequence of hidden metastable states $(q_t)$ for given optimal parameters, and

(3) determine the number of important metastable states; up to now we also simply assumed that the number $M$ of hidden states is a priori given.

Before we go into the details of maximizing the likelihood we will shortly summarize the general framework of the algorithmic realization of steps (1)-(3); details are to be found in [14].

**Problem (1): Optimal parameters.** To solve problem (1) we will use the *expectation-maximization* (EM) algorithm. The EM algorithm is a learning algorithm: it alternately iterates two steps, the Expectation step and the Maximization step. Starting with some initial parameter set $\lambda_0$ the algorithm iteratively refines the parameter set, i.e., in step $k$ the present parameter set $\lambda_k$ is refined to $\lambda_{k+1}$. We will work out the details of the EM algorithm for the problem under investigation by following the general framework given in [2].

Since the data whose likelihood we want to maximize is not fully observable, we have to average over the hidden part. The key object of the EM algorithm is the expectation

$$Q(\lambda, \lambda_k) \;=\; \mathbf{E}\Big( \log p(O, q|\lambda) \;|\, O, \lambda_k \;\Big) \qquad (5)$$

of the complete-data likelihood $\mathcal{L}(\lambda|O,q) = p(O,q|\lambda)$ (in our case given by (4)) wrt. the hidden sequence $q$ given the observation sequence and the current parameter estimate $\lambda_k$. One step of the EM algorithm then realizes the following two steps:

- Expectation-step: This step evaluates the expectation value $Q$ based on the given parameter estimate $\lambda_k$.
- Maximization-step: This step determines the refined parameter set $\lambda_{k+1}$ by maximizing the expectation:

  $$\lambda_{k+1} = \underset{\lambda}{\operatorname{argmax}}\, Q(\lambda, \lambda_k). \tag{6}$$

  The maximization guarantees that $\mathcal{L}(\lambda_{k+1}) \geq \mathcal{L}(\lambda_k)$.

According to [2] (Chap. 4.2) the expectation value $Q$ as defined in (5) can be rewritten as

$$Q(\lambda|\lambda_k) \quad = \sum_{q=(q_t)\in S^{T+1}} p(O,q|\lambda_k) \log\left(p(O,q|\lambda)\right), \tag{7}$$

where $S$ denotes the state space of the hidden states. As we will see below this form will allow us to find very efficient maximizers. Due to Baum et al. [1] the function $Q(\cdot|\lambda_k)$ exhibits an unique maximum such that the new parameter iterate $\lambda_{k+1}$ is uniquely determined by (6).

**Problem (2): Optimal sequence of hidden states.** With the generalization of the EM algorithm as discussed below, problem (2) from page 5 can be solved by applying the standard Viterbi algorithm [22]. For given $\lambda$ and $O$ this algorithm computes the most probable hidden path $Q^* = (q_1^*, \ldots, q_T^*)$. This path is called the *Viterbi path*. For an efficient computation see [14]; for more details see [11].

**Problem (3): Optimal number of metastable states.** In the setup of HmmSde for a given observation sequence one is confronted with the task to select *in advance* the number $M$ of hidden states. There are no general solutions to this problem, and the best way to handle this problem often is a mixture of insight and preliminary analysis. However, since our goal is to identify metastable states we can proceed as suggested in [11]: Start the EM algorithm with some sufficient number of hidden states, say $M$, that should be greater than the expected number of metastable states. After termination of the EM algorithm, take the resulting transition matrix $A$ and aggregate the $M$ hidden states into $M_{\text{meta}} \leq M$ metastable states by means of an aggregation techniques that is designed to detect metastability. Herein, we use an aggregation technique called "Perron-cluster cluster analysis" (PCCA), see [5, 6]. The resulting aggregates of hidden states will then allow an interpretation of the results in terms of metastable states. Details can be found in [11, 14].

## 2.2 Fokker-Planck Equation and Likelihood optimization

For the sake of simplicity, we will present the derivation for a one-dimensional state space. As is pointed out in [14], this restriction is not necessary.

**Propagation of probability density.** Let us first assume that the jump process in the HmmSde model (1) is fixed to one state, say $q(t) = q$ for the times $t$ considered. Looking at a statistical density function $\rho(x,t)$ of an ensemble of SDE solutions (1) for different realizations of the stochastic process $W$ we get an equivalent representation of the dynamics in terms of the Fokker-Planck operator:

$$\partial_t \rho = \triangle_x V^{(q)}(x)\rho + \nabla_x V^{(q)}(x) \cdot \nabla_x \rho + \frac{1}{2}(\sigma^2)^{(q)} \triangle_x \rho, \tag{8}$$

where $(\sigma^2)^{(q)} \in \mathbf{R}^1$ denotes the variance of the white noise (for $\mathbf{R}^d$ it is the trace of a positive definite selfadjoint matrix). In the case of harmonic potentials this partial differential equation can be solved analytically whenever the initial density function can be represented as a superposition of Gaussian distributions: the solution of the Fokker-Planck equation (8) remains to be a sum of Gaussians whenever the initial probability function $\rho(\cdot, t = 0)$ is. Therefore, let us apply the variational principle (Dirac-Frenkel-MacLachlan principle [9]) to (8) restricted to functions $\rho$ of the form

$$\rho(x,t) = \nu(t) \exp\left(-(x - y(t))^T \Sigma(t)(x - y(t))\right).$$

This leads to an explicit solution on the time-interval $(t, t+\tau)$ where the hidden jump process $q(t)$ is fixed in the state $q$ [14]:

$$\begin{aligned}
y(t+\tau) &= \mu^{(q)} + \exp\left(-D^{(q)}\tau\right)(y(t) - \mu^{(q)}), \\
\Sigma(t+\tau) &= \left(D^{(q)^{-1}}(\sigma^2)^{(q)} - \exp\left(-2D^{(q)}\tau\right)\left(D^{(q)^{-1}}(\sigma^2)^{(q)} - \Sigma(t)^{-1}\right)\right)^{-1}, \\
\nu(t+\tau) &= \frac{1}{\sqrt{\pi}}\,\Sigma(t+\tau)^{1/2}, \qquad (9)
\end{aligned}$$

In case of initial states that are sums of Gaussians, each Gaussian would move independently according to (9) and we would get the solution of (8) by superposition.

However, in the case considered herein, we are interested in the probability of output $O(t_{j+1})$ in metastable state $q_{t_{j+1}}$ under the condition that the system has been in state $O_{t_j}$ at time $t_j$. For this, we can now use (9) with $y(t_j) = O_{t_j}$ and $\Sigma(t_j)^{-1} = 0$. Therefore, the output probability distribution results to be

$$\rho(O_{t_{j+1}}|q_{t_j}, O_{t_j}) = \nu(t_{j+1}) \exp\left(-(O_{t_{j+1}} - y(t_{j+1}))\Sigma(t_{j+1})(O_{t_{j+1}} - y(t_{j+1}))^T\right),$$

with

$$\begin{aligned} y(t_{j+1}) &= \mu^{(q)} + \exp\left(-D^{(q)}\tau\right)(O(t_j) - \mu^{(q)}), \\ \Sigma(t_{j+1}) &= \left(D^{(q)^{-1}}(\sigma^2)^{(q)} - \exp\left(-2D^{(q)}\tau\right) D^{(q)^{-1}}(\sigma^2)^{(q)}\right)^{-1} \\ &= \left(1 - \exp\left(-2D^{(q)}\tau\right)\right)^{-1} D^{(q)}(\sigma^2)^{(q)^{-1}}, \\ \nu(t_{j+1}) &= \frac{1}{\sqrt{\pi}}\,\Sigma(t_{j+1})^{1/2}, \end{aligned} \tag{10}$$

for metastable state $q = q_{t_{j+1}}$ and with $\tau = t_{j+1} - t_j$.

**Euler-Murayama discretization: explicit estimator.** The formula (10) for the parameters of the output distribution can be further simplified by the assumption that we only want to know about the evolution of the system within a *short* time interval $[t, t+\tau)$. We can then apply an Euler discretization resulting in

$$y(t+\tau) = O_t - D^{(q)}(O_t - \mu^{(q)})\tau \tag{11}$$

$$\Sigma(t+\tau) = \frac{1}{2\tau}\,(\sigma^2)^{(q)^{-1}} \tag{12}$$

$$\nu(t+\tau) = \frac{1}{\sqrt{\pi}}\,\Sigma^{1/2}(t+\tau), \tag{13}$$

which simplifies the following steps significantly.

Therefore for given model parameters $\lambda$ we have the following joint probability distribution for the observation and hidden state sequences:

$$\begin{aligned} p(O,q|\lambda) &= v(q_0)\rho(O_0|q_0)\prod_{t=1}^{T} a_{q_{t-1}q_t}\rho(O_t|q_t, O_{t-1}) \\ &= v(q_0)\nu^{(q_0)}(t)\exp\left(-(O_0 - y^{(q_0)}(0))\Sigma^{(q_0)}(0)(O_0 - y^{(q_0)}(0))^T\right) \\ &\quad \prod_{t=1}^{T} a_{q_{t-1}q_t}\nu^{(q_t)}(t)\exp\left(-(O_t - y^{(q_t)}(t))\Sigma^{(q_t)}(t)(O_t - y^{(q_t)}(t))^T\right). \end{aligned}$$

Due to (11) - (13) the Gaussian observation likelihood reduces to

$$\begin{aligned} &\rho(O_t|q_t, O_{t-1}, ..., O_1) = \rho(O_t|q_t, O_{t-1}) = \\ &= \frac{1}{(4\pi\tau^2)^{1/4}}\,((\sigma^2)^{(q_t)})^{-1}\,\exp\left(-(O_t - y^{(q_t)})\frac{1}{2\tau}(\sigma^{(q_t)})^{-1}(O_t - y^{(q_t)})^T\right), \end{aligned} \tag{14}$$

with

$$y^{(q_t)} = (O_{t-1} - D^{(q_t)}(O_{t-1} - \mu^{(q_t)})\tau).$$

and the parameter-tuple is $\lambda = (v, A, \mu, D, \sigma^2)$.

This now has to be inserted into formula (7) in order to get an explicit formula for the functional $Q$ that has to be maximized in each step of the EM iteration. To simplify notation we will use the notation $\lambda = (v, A, \mu, D, \sigma^2) = \lambda_k$ and $\hat{\lambda} = \lambda_{k+1}$ for the old and new parameter iterate, respectively.

In order to identify $\hat{\lambda} = (\hat{v}, \hat{A}, \hat{\mu}, \hat{D}, \hat{\sigma}^2)$ we have to find the zeros of the partial derivatives of $Q$ wrt. $\hat{v}, \hat{A}, \hat{\mu}, \hat{D}$, and $\hat{\sigma}^2$. Calculations and representation of these derivatives can be found in [14] together with *explicit* formulas for the optimal new parameter iterate $\hat{\lambda} = (\hat{v}, \hat{A}, \hat{\mu}, \hat{D}, \hat{\sigma}^2)$.

**Exponential discretization: semi-implicit estimator.** The main advantage of the Euler-based estimator is that given the observation sequence $(O_t)_{t=0,\ldots,T}$ and the corresponding sequence of the hidden Markov states $q_t$, the formulas for the new parameter iterate allow direct, explicit estimation of the SDE-parameters. The main assumption used in the derivation of these formulas was that the time interval $\tau$ between distinct observations of the stochastic process is short enough to validate the application of the Euler-Murayama discretization of SDE-dynamics. However, this assumption may be violated in cases where the observation sequence is coarsely spaced (i.e., $\tau$ is not small relative to the fastest time scales of complex system).

Avoiding any discretization of the SDE dynamics and thus considering the exact solution (9) of the Fokker-Planck equation (8), the log-likelihood function is $Q(O, q|\lambda) = \sum_{q=(q_t)\in S^{T+1}} p(O, q|\lambda_k) \log\left(p(O, q|\lambda) p(O, q|\lambda)\right)$ with $p$ defined as:

$$\begin{aligned} p(O, q|\lambda) &= \log(v(q_0)\nu^{(q_0)}(t)) - (O_0 - y^{(q_0)}(1))^T \Sigma^{(q_0)}(0)(O_0 - y^{(q_0)}(0)) \\ &\quad + \sum_{t=1}^{T} \log\left(a_{q_{t-1}q_t} \nu^{(q_t)}(t)\right) \\ &\quad - \sum_{t=1}^{T} \left((O_t - y^{(q_t)}(t))^T \Sigma^{(q_t)}(t)(O_t - y^{(q_t)}(t))\right), \end{aligned} \tag{15}$$

where

$$\begin{aligned} y^{(q)}(t+1) &= \mu^{(q)} + \exp\left(-D^{(q)}\tau\right)(O_t - \mu^{(q)}), \\ \Sigma^{(q)}(t+1) &= \left(1 - \exp\left(-2D^{(q)}\tau\right)\right)^{-1} D^{(q)} (\sigma^2)^{(q)^{-1}}, \\ \nu^{(q)}(t+1) &= \frac{1}{\sqrt{\pi}}\,\Sigma^{(q)}(t+1)^{1/2}. \end{aligned}$$

For *fixed* hidden sequence $q_t$ this can be re-written as a functional depending on the SDE-parameters $(\mu, D, \sigma^2) = (\mu^{(q)}, D^{(q)}, (\sigma^2)^{(q)})$. All of the second partial derivatives of this functional wrt. these parameters are negative functions, which implies that the likelihood function $\mathcal{Q}(O, q|\mu, D, (\sigma^2))$ is a *concave* functional and it has a unique maximum. In order to find this maximum, we have to find the solution of the following system of algebraic equations (see Appendix):

$$\mathcal{Q}_\mu = \frac{\partial}{\partial \mu} Q(O, q|\mu, D, \sigma^2) \quad = \quad 0 \tag{16}$$

$$\mathcal{Q}_D = \frac{\partial}{\partial D} Q(O, q|\mu, D, \sigma^2) \quad = \quad 0 \tag{17}$$

$$\mathcal{Q}_{\sigma^2} = \frac{\partial}{\partial (\sigma^2)} Q(O, q|\mu, D, \sigma^2) \quad = \quad 0 \tag{18}$$

Furthermore, explicit evaluation of (16) exhibits that it is possible to express the maximizer $\hat{\mu}$ of (16) as an *explicit* function of $D$. Substituting this expression into (17) and (18) one can get the maximizers $\hat{D}$ and $\hat{\sigma^2}$ just by solving the resulting nonlinear system of two equations with two unknowns.

Numerically this can be efficiently done by an application of Newton's method with starting point determined by the explicit Euler-based estimator. The whole procedure can be implemented in the form of a *predictor-corrector* scheme, i.e., the initial guess is predicted by the explicit Euler scheme and approved (corrected) by a subsequent Newton iteration. Whenever the initial guess provided by the explicit Euler-based estimator is good enough, the Newton method will require only few iterations.

**Direct comparison of estimators for a single SDE.** In order to compare the accuracy of the two estimators for the pure determination of SDE parameters we should consider a model with only *one* hidden Markov state and given SDE parameters, e.g., $\{\mu, D, \sigma^2\} = \{9, 50, 100\}$. We generate a typical realization of this process for time step $\tau = 0.005$ and total length $T = 25$ (see Fig. 1). We then take this time series and produce two coarser spaced ones by only taking each fourth point of the original one, and each tenth point, respectively. By this, we generate two additional time series with observation stepsize $\tau = 0.02$ and $\tau = 0.05$, respectively. The results of parametrization via HmmSde with $M = 1$ and our two estimators is displayed in Fig. 2. On one hand, we observe that the shorter the observation stepsize, the smaller the error of the parametrization with both estimators; here the error is the distance between the estimated parameters and the *exact* value of the model parameters used for a generation of the timeseries. On the other hand, the error of the explicit Euler-based estimator increases much faster with increasing $\tau$ than the error of the semi-implicit one (which stays relatively small). But since the result from the explicit estimator enters the evaluation of the semi-implicit one as initial value for the Newton iteration, we observe that the number of Newton iterations increases with increasing $\tau$: whereas for $\tau = 0.005$ the Newton method converges within two iterations, for $\tau = 0.04$, e.g., the Newton method needs ten iterations (see Fig. 2, right, and be aware of the steep gradients of the likelihood landscape).

**Complexity and Convergence.** How does the numerical effort of the two algorithmic realizations scale with the size of the problem, i.e., with the length of

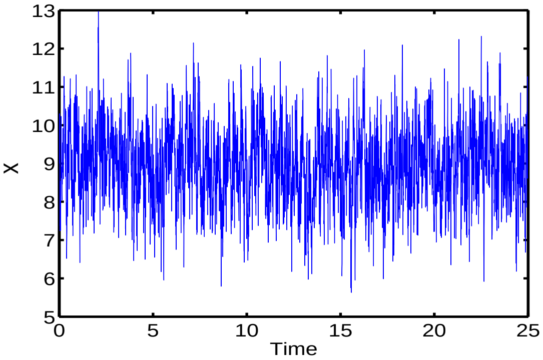

Figure 1: Realization of HmmSde model for one hidden state and SDE parameters $\{\mu, D, \sigma^2\} = \{9, 50, 100\}$ generated with time step $\tau = 0.005$.

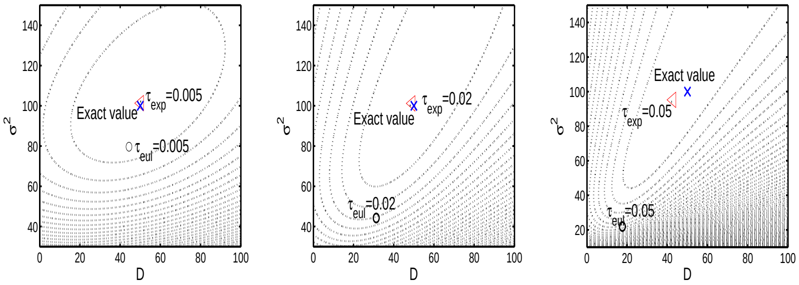

Figure 2: Exponential likelihood landscape associated with the full time series ($\tau = 0.005$) of Fig. 1 as a function of parameters $(D, \sigma^2)$ (left) and the likelihood landscapes of the two sub-sampled time series with $\tau = 0.02$ and $\tau = 0.05$ (middle and right; for details of their construction, see text). Due to steep gradients all three sub-figures show contour lines of the double logarithm $\log\log \mathcal{L}$ of the exact likelihood $\mathcal{L}$ due to (15). The marks in the three sub-figures indicate the respective estimated parameters resulting from HmmSde with the *explicit* estimator (circles) and semi-implicit estimator (triangles). The *exact* parameters used for generation of the time series are indicated by crosses.

the observation sequence $T$, and the number $M$ of hidden states? The literature on the application of EM and Viterbi algorithms to the parametrization of HMMs demonstrates that *one step* of EM and the entire Viterbi algorithm scales linearly in $T$ and quadratically in $M$; this is still true for the specific HmmSde procedure. Putting all terms together for the herein considered one-dimensional cases one finds asymptotic estimates of the following form [14]: For the explicit Euler-based estimator

$$\mathcal{O}\Big(M^2T\Big) \cdot \text{number of EM iterations.}$$

For the semi-implicit exponential estimator

$$\mathcal{O}\Big(M^2T\Big) \cdot \sum_{\#EMsteps} \text{number of Newton iterations per EM step.}$$

Here, the necessary number of iterations of the EM and Newton procedures should be determined by an accuracy requirement on the error of the underlying optimization problems. In this article convergence is controlled by the following termination criterions: (1) When the increase in likelihood in the last EM iteration does not exceed a certain preset threshold level $\text{tol}_{EM}$, the iteration is stopped. (2) When the residuum of Newton's iteration no longer is larger than a certain preset threshold level $\text{tol}_{Newton}$, the Newton iteration is terminated. In the section on numerical experiments below the two threshold values have been chosen to be identical: $\text{tol}_{EM} = \text{tol}_{Newton} = 10^{-7}$.

# 3 Numerical Experiments

In order to test and compare the likelihood estimators, we first apply them to time series generated from direct realizations of given models with known parameters of SDEs, known rate matrix, and known hidden Viterbi path.

For the first test case we use direct realizations of models of type (1) (parameters $(v, A, \mu, D, \sigma^2)$ known). For the second test case we consider a system with two perturbed metastable states, and test our two estimators for different amplitudes of perturbation. Based on the output sequence of such realizations we re-identify the parameters by application of the different HmmSde identification algorithms based on the results of the last section. The general aim is to allow to compare both approaches (Euler-based explicit estimator and semi-implicit exponential estimator). Finally, we demonstrate the application of both estimators to a 100 ns B-DNA time series.

In all numerical experiments the initial parameter guesses are based on the same procedure: The initial $M \times M$ transition matrix is chosen to be a stochastic matrix with off-diagonal entries 0.001 and identical diagonal entries. The remaining part of the parameters is obtained by the re-estimation formulas, where the probabilities $P(O_t|q_t, O_{t-1})$ are chosen uniformly distributed on $[0, 1]$, for details see [14].

Each of the examples presented in the following needed 20-70 EM iterations and 2-12 Newton steps (in case of the semi–implicit estimator).

### 3.1 Case 1: Two coupled metastable sets.

For the first test case we compute a realization of (1)+(2) with harmonic SDE potentials of the form (3) with $M = 2$ hidden states of the jump process with rate matrix

$$\mathbf{R} = \begin{pmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{pmatrix}.$$

The parameters $(\mu^{(q)}, D^{(q)}, (\sigma^2)^{(q)})$ of the two associated SDEs can be found in Table 1.

| model | First SDE | Second SDE |
|---|---|---|
| $\mu$ | 8 | 9 |
| $D$ | 50 | 50 |
| $\sigma^2$ | 100 | 100 |

Table 1: Parameters of the SDEs for the first test case.

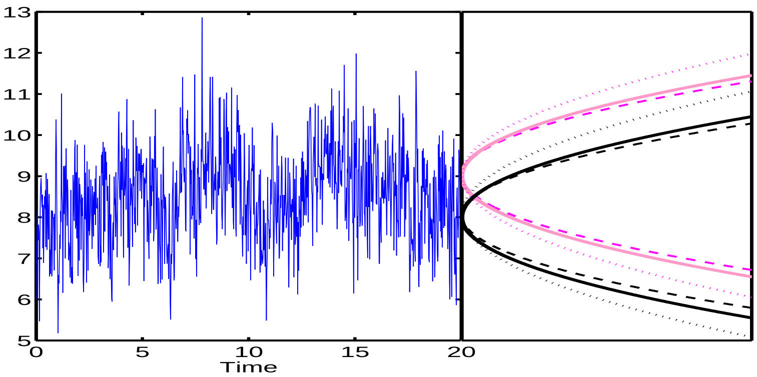

Figure 3: Left: Realization of HmmSde for the first test system with two metastable sets with parameters given in Table 1 against time. Right: Comparison of the original SDE potentials $V$ (solid) with the potentials estimated from the time series with help of the explicit Euler-Murayama likelihood estimator (dotted) and the semi-implicit exponential estimator (dashed).

Fig. 3 shows a typical realization: we observe that from the given data one cannot directly see metastability. Furthermore we observe that the SDE parameter estimation by means of the semi-implicit exponential estimator is significantly more accurate than by means of the explicit Euler-based one. Fig. 4 displays the original (hidden) path of the Markov jump process and the two paths computed via the Viterbi algorithm for the two types of estimator. Although the jumps between hidden states are not obvious from the observation sequence, the semi-implicit estimator almost perfectly identifies the hidden path while for the explicit estimator certain differences can be seen.

**Accuracy.** Figs. 5 and 6 display the accuracy of HmmSde with the two different likelihood estimators in its dependence on the observation stepsize $\tau$. We

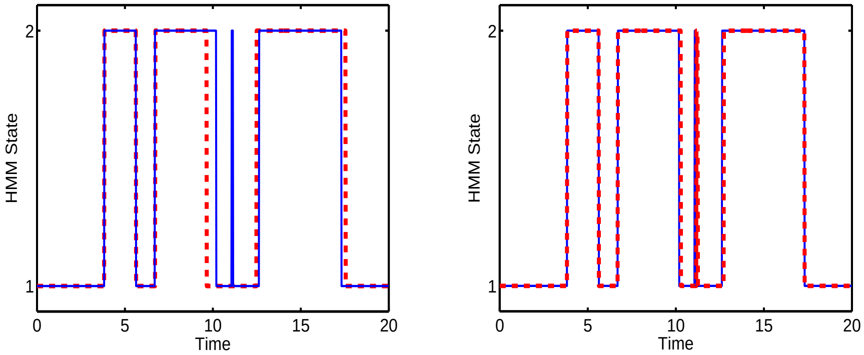

Figure 4: Comparison of the exact Viterbi path (solid) with the Viterbi path resulting from the explicit Euler-based estimator (left, dashed) and with the one resulting from the semi-implicit exponential estimator (right, dashed).

measure accuracy as follows: For each value of $\tau$ we calculate 1.000 realizations of the original dynamics (always for the same total integration time span), and compute the mean of the relative difference between the HmmSde results and the exact parameters for each of these 1.000 realizations. In Fig. 5, it can be seen clearly that, for increasing observation stepsizes $\tau$ the mean relative error of the explicit Euler-based estimator significantly increases for parameters $D$ and $\sigma^2$, while the relative error of the semi-implicit estimator shows a slight increase in $\tau$ only.

In addition to the mean error we should also be interested in the decay of variance of the error with increasing total length $T$ (number of observations) of the time series (for fixed observation stepsize $\tau$, say $\tau = 0.02$). Fig. 6 allows to observe that, for the semi-implicit estimator, the variance of the relative error for $D$ and $\sigma^2$ decreases with $T$ (almost proportional to the inverse square root of $T$). For $\mu$ there is no improvement of accuracy with longer observation sequences; this is not surpring since the parameter $\mu$ simply represents the statistical mean of the distribution of data within the respective metastable state, therefore is not directly related to any dynamical effect, and thus is already determined by few observation points.

**Comparison to standard SDE parametrization.** Since we observed above that both HmmSde estimators are able to uncover the hidden metastability in the time series of case 1, we should ask whether parametrization of a *single* SDE with a more complex, i.e., nonharmonic potential also allows to detect this metastability. We therefore take the time series of case 1 (exactly the one that we considered above with $\tau = 0.005$), and follow [20] to directly estimate the parameters of the following SDE:

$$\begin{aligned} dx &= -D_x U(x)dt + \sigma dW \\ U(x) &= a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0 \end{aligned}$$

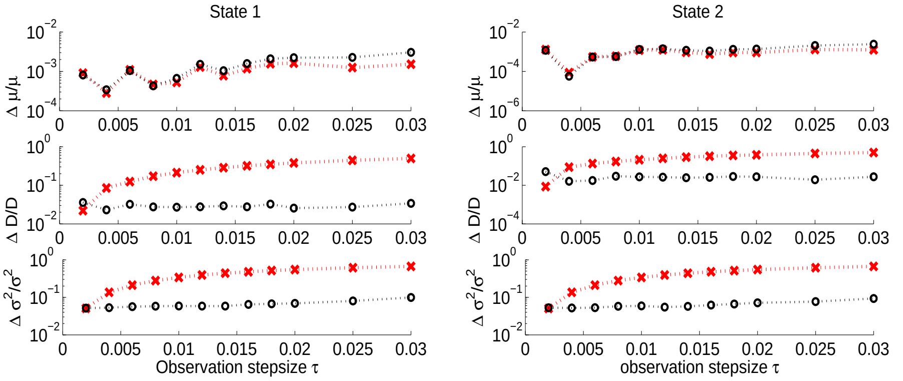

Figure 5: Mean of relative estimation errors of HmmSde results for time series of length $T = 10.000$ for the explicit Euler estimator (crosses), and semi-implicit exponential estimator (circles). The mean is computed by averaging of HmmSde results over 1.000 different realizations of the model with parameters given in Table 1

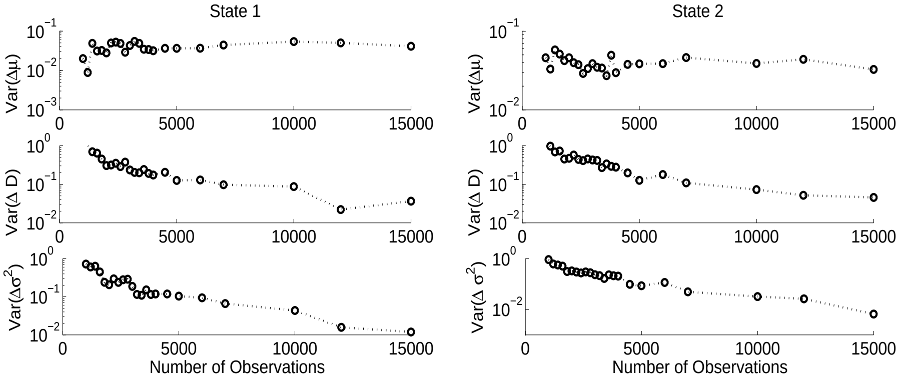

Figure 6: Variance in the relative estimation errors of HmmSde results for time series of different numbers of observations $T$ (always observation stepsize $\tau = 0.02$) for the semi-implicit exponential estimator. The variance is computed by averaging of HmmSde results over 1.000 different realizations of the model with parameters given in Table 1

The resulting potential $U$ is shown in Fig. 7: The direct estimator *fails* to uncover the hidden metastability but results in a simple quartic potential with just one well instead of a double well landscape.

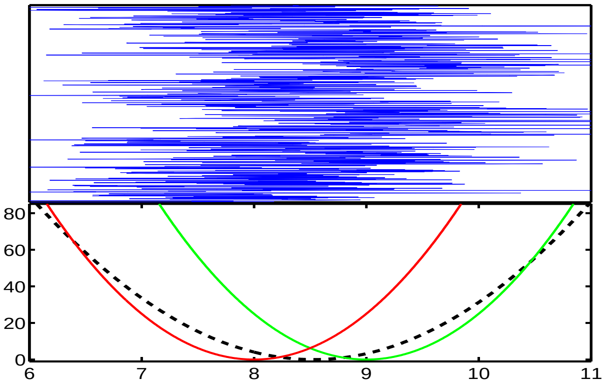

Figure 7: Results of direct parametrization of a single SDE with nonharmonic potential based on the time series (top) of case 1 as described in the text. Comparison of the exact potentials (solid, lower plot) used for the generation of the time series with the one resulting from direct parametrization (dashed, lower plot) [20].

### 3.2 Case 2: Two perturbed metastable subsets.

We now want to inspect the robustness of the detection of metastability by HmmSde wrt. perturbations. Therefore, for our second test case we again take the previous model but add perturbations to its harmonic SDE potentials in the following form:

$$V^{(q)}(x) \quad = \quad \frac{1}{2}D^{(q)}(x-\mu^{(q)})^2 + V_0^{(q)} + \epsilon^{(q)} \sin(\omega^{(q)}x). \tag{19}$$

More precisely, the time series shown in Fig. 8 results for a model of type (1)+(2) with potentials of the above form, parameters as given in Table 2, and rate matrix

$$\mathbf{R} = \begin{pmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{pmatrix}.$$

| model | First SDE | Second SDE |
|---|---|---|
| $\mu$ | 8 | 9 |
| $D$ | 50 | 50 |
| $\sigma^2$ | 100 | 100 |
| $\epsilon$ | 5 | 5 |
| $\omega$ | 50 | 50 |

Table 2: Parameters of the SDEs for test case 2.

Figs. 8 and 9 show that, although the model is perturbed, our two estimator both succeed to detect the hidden metastability correctly, i.e., they both yield reliable parameter estimations and Viterbi paths even for relatively short observation sequences (see the right panel of Figure 8).

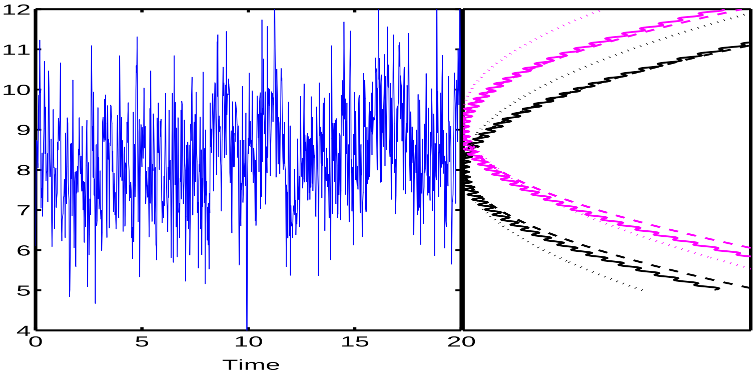

Figure 8: Left: Time series for case 2 with two perturbed metastable states with parameters given in Table 2 against time (length 1.000 with observation time step $\tau = 0.02$). Right: Comparison of the exact SDE potentials (solid) with the potentials estimated from this time series by means of the explicit Euler-Murayama based likelihood estimator (dotted) and semi-implicit exponential estimator (dashed).

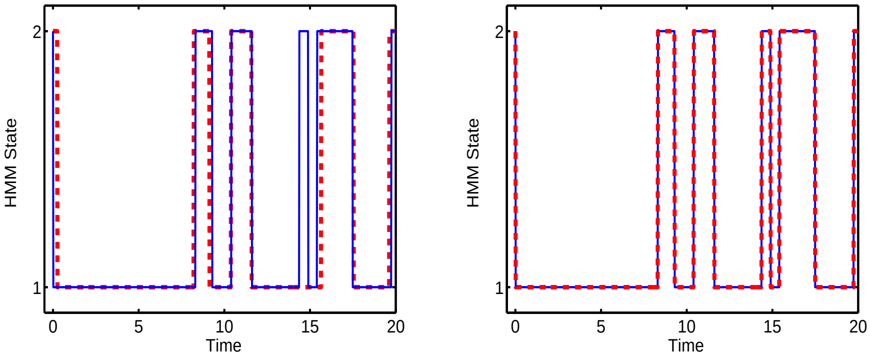

Figure 9: Comparison of the exact Viterbi path (solid) with the Viterbi path resulting from the explicit Euler-based estimator (left, dashed) and with the one resulting from the semi-implicit exponential estimator (right, dashed).

Finally, we want to compare the robustness of the two estimators for increasing perturbation, i.e., for increasing value of the perturbation amplitudes $\epsilon$ (all parameters as before, only $\epsilon^{(1)} = \epsilon^{(2)} = \epsilon$ are varied). We are mainly interested in the following aspect: HmmSde approximates the effective dynamics by fitting *harmonic* potentials. With increasing $\epsilon$, the perturbed potential will be a kind of “rugged” or “noisy” harmonic potential. When trying to find a fitting harmonic potential, then the dynamical effects, that this spatial ruggedness introduces, should be represented as additional temporal noise, i.e., we should expect to find that the optimal estimated noise intensity $\sigma_{est}$ almost coincides with the noise intensity $\sigma_{org}$ used for computing the time series for small $\epsilon$ but increases with increasing $\epsilon$.

Fig. 10 shows the results for our two estimators based on time series of equal lengths and step size but with different values of $\epsilon$. The deviation $(\Delta\mu, \Delta D, \Delta\sigma^2)$ of the estimated parameters is computed wrt. the parameters of the original model used for computation of the time series. As it can be seen, the semi-implicity exponential estimator determines the values of $D$ with much less relative error, almost independently from spatial noise intensity $\epsilon$. However, the relative error of the $\sigma^2$ estimation is increasing with increasing amplitude $\epsilon$ (as expected). In addition, the explicit estimator clearly gets significantly wrong estimates of the parameters for all values of $\epsilon$ considered.

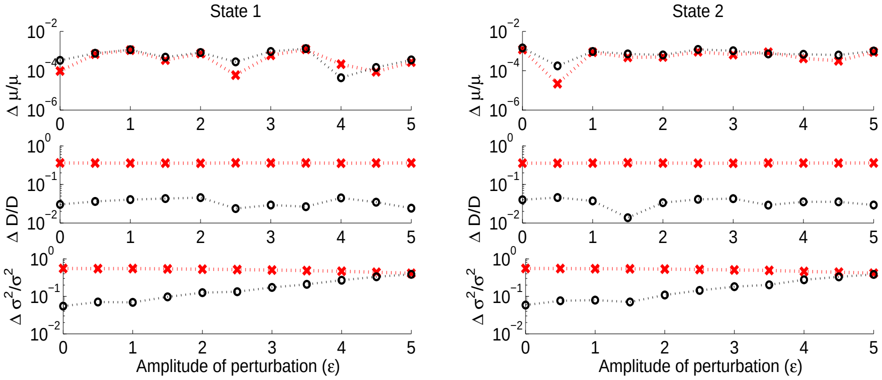

Figure 10: Mean of relative estimation errors of HmmSde results for time series of length $T = 2.000$ and observation stepsize $\tau = 0.02$ but different perturbation amplitudes $\epsilon$ for the explicit Euler estimator (crosses), and semi-implicit exponential estimator (circles). The mean is computed by averaging of HmmSde results over 1.000 different realizations of the model of case 2.

## 3.3 Case 3: Torsion dynamics of a B-DNA oligomer

Finally we consider a time series originating from a long term molecular dynamics simulation of a 15-AT B-DNA oligonucleotide, see Fig. 11. The 100 ns AMBER(parm96) force field simulation with explicit water and potassium ions was conducted in the group of J. Maddocks (EPFL), for details visit [15]. The raw data are given in the form of the backbone torsion angles time series with an observation step size $\tau = 1$ ps which is rather large compared to the fastest time scales in the underlying dynamics.

Spectral analysis of the corresponding Fisher-matrix [10, 13] (which is the analog of the covariance matrix for circular data on the torus) shows two dominant modes (see Fig. 12). Due to [15] the *essential dynamics* of the system can

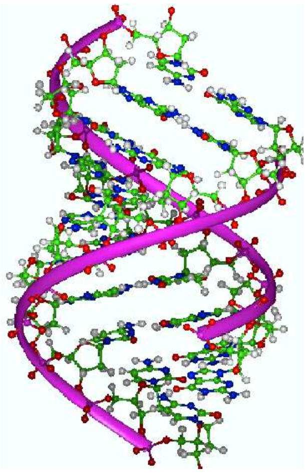

Figure 11: Illustration of the 15-AT B-DNA oligonucleotide in atomic resolution. The attached violett (grey) stings indicate the backbones. The molecular dynamics simulation referred to herein includes solvent (water and counter-ions) which is not shown here.

be investigated by projecting the overall DNA dynamics onto these dominant modes (the so-called *principal modes*).

In general, projection of the time series onto few principal modes allows for approximating the so-called *free energy landscape* of the system in terms of these modes, for details see [15, 21]. The resulting free energy landscape in terms of the first principal mode is shown in Fig. 12. Whenever the effective dynamics can be expressed correctly in terms of the principal modes then the wells in the free energy landscape can roughly be identified with metastable states of the system.

We now compare the results of HmmSde for each of our two estimators with the structure of the free energy landscape. We therefore apply the two variants of HmmSde , each with $M = 7$ hidden states, to the time series resulting from projection of the DNA dynamics onto the first principal mode. As it can be seen from Fig. 13, for both estimators the minima of the resulting harmonic potentials can be found in the vicinity of the local minima of the free energy surface. However, for the semi-implicit exponential estimator the correspondence between minima of the harmonic potentials and wells in the free energy landscape is much closer and much more reliable while the explicit estimator seems to misplace at least one of its harmonic potentials.

# Conclusion

This paper has been concerned with a novel approach to model reduction for metastable systems that has first been presented in [14]. Its conceptual core is to represent the dynamical behavior of a complex system (given via some suffi-

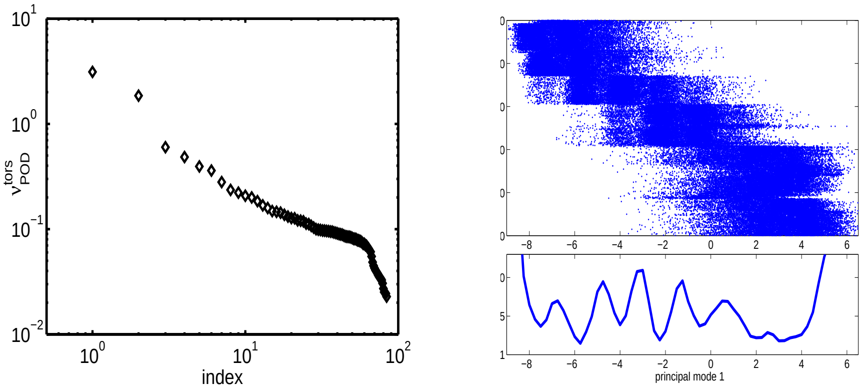

Figure 12: Left: Spectrum of the torsion angles covariance matrix (Fisher matrix) (dominant eigenvalues $\nu_1 = 3.12, \nu_2 = 1.85, \nu_3 = 0.60, \nu_4 = 0.48$). Right: projection of the overall DNA dynamics onto the first principle mode (top) and free energy landscape computed from this time series (bottom).

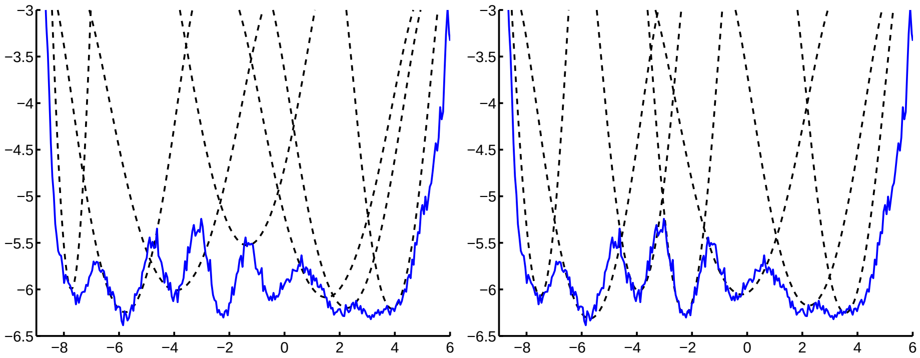

Figure 13: Comparison of the free energy landscape in terms of the first principal mode (solid) with the potentials identified by HmmSde with $M = 7$ hidden states. Left: with explicit Euler-based estimator. Right: with semi-implicit exponential estimators. Note that for the semi-implicit estimator the minima of the harmonic SDE potentials are much closer to the local minima of the free energy than for the explicit one.

ciently long time series) in terms of a Markov jump process between metastable states with internal stochastic dynamics (SDEs). Its algorithmic core is the optimal estimation of the parameters of the jump process and of the internal SDEs via likelihood maximization. We presented a novel algorithmic approach to this parameter optimization problem that includes (a) the derivation of a likelihood function that is not based on any discretization of the dynamics, and (b) a semi-implicit Newton-based technique for the optimization of this functional in the framework of the expectation-maximization algorithm. We presented several examples that document that this novel semi-implicit estimator is more accurate than the explicit estimator presented in [14], especially if applied to coarsely-spaced time series.

The main algorithmic problem left open in this article is the problem of extending the presented technique to multi-dimensional time-series. This generalization is possible along the lines presented herein: the likelihood functional can be constructed accordingly and the parameter optimization can also be based on the EM algorithm and Newton's method. However, there are several additional questions related to the essentially larger number of parameters involved, and the convergence of Newton's method in the therefore highly-dimensional parameter space. A detailed investigation will be presented in [16].

However, even in highly-dimensional cases, it may surprisingly be sufficient to have a reliable estimator for the one-dimensional case. As introduced in [7] and illustrated in [15], one can take HmmSde to evaluate the Viterbi paths for each single dimension of alone (successively), and then cluster the resulting Viterbi paths in order to construct a global multi-dimensional Viterbi path.

# References

[1] L. E. Baum. An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. *Inequalities*, 3:1–8, 1972.

[2] J. A. Bilmes. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and Hidden Markov Models. Technical report, International Computer Science Institute, Berkeley, 1998.

[3] M. Dellnitz and O. Junge. On the approximation of complicated dynamical behavior. *SIAM J. Num. Anal.*, 36(2):491–515, 1999.

[4] M. Dellnitz and R. Preis. Congestion and almost invariant sets in dynamical systems. In *Proceedings of Symbolic and Numerical Scientific Computation (SNSC'01)*, LNCS 2630, pages 183–209, 2003.

[5] P. Deuflhard, W. Huisinga, A. Fischer, and C. Schütte. Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains. *Lin. Alg. Appl.*, 315:39–59, 2000.

[6] P. Deuflhard and M. Weber. Robust Perron cluster analysis in conformation dynamics. ZIB-Report 03-19, Zuse Institute Berlin, 2003.

[7] E. Dittmer. Projizierte Hidden-Markov-Modelle in der Metastabilittsanalyse hochdimensionaler Zeitreihen. Diploma thesis, Freie Universität Berlin, 2004.

[8] R. Elber and M. Karplus. Multiple conformational states of proteins: A molecular dynamics analysis of Myoglobin. *Science*, 235:318–321, 1987.

[9] E. Faou and C. Lubich. A Poisson integrator for Gaussian wavepacket dynamics. *submitted to Comput. Visual. Sci.*, 2001.

[10] A. Fischer, F. Cordes, and C. Schütte. Hybrid Monte Carlo with adaptive temperature in a mixed–canonical ensemble: Efficient conformational analysis of RNA. *J. Comput. Chem.*, 19:1689–1697, 1998.

[11] A. Fischer, S. Waldhausen, and C. Schütte. Identification of biomolecular conformations from incomplete torsion angle observations by Hidden Markov Models. *submitted to J. Comput. Phys.*, 2004.

[12] H. Frauenfelder, P. J. Steinbach, and R. D. Young. Conformational relaxation in proteins. *Chem. Soc.*, 29A:145–150, 1989.

[13] W. Huisinga, C. Best, R. Roitzsch, C. Schütte, and F. Cordes. From simulation data to conformational ensembles: Structure and dynamic based methods. *J. Comp. Chem.*, 20(16):1760–1774, 1999.

[14] I. Horenko, E. Dittmer, A. Fischer, and C. Schütte. Automated Model Reduction for Complex Systems exhibiting Metastability. submitted to MMS

[15] I. Horenko, F. Lankas, J. Maddocks, and C. Schütte. Macroscopic Dynamics of Complex Systems Exhibiting Metastability: Theory, Algorithms, and Application to B-DNA. submitted to SIADS

[16] I. Horenko, E. Dittmer, and C. Schütte. Reduced SDE models for complex systems: The problem of dimensionality. manuscript in preparation

[17] G. U. Nienhaus, J. R. Mourant, and H. Frauenfelder. Spectroscopic evidence for conformational relaxation in Myoglobin. *PNAS*, 89:2902–2906, 1992.

[18] C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard. A direct approach to conformational dynamics based on hybrid Monte Carlo. *J. Comput. Phys.*, 151:146–168, 1999.

[19] C. Schütte and W. Huisinga. Mathematical analysis and simulation of conformational dynamics of biomolecules. In P. G. Ciaret and J.-L. Lions, editors, *Handbook of Numerical Analysis*, volume Computational Chemistry. North–Holland, 2002. in preparation.

[20] V. Smelyanskiy, D. Timucin, A. Brandrivskyy, and D. Luchinsky. Model reconstruction of nonlinear dynamical systems driven by noise. *Submitted to Physical Review Letters*, 2004.

[21] Y. Mu, P. H. Nguen, and G. Stock. Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Proteins*, 58:45–52, 2004.

[22] A. J. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. on Inform. Theory*, IT-13:260–269, 1967.

# Appendix

Herein we derive the partial derivatives of the Log-Likelihood $Q$ due to (7)+(15) wrt. the parameters of SDE associated with the Markov state $i$. We get:

$$\begin{aligned}
\frac{\partial Q}{\partial \mu^{(i)}} &= \sum_{t=2}^{T} \underbrace{\sum_{q \in S^T} \mathcal{L}(O, q_1, ..., q_t = i, ..., q_T | \bar{\lambda})}_{=\alpha_t(i)\beta_t(i)} \\
&\qquad \cdot \frac{2D^{(i)}(e^{-D^{(i)}\tau} - 1)((e^{-D^{(i)}\tau} - 1)\mu^{(i)} - e^{-D^{(i)}\tau}O_{t-1} + O_t)}{(1 - e^{-D^{(i)}\tau})(\sigma^2)^{(i)}} \\
\frac{\partial Q}{\partial D^{(i)}} &= \sum_{t=2}^{T} \alpha_t(i)\beta_t(i)\big[ - (D^{(i)})^{-1} - \frac{2e^{-2D^{(i)}\tau}\tau}{1 - e^{-2D^{(i)}\tau}} \\
&+ \frac{1}{(e^{2D^{(i)}\tau} - 1)^2(\sigma^{(i)})^2} \ \ (O_{t-1} - \mu^{(i)} + e^{D^{(i)}\tau}(\mu^{(i)} - O_t)) \\
&\cdot(\mu^{(i)} + e^{2D^{(i)}\tau}(2D^{(i)}\tau - 1)(\mu^{(i)} - O_{t-1})) - O_{t-1} + e^{3D^{(i)}}(\mu^{(i)} - O_t) \\
&- e^{D^{(i)}\tau}(2D^{(i)}\tau + 1)(\mu^{(i)} - O_{t-1})) \ \big] \\
\frac{\partial Q}{\partial (\sigma^2)^{(i)}} &= \big[\sum_{t=1}^{T} \alpha_t(i)\beta_t(i)((\sigma^2)^{(i)})^{-1} \\
&- \sum_{t=2}^{T} \alpha_t(i)\beta_t(i) \frac{D^{(i)} \ \ (e^{-D^{(i)}\tau} - 1)\mu^{(i)} - e^{-D^{(i)}\tau}O_{t-1} + O_t \ \ ^{2}}{(1 - e^{-2D^{(i)}\tau})(\sigma^2)^{(i)}} \big]
\end{aligned}$$

The maximum is computed via the zeros of this derivatives. From the first equation we can express $\mu^{(i)}$ as an *explicit* function of $D^{(i)}$. Substituting $\mu^{(i)}(D^{(i)})$ into the second and third equations we get a nonlinear system of two equations with two unknowns $D^{(i)}$ and $(\sigma^2)^{(i)}$.

$$\begin{aligned}
\frac{\partial Q}{\partial D^{(i)}}|_{\mu^{(i)} = \mu^{(i)}(D^{(i)})} &= 0 \\
\frac{\partial Q}{\partial (\sigma^2)^{(i)}}|_{\mu^{(i)} = \mu^{(i)}(D^{(i)})} &= 0
\end{aligned} \tag{20}$$