From hidden diffusion processes to hidden Markov models

A. Benabdallah,[1,2] A. Löser,[1] and G. Radons[1]

[1] *Institute of Physics, Chemnitz University of Technology, 09107 Chemnitz, Germany*

[2] *Max Planck Institut für Physik komplexer Systeme, Nöthnitzer Strasse 38, 01187 Dresden,Germany*

(Dated: December 9, 2004)

Diffusion processes, continuous in time and space, often cannot be observed directly but only via functions reflecting the measurement process. Within this work we investigate the parameter estimation problem for hidden diffusion processes in a periodic potential. We propose a discrete model as an approximation of continuous partially observed diffusion processes. Using the Baum-Welch algorithm we estimate the drift and the diffusion parameters which characterize the diffusing state variable via an approximation by a discrete random walk. Numerical results confirm our analytical estimation.

PACS numbers: 05.45.Tp, 07.05.Kf

I. INTRODUCTION

Hidden Markov models (HMM), also know as probabilistic functions of Markov chains [2], are widely used in speech recognition [27, 28], biological sequences alignment [12], ion channel analysis [5, 8, 15], etc. These models are a special case of partially observed dynamical systems, where the state of a finite Markov chain should be estimated from a sequence of noisy observations.

Despite many existing variants for modeling and analyzing time series with HMM, e.g. with respect to the output or observation processes (discrete, normal, binomial, Poisson distributions, etc.) or with respect to the underlying Markov processes (e.g. different topologies of the transition graphs, transient or ergodic processes), all existing models have in common the discreteness of the states of the Markov process [13, 17, 24]. This holds for the standard models in discrete time, but also for the existing models in continuous time (transition rates replacing transition probabilities), as they are used in the analysis of ion channels [34–36].

The continuum limit, where the underlying Markov process is described by a Fokker-Planck or a Langevin equation [30], has not been treated within the corresponding framework. There are only approaches, which estimate drift and diffusion coefficients from discrete experimental data for the diffusing variable itself [31], but without giving meaning to the observation process, nor considering it explicitly. The observation process, however, is of great importance, especially for temporally and spatially continuous processes, because in practice measurements are samples discrete in space and time (for a discussion see e.g. [32]). This means that an underlying process, which is continuous in space and time, cannot in principle be observed directly.

Despite the fact that recently functions over Brownian paths received increasing interests (e.g. in the theory of economic processes), there exists at the moment no statistically founded theory for estimating the underlying Markov processes from time series comparable to the state of the art for discrete hidden Markov models. A known strategy for estimating parameters within filtering problems is to apply the EM algorithm between observations to estimate the parameters as in Dembo and Zeitouni [10] Campillo and LeGland [7]. However, this relies on an iterative convergence between the observations with no fixed upper bound on the computations required, it is only know to apply to a specific class of model, and a general consistency result for the parameter estimates has not been proven.

Diffusion in a periodic potential has many applications in physics, biology, and for nonlinear dynamical systems in general, e.g. in superionic conductors [6], in oscillatory neurons [9, 21, 26], in phase description of dynamical systems [16] and in systems of Josephson junctions [1, 22]. Especially, the latter example is an excellent candidate for physical applications of ideas of hidden diffusion processes. In the experiment direct measurements of the time dependence of the voltage in a Josephson junctions are usually impossible as the characteristic time scale of voltage variations is too short ($\sim$ picoseconds) provides a means for tunable oscillators at frequencies above 100 GHz. One usually can measure the Josephson radiation emission in some narrow frequency range, also in the chaotic region [33]. But in this case one cannot see higher harmonics which are required to fully reconstruct the voltage time evolution.

The paper proceeds as follows. First, we define hidden diffusion processes in a periodic potential, approximated by discrete models. Second from the standard Baum-Welch algorithm we derive the modified Baum-Welch algorithm. We give an explicit solution of the modified Baum-Welch algorithm in the case of homogeneous random walks. Numerical investigations of the comparison between the two algorithms are done for space homogeneous and inhomogeneous random walks.

II. MATHEMATICAL MODEL FOR HIDDEN DIFFUSION PROCESSES

Diffusion processes are usually modeled by the evolution equation of the probability density function $\mathcal{P}$ which is governed by the continuous Fokker-Planck equation, which in one dimension reads

$$\frac{\partial}{\partial t}\mathcal{P}(x,t) = -\frac{\partial}{\partial x}\left[F(x)\mathcal{P}(x,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[D(x)\mathcal{P}(x,t)\right] \ , \tag{1}$$

where F is the drift and D is the diffusion coefficient. The eq. (1) equivalently be described by a Langevin equation interpreted in Itô sense

$$\dot{x} = F(x) + \sqrt{D(x)}\,\epsilon(t) \ , \tag{2}$$

where $\epsilon(t)$ is a intrinsic Gaussian, white noise. In the latter representation hidden diffusion processes are most easily defined. The state variable $x(t)$ can usually not be observed directly but only via a measurement process, which is modeled via nonlinear function h as

$$y(t) = h(x(t), \eta(t)) \ , \tag{3}$$

where $\xi(t)$ is an observation noise. The eq. (1) or (2) together with eq. (3) define a hidden diffusion processes.

As mentioned it in the introduction our main goals can be summarize as the following

- Identify the underlying continuous diffusion process from measured signal $y(t)$ (continuously observed or discretely sampled)
- Identify the deterministic or probabilistic measurement function
- Identify the drift and diffusion coefficient from $y(t)$ from discrete model we want to tend to continuous model.

We investigate these tasks, first by starting from the discrete model which is related to so called hidden Markov model (HMM). Consider a discrete HMM with N states and parameters $(\pi, \mathbf{A}, \mathbf{B})$, where the transition matrix $\mathbf{A}$ gives the transition between states, the emission matrix $\mathbf{B}$ connects the states and the observation and the row vector π is the distribution of the initial state.

III. DIFFUSION IN A PERIODIC POTENTIAL

In this section we consider a continuous hidden diffusion processes and we approximate it by a discrete models namely by a hidden Markov models. Typically this is done in two step. First we approximate the diffusion processes by random walk for simplicity we consider only the nearest neighbors interaction. We define the random variable as

$$x_t(\omega) = \begin{cases} s_{i+1}, & \omega = \text{right} \\ s_i, & \omega = \text{center} \\ s_{i-1}, & \omega = \text{left} \ , \end{cases} \tag{4}$$

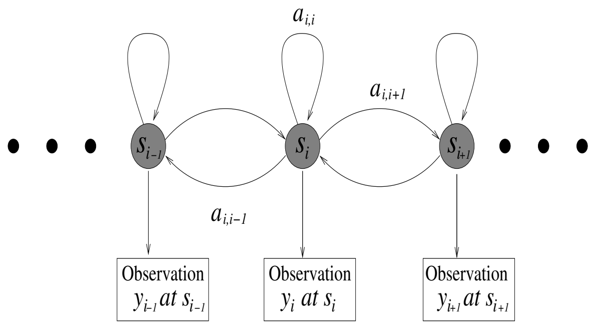

FIG. 1: Schematically approximation of hidden diffusion processes by a discrete model

so that the probability

$$\begin{aligned} P(x_{t+1} = s_{i+1}|x_t = s_i) &= a_{i,i+1}, \\ P(x_{t+1} = s_{i-1}|x_t = s_i) &= a_{i,i-1}, \\ P(x_{t+1} = s_i|x_t = s_i) = a_{i,i} &= 1 - a_{i,i+1} - a_{i,i-1} \ . \end{aligned} \tag{5}$$

The second step is to define the observation appropriately which is well defined in the continuous time limit.

Fig. 1 shows schematically the approximation of the continuous hidden diffusion processes by a discrete model. The states are represented by s_i and the observation function is given by some nonlinear function at discrete time instant t_i. Throughout this article we consider the case of a constant diffusion coefficient $D(x) = D$.

To illustrate the problems arising in the parameter estimation we consider as a first simple example of a particle subjected to a constant drift and some constant random white noise. The probability density $\mathcal{P}(x,t)$ of finding the particle at point x at some time t is described by the Fokker-Planck equation

$$\frac{\partial \mathcal{P}}{\partial t} = -F_0\frac{\partial \mathcal{P}}{\partial x} + \frac{D}{2}\frac{\partial^2 \mathcal{P}}{\partial x^2} \ . \tag{6}$$

The movement of the particle in the discrete state space $x_t \in \{s_i\}$ with the constant spacing interval $\Delta x = s_{i+1} - s_i$ can be described by a discrete version of eq. (2) If the time and the fluctuations are small, we can restrict the change in state space for one time step to nearest neighbors, $\Delta x_t \in \{\Delta x, 0, -\Delta x\}$.

The transition probabilities of eq. (5) for the change in state space define the discrete Markov model, which we use as the numerical approximation. To connect the transition probabilities to the drift and diffusion one can start from eq. (6). By a simple approximation of the derivatives in eq. (6) with difference quotients we get a tridiagonal structure for the transition matrix $\mathbf{A}$, with the probability for moving right $a_{i\,i+1} = p_0$, left $a_{i-1\,i} = q_0$ and remaining at the current position $a_{ii} = r_0$. With a constant time interval Δt for sampling, we get as approximation for the drift $F_0 = (p_0 - q_0)\Delta x/\Delta t$ and for the diffusion coefficient $D = (p_0 + q_0)(\Delta x)^2/\Delta t$.

To ensure a fixed diffusion coefficient for a resampling with a changed time interval, the sampling ratio $(\Delta x)^2/\Delta t$ has to be fixed. Because the drift should remain constant as well, the difference between the probabilities of the particle moving right respectively left has to vanish with Δx. This decreasing difference between the parameters of the Markov process with decreasing time interval of the sampling is one potential problem of the standard procedures for estimating the parameters of the Markov model.

Another problem connected to this approximation of the diffusion is that the obtained Markov model is not unique. Because $(p_0 - q_0)$ vanishes as Δx becomes small, the diffusion coefficient is defined only up to terms of this order. So in addition to the problem stated above for small Δt, there is the question how to choose the approximation for the case of coarse sampling.

This problem can readily be solved by considering the variance $var(x(t))$ of the diffusion compared to the variance $var(x_t)$ of the the Markov model. For the simple case of eq. (6) this gives a diffusion coefficient $D = \left[p_0 + q_0 - (p_0 - q_0)^2\right](\Delta x)^2/\Delta t$. For this as well as the numerical calculations we choose always a fixed initial state (s_1), which implies $p_0(i) = \delta_i$ as initial condition for eq. (6).

The second example deals with an additional harmonic modulation of the force field. This example is designed as a first step towards an harmonic decomposition of general force function. The last example is given by an anharmonic drift and demonstrates the versatility with respect to the functional form of the drift.

The drift approximation $F(x) \approx (p_i - q_i)\Delta x/\Delta t$ can be calculated by substitution of the derivatives in eq. (6). The diffusion coefficient can be fixed by considering the relative change in position for a single time step $\zeta_t = x_{t+1} - x_t$. The idea behind is that even for some poor approximation with large deviations from diffusion on large time scales, there should be at least for small time scales a good approximation of eq. (6) by the Markov process. For a defined position s_i at time t the variance is $var(\zeta) = \left[p_i + q_i - (p_i - q_i)^2\right](\Delta x)^2/\Delta t$.[18] The parameters for the teacher model can be expressed in terms of the drift and diffusion coefficient (eq. B14).

All three examples have in common some periodic drift. Therefore periodic boundary conditions are used for the discrete Markov models. This is by no means necessary, e.g. fixed boundary conditions for a particle moving in a box were tested as well with similar numerical results.

The periodic boundary condition yields to a deviation from the simple tridiagonal structure of the transition Matrix $\mathbf{A}$, e.g. for constant drift we have $a_{1,N} = a_{N-1,N} = q_0$ and $a_{N,1} = a_{N,N-1} = p_0$.

Most important for practical applications is the introduction of the measurement function in eq. (3). In an experimental setup this function may be defined by the physical measurement process and by properties of the measurement device, e.g. some smoothing of the signal.

As an example we have chosen an harmonic modulation of the states, where the observation is given by

$$y(t) = y_0 \cos 2\pi \frac{x(t)}{L} \ . \tag{7}$$

The period length is given by L. For the second example this represents the measurement of the potential.

In the discrete case we have some fixed number of possible signal observations. We assume that the period length L of the observation coincides with the period of the model. In this simplest case the number of possible observations is half the number of states N and connected to the state s_k by

$$y_t \ = \ V_k = \cos[2\pi(k - 1/2)/N] + \xi_k, \quad \text{if } x_t = s_k \ . \tag{8}$$

In general the shown approach can be applied to noisy measurements with $\xi_k \sim \mathcal{N}(0, \eta^2)$. The numerical calculations for constant drift and small noise showed no apparent difference in the quality of the approximation. This and other problems will be addressed in more detail in a subsequent paper dealing with practical problems, where the additional parameters are more meaningful.

The representation of the emission matrix $\mathbf{B}$ is given by

$$b_{j,k} = \begin{cases} \delta_{j,k} & \text{for} \quad 1 \le j \le N/2 \\ \delta_{N-j,k} & \text{for} \quad N/2 + 1 \le j \le N \end{cases} . \tag{9}$$

The potentially challenging problem becomes clear. Every two states are coded with one observation value. Even in the deterministic case without noise no easy distinction of the states is possible and therefore the dynamics is called *hidden*.

Of course by taking two successive observation into account a distinction is possible. For the probabilistic problem in general a larger sequence of observations is necessary.

From this qualitative picture we can predict the most difficult to estimate states. These are neighboring states which generate the same output symbol. Because most of the subsequent transitions generate similar output, the transition probabilities connected to these most hidden states are most difficult to estimate.

IV. STANDARD AND MODIFIED BAUM-WELCH ALGORITHMS

Before going to derive the modified Baum-Welch algorithm we give here a short overview of the evaluation of the Likelihood function and the Baum-Welch algorithm. More details can be found in the Rabiner Tutorial [27].

A. Evaluation of Likelihood function

Let's consider the HMM models $\lambda = (\pi, \mathbf{A}, \mathbf{B})$ and the observation sequence $\vec{y} = (y_1, y_2, \cdots, y_T)$ assumed to be generated by this HMM. The Likelihood function is defined by

$$L(\vec{y}) \equiv P(\vec{y}|\lambda) \ .$$

Since the model is "Markovian": i) the t^{th} hidden state, given the $(t-1)^{st}$ hidden state, is independent of previous states, and ii) the t^{th} observation, given the t^{th} hidden state, is independent of the other states then

$$L(\vec{y}) = \sum_{\vec{s}} \left[\pi_{s_1} \left(\prod_{t=1}^{T-1} a_{s_t s_{t+1}} \right) \left(\prod_{t=1}^{t=T} b_{s_t y_t} \right) \right] \ , \qquad (10)$$

where the vector $\vec{s} = (s_1, \cdots, s_N)$ denotes the hidden state sequence.
As shown in the formula (10) to evaluate $L(\vec{y})$ we can (in principle) just postmultiply a successively by $b_{s_1 y_1}$,$b_{s_2 y_2}$, etc. Because there is a numerical complication of the formula (10) which due to the large number of paths we will use so call "forward-backward algorithm". The forward-backward algorithm consists essentially of the computation of the forward probabilities $\alpha_t(i)$ and backward probabilities $\beta_t(i)$ These probabilities are defined as follows for all states i of the Markov chain, and all t from 1 to T:

$$\begin{aligned} \alpha_1(j) &= \pi_j b_{j y_1} \\ \alpha_{t+1}(j) &\equiv P(y_1, \cdots, y_{t+1}, x_t = s_j|\lambda) \\ &= b_{j y_{t+1}} \textstyle\sum_{i=1}^{N} \alpha_t(i) a_{ij} \ , \end{aligned}$$

$$\begin{aligned} \beta_T(j) &= b_{j y_T} \\ \beta_t(j) &\equiv P(y_{t+1}, \cdots, y_T, x_t = s_j|\lambda) \\ &= b_{j y_{t+1}} \textstyle\sum_{i=1}^{N} a_{ij} \beta_{t+1}(j) \ . \end{aligned} \qquad (11)$$

The Likelihood function is then given by [27]

$$L(\vec{y}) = \sum_{i=1}^{N} \alpha_t(i) \beta_t(i) \quad \text{for all } \ t \ . \qquad (12)$$

B. Standard Baum-Welch algorithm

The Baum-Welch algorithm [3, 4] is a special case of the EM algorithm [11] and consists of maximizing the following sum over all possible paths $\vec{s}$ with respect to λ:

$$Q(\lambda', \lambda) = \sum_{\vec{s}} P(\vec{s}, \vec{y} \,|\, \lambda'\}) \log P(\vec{s}, \vec{y} \,|\, \lambda) \ . \qquad (13)$$

In the following we will be interested only in the re-estimation of transition parameters. The other parameters such as the emission matrix $\mathbf{B}$ will be given by the observation (measurement) process (3) and the initial state distribution will be fixed to $\pi = (1, 0, \cdots, 0)$.
To this end let's consider the conditional probability at iteration n

$$\Psi_{ij}(\vec{y}, \{a_{kl}^{(n)}\}) = \sum_{t=1}^{t=T} \Phi_{ij}(t, \vec{y}, \{a_{kl}^{(n)}\}) \ ,$$

where $\Phi_{ij}(t, \vec{y}, \{a_{kl}^{(n)}\}) = P(x_t = s_i, x_{t+1} = s_j|\vec{y}, \{a_{kl}^{(n)}\})$. This quantity is evaluated in the appendix A.
By differentiating of the Baum's auxiliary function Q with respect to the $\{a_{ij}^{(n)}\}$ one obtains the update of the coefficient of the transition matrix explicitly as [27]

$$a_{ij}^{(n+1)} = \frac{\Psi_{ij}(\vec{y}, \{a_{kl}^{(n)}\})}{\sum_{i,j} \Psi_{ij}(\vec{y}, \{a_{kl}^{(n)}\})} = \frac{\sum_{t=1}^{T} \alpha_{t-1}(i) a_{ij}^{(n)} b_{j y_t} \beta_t(j)}{\sum_{t=1}^{T} \alpha_{t-1}(i) \beta_{t-1}(i)} \ . \qquad (14)$$

From this formula we obtain an iteration process $\{a_{kl}^{(n)}\} \to \{a_{kl}^{(n+1)}\} \to \cdots \to \{a_{kl}^{*}\}$ and one can show that [11]

$$P(\vec{y}|\{a_{kl}^{(n+1)}) \ \geq \ P(\vec{y}|\{a_{kl}^{(n)}) \ .$$

C. Modified Baum-Welch algorithm

The Baum-Welch algorithm therefore guarantees the convergence to a "local" maximum of the likelihood function (10) [37]. However the Baum-Welch algorithm can fail when the HMM has a large number of states which means that many parameters $\{a_{ij}\}$ should be re-estimated by using the formula (14). In this case we expect that the standard Baum-Welch algorithm not applicable
To avoid this problem one can parameterize the transition probabilities as $\{a_{ij}\} = \{a_{ij}(\vec{\lambda})\}$ this can be done for example by Fourier expansion in space of the element of the transition matrix. In this case the modified derivation of Maximize Baum's auxiliary function is given by

$$\begin{aligned} &Q(\{a_{ij}(\vec{\lambda}')\}, \{a_{ij}(\vec{\lambda})\}) = \\ &\textstyle\sum_{\vec{s}} P(\vec{s}, \vec{y} \,|\, \{a_{ij}(\vec{\lambda}')\}) \log P(\vec{s}, \vec{y} \,|\, \{a_{ij}(\vec{\lambda})\}) \ . \end{aligned} \qquad (15)$$

Following the work of Michalek *et al.* [25], one differentiates the modified auxiliary function (15) with respect to $\vec{\lambda} = (\lambda_1, \lambda_2, \cdots, \lambda_K)$ and obtain the modified Baum-Welch algorithm

$$0 = \sum_{i \neq j} \frac{\partial a_{ij}(\vec{\lambda}^{(n+1)})}{\partial \lambda_{\kappa}^{(n+1)}} \left(\frac{\Psi_{ij}(\vec{y}, \{a_{kl}(\vec{\lambda}^{(n)})\})}{a_{ij}(\vec{\lambda}^{(n+1)})} - \frac{\Psi_{ii}(\vec{y}, \{a_{kl}(\vec{\lambda}^{(n)})\})}{a_{ii}(\vec{\lambda}^{(n+1)})} \right) , \qquad (16)$$

for $\kappa = 1, \cdots, K$. Thus we obtain K nonlinear implicit equations to solve which in principle we can solve it numerically by rewriting the equation (16) as $\vec{\lambda}^{(n+1)} = \vec{F}(\vec{\lambda}^{(n)}, \vec{y})$ and by using the Newton methods for example one can found the fix point $\vec{\lambda}^{*}$ solution of the equation (16).

D. Example: Homogeneous random walk

As an application of the modified Baum-Welch algorithm (16) let's consider the simplest example of the homogeneous random walk with nearest neighbors interaction. Therefore the transition matrix $\mathbf{A}$ reads

$$\mathbf{A} = \left\{ \begin{array}{ccccc} r_0 & p_0 & 0 & \cdots & q_0 \\ q_0 & r_0 & p_0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & q_0 & r_0 & p_0 \\ p_0 & 0 & \cdots & q_0 & r_0 \end{array} \right\} \tag{17}$$

Thus the modified Baum-Welch algorithm (16) can be written as:

$$\begin{aligned} 0 &= \sum_i \left(\frac{\Psi_{i,i+1}(\vec{y},(p_0^{(n)},q_0^{(n)}))}{p_0^{(n+1)}} - \frac{\Psi_{i,i}(\vec{y},(p_0^{(n)},q_0^{(n)}))}{1-p_0^{(n+1)}-q_0^{(n+1)}} \right) , \\ 0 &= \sum_i \left(\frac{\Psi_{i,i-1}(\vec{y},(p_0^{(n)},q_0^{(n)}))}{q_0^{(n+1)}} - \frac{\Psi_{i,i}(\vec{y},(p_0^{(n)},q_0^{(n)}))}{1-p_0^{(n+1)}-q_0^{(n+1)}} \right) . \end{aligned} \tag{18}$$

In this case we can obtain an explicit iteration of the equation (18) by defining the abbreviations

$$\begin{aligned} \psi^+ &= \textstyle\sum_i \Psi_{i,i+1}(\vec{y},(p_0^{(n)},q_0^{(n)})), \\ \psi^- &= \textstyle\sum_i \Psi_{i,i-1}(\vec{y},(p_0^{(n)},q_0^{(n)})), \\ \psi^0 &= \textstyle\sum_i \Psi_{i,i}(\vec{y},(p_0^{(n)},q_0^{(n)})) . \end{aligned}$$

Then the solutions of the modified Baum-Welch algorithm (18) can be obtained

$$\begin{aligned} p_0^{(n+1)} &= \frac{\psi^+}{\psi^+ + \psi^- + \psi^0} \\ q_0^{(n+1)} &= \frac{\psi^-}{\psi^+ + \psi^- + \psi^0} . \end{aligned} \tag{19}$$

One can see that with the modified Baum-Welch algorithm (19), there are only 2 parameters to reestimate.

V. NUMERICAL RESULTS

In order to test numerically the modified Baum-Welch algorithm given by the expression (19). We consider a "teacher" HMM with given parameters $\lambda = (\pi, \mathbf{A}, \mathbf{B})$. This HMM is used to generate observation sequences, which correspond to hidden diffusion processes with known drift and diffusion coefficient (for this correspondence see appendix B).
Once we have the observation sequences we apply the two algorithms: the standard estimation formula (14) and the modified Baum-Welch algorithm (19). The criteria for the convergence of both algorithms is such that the condition $\log P(\vec{y}|\lambda^{(n)}) - \log P(\vec{y}|\lambda^{(n-1)}) \leq \delta$ is satisfied (typically $\delta = 10^{-6)}$). Then we will obtain $\mathbf{A}^*$ as "local" minimum of the Baum-Welch auxiliary function, an estimate of the correct values of $\mathbf{A}$ and in noisy case of observation process we will obtain $\mathbf{B}^*$ an estimate $\mathbf{B}$. Throughout our numerical simulation we define the transition matrix $\mathbf{A}$ of our 'teacher" - HMM to be tridiagonal so that the upper-diagonal, lower-diagonal and diagonal of their elements are $p(i)$, $q(i)$ and $r(i)(\equiv 1 - p(i) - q(i))$ respectively. The boundary conditions and the observation function are given in section III

A. Space homogeneous random walk

The space homogeneous random walk is the zero order approximation of periodic potential. In this case we obtained an explicit solution given by the formula (19). The fixed 'teacher"- HMM is trained with $N = 10$ (number of states) and two length of observation sequences $T = 10^4$ and $T = 10^5$.
Fig. 2 shows the parameters $p_0^{(n)}$, $q_0^{(n)}$ and $r_0^{(n)}$ as function of the numbers of iteration n with a deterministic observation process ($\xi_k = 0$ in eq. (8)) and 10 random initial parameter sets. In this figure we have used the standard Baum-Welch algorithm (14). Notice here that we have only plotted in Fig. 2 the result generated by 2 initial parameters and according to the convergence of the Baum-Welch algorithm one can check that we obtain the same results for the other 8 initial parameters set.
One can see from this figure that the convergence of standard Baum-Welch algorithm (14) is obtained for the 2 initial parameters. By increasing the length of observation sequences from $T = 10^4$ to $T = 10^5$ there is only little improvement of the convergence to the correct parameter values. As noted before the estimation of parameters is worst for neighboring states with the same observation values (best seen for $r_0^{(n)}$). Fig. 3 shows the parameters $p_0^{(n)}$, $q_0^{(n)}$ and $r_0^{(n)}$ as function of iteration numbers n with the same parameters as in Fig. 2. Here we have used the modified Baum-Welch algorithm (19). Up to permutation between p_0 and q_0, one can see from this figure the convergence of modified Baum-Welch algorithm to the correct parameter values was obtain for all the starting point. Moreover this convergence is very fast $n \approx 15$ whereas as shown in Fig. 2 the standard Baum-Welch algorithm needs a large number of iteration (around $n \approx 2500$) until the convergence is obtained. Therefore the time consuming is much important for standard Baum-Welch algorithm than for modified Baum-Welch algorithm.
The failure of standard Baum-Welch algorithm is due to the large number of reestimated parameters. At each iteration the $2N$ coefficients of the transition matrix have to be reestimated by the formula (14). For the modified Baum-Welch algorithm (19) only 2 parameters are to be determined independent of the number of states (N), which is important in case of smaller time intervals. Calculations with fixed and absorbing boundary conditions (not shown) for many short observation sequences

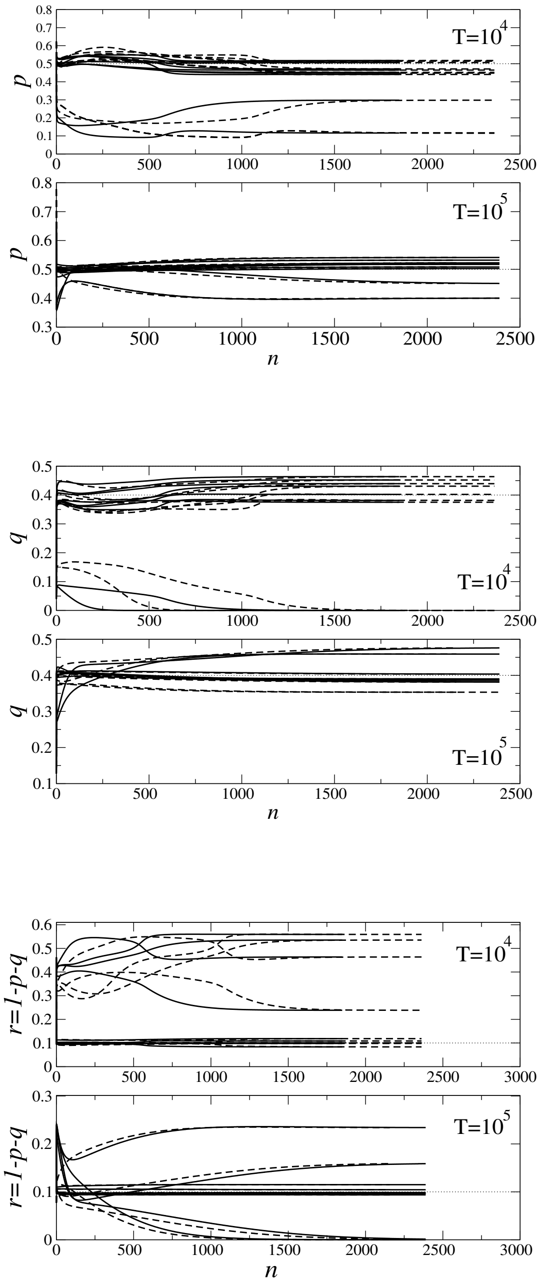

FIG. 2: The random walk parameters ($p_0^{(n)}$, $q_0^{(n)}$, $r_0^{(n)}$) vs. the number of iterations with 2 initial parameter set: $p_0^{(0)} = .7776, q_0^{(0)} = .1450$ (solid lines) and $p_0^{(0)} = .6517, q_0^{(0)} = .1280$ (dashed lines) and for two the length of observation sequence ($T = 10^4$, $T = 10^5$). The values $p(i) = p_0 = 0.5, q(i) = q_0 = 0.4$ and $r(i) = r_0 = 0.1$ for the teacher are shown by dots. The standard Baum-Welch algorithm (14) is used.

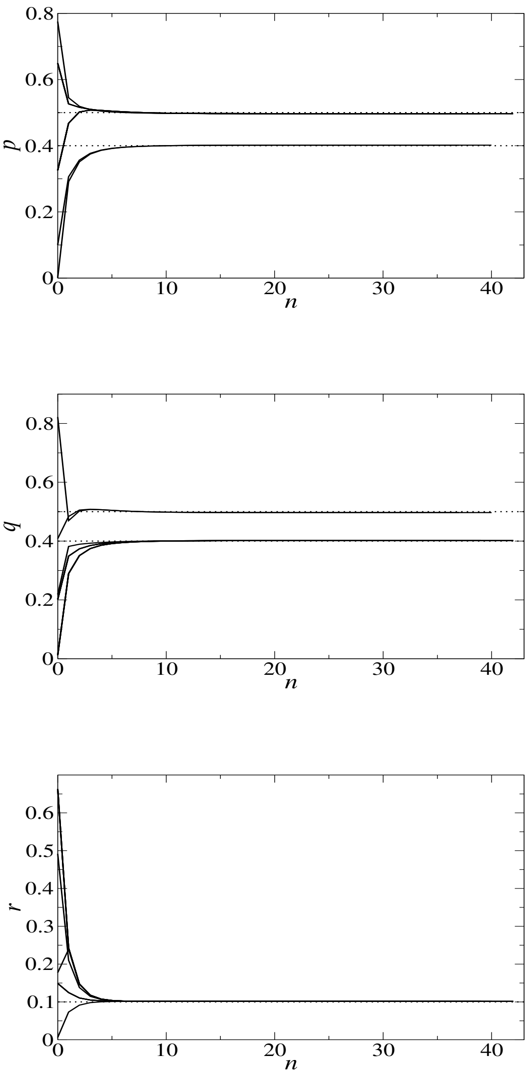

FIG. 3: The random walk parameters ($p_0^{(n)}$, $q_0^{(n)}$, $r_0^{(n)}$) vs. the number of iterations with 10 initial parameter set and $T = 10^5$ (the length of observation sequence). The values $p(i) = p_0 = 0.5, q(i) = q_0 = 0.4$ and $r(i) = r_0 = 0.1$ for the teacher are shown by dots. The modified Baum-Welch algorithm (19) is used. Convergence to the correct parameter values is observed.

have revealed no qualitative improvement of the standard Baum-Welch algorithm.

Notice in Fig. 3 that by the model there is no distinction between left and right movement possible. Therefore the convergent values p^* and q^* can be interchanged. We can see this by calculating $F = p - q$.

For a noisy observation process from the "teacher". Fig. 4 shows the deviation of the parameter r from the correct value ($\Delta r = r_0^{(n)} - 0.1$) as a function of the level observation noise η. Here the modified Baum-Welch algorithm (19) are used. One can see a small deviations from the

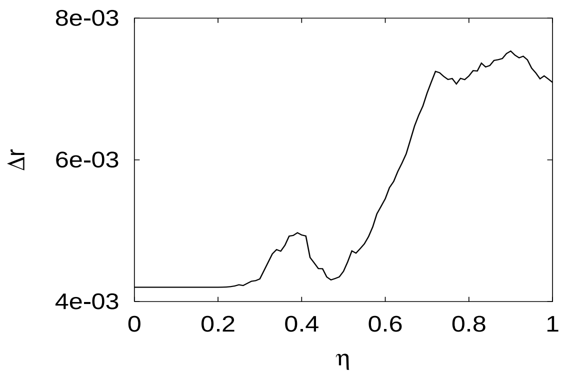

FIG. 4: Deviations ($\Delta r = r_0^{(n)} - 0.1$) from the correct values of $r = 1 - p - q$ vs. η.

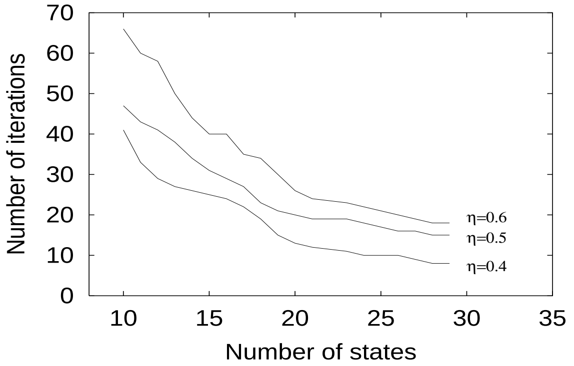

FIG. 5: The number of iterations vs. the number of states for different values of η.

correct value the algorithm of modified Baum-Welch algorithm still gives a good result. Fig. 5 shows the number of iterations as function of the number of states for the modified Baum-Welch algorithm with three parameters of level noise $\eta = 0.4, 0.5, 0.6$ (we recall η is the deviation of noisy measurement/observation processes). Here the parameters of "teacher"-HMM are the same as in Fig. 4 except for the number of states N. According that for large number of states, the discrete Hidden Markov model will converges to the continuous hidden diffusion processes. Since the coefficient of the diffusion parameter D is constant, therefore the time step must be proportional to space step see the formula (B9) given in the appendix B. Thus the length of the observation sequences T must be proportional to N^2. The numerical simulations are done as follow. Let's for $N_0 = 10$ we generate a observations sequence with length proportional to N_0^2 for example $T_0 = C\,N_0^2 (= 10^4)$ now if we change N_0 by N_1 then we will take the length of observations T_1 for N_1 is given by

$$T_1 = T_0 \left(\frac{N_1}{N_0}\right)^2 ,$$

and so on. This figure shows clearly that the decay of the number of iterations vs. the number of states for different values of η. One can deduce that for large number of states which approach the problem of the continuous hidden diffusion processes we need a small number of iterations.

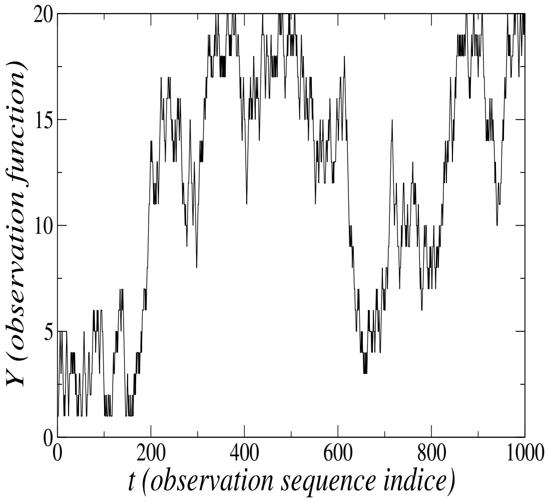

FIG. 6: Observation sequence vs. time generated from the Langevin equation where the drift $F(x) = -0.1 + 0.1\sin(\frac{2\pi x}{L})$ and the diffusion coefficient $D = 0.8$. The space length $L = 40$

B. Space inhomogeneous random walk

In this subsection we are interested in estimating the higher order Fourier components of the drift coefficient and diffusion's constant. Here we are considering a concrete problem of hidden diffusion processes where the the observation sequence is generated via the observation function given by the formula (7) and where the hidden state x is solution of the Langevin equation (2) subjected to a periodic force given by

$$F(x) = F_0 + F_1 \sin\left(\frac{2\pi x}{L}\right) , \qquad (20)$$

and a constant diffusion coefficient D. In eq. (20) L is the space length. While the explicit solution of the equation (2) is not available we have to integrate it numerically. We used the Euler scheme and then generated the observation sequence by using the formula (7).
Fig. 6 shows an example of the observation sequence as function of time. Once the observation sequence is available we have implement the modified Baum-Welch algorithm describe in the section IV.

Notice that according to the appendix B, we choose the initial data of HMM parameters as

$$\tilde{p}_i - \tilde{q}_i = \tilde{F}_0 + \tilde{F}_1 \sin\left(\tfrac{2\pi i}{N}\right) \tag{21}$$

$$(\tilde{p}_i - \tilde{q}_i) + (\tilde{p}_i - \tilde{q}_i)^2 = \tilde{D} \quad , \tag{22}$$

where $\tilde{F}_0, \tilde{F}_1$ and $\tilde{D}$ are randomly chosen. The equation of (22) can be interpreted that the diffusion coefficient $\tilde{D}$ for the initial HMM is chosen to be constant. From the equation of (21)- (22) one can have an explicit form of the initial data of HMM parameters

$$\tilde{p}_i \;=\; G_0 + G_1 \sin\left(\frac{2\pi i}{N}\right) + G_2 \sin\left(\frac{2\pi i}{N}\right)^2 \tag{23}$$

$$\tilde{q}_i \;=\; H_0 + H_1 \sin\left(\frac{2\pi i}{N}\right) + H_3 \sin\left(\frac{2\pi i}{N}\right)^2 \tag{24}$$

$$\tilde{r}_i \;=\; 1 - p_i - q_i\,, \tag{25}$$

where

$$G_0 = \frac{1}{2}(\tilde{D} + \tilde{F}_0 + \tilde{F}_0^2),\; G_1 = \frac{1}{2}(\tilde{F}_1 + 2\tilde{F}_0\tilde{F}_1),\; G_2 = \frac{\tilde{F}_1^2}{2}$$

and

$$H_0 = \frac{1}{2}(\tilde{D} - \tilde{F}_0 + \tilde{F}_0^2),\; H_1 = \frac{1}{2}(-\tilde{F}_1 + 2\tilde{F}_0\tilde{F}_1),\; H_2 = \frac{\tilde{F}_1^2}{2}$$

The boundary conditions and the observation function are the same as in the homogeneous random walk case. By using Newton's method one can calculate the numerical solution of nonlinear systems (16) with $K = 6$. Notice that at each iteration j of the modified Baum-Welch algorithm the transition matrix is given by

$$\mathbf{A}^{(\mathbf{j})} = \left\{ \begin{array}{ccccc} r_1^{(j)} & p_1^{(j)} & 0 & \cdots & q_1^{(j)} \\ q_2^{(j)} & r_2^{(j)} & p_2^{(j)} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & q_{N-1}^{(j)} & r_{N-1}^{(j)} & p_{N-1}^{(j)} \\ p_N^{(j)} & 0 & \cdots & q_N^{(j)} & r_N^{(j)} \end{array} \right\}, \tag{26}$$

where the coefficients $p_i^{(j)}$, $q_i^{(j)}$ and $r_i^{(j)}$ are determined by the relation (23-24-25). Using eq. (21-21) we calculate $F_{0,i}^{(j)}$, $F_{1,i}^{(j)}$ and $D_i^{(j)}$.

Fig. 7 shows the drift function and the diffusion coefficient as function of the states after applying the standard and modified Bum-Welch algorithm with $N = 10$ together with their correct and initial values. One can see a qualitative agreement with the correct values especially for the diffusion coefficient. By increasing the number of states from $N = 10$ to $N = 40$ we have perform this result see Fig. 8. We have done a systematic analysis by looking at the mean square error between the estimated parameters and the correct values. Thus once the observation sequence's length is fixed the mean square error for the drift function is defined as

$$E_F = \frac{1}{N}\sum_j (F(x_i) - \hat{F}_i)^2\,, \tag{27}$$

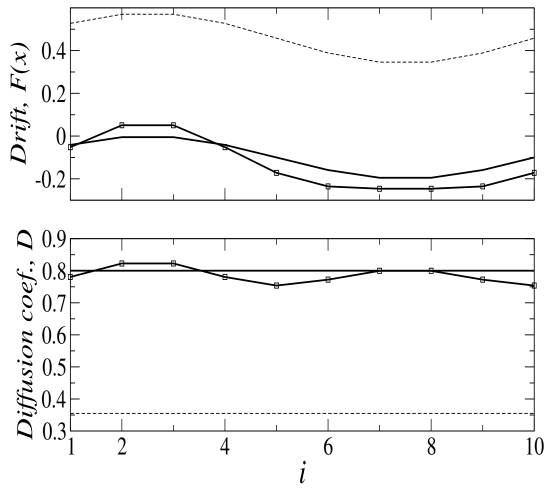

FIG. 7: The drift $F(x)$ (top panel) and the diffusion coefficient D (bottom panel) vs. the state i. The correct values of the drift $F(x) = -0.1 + 0.1\sin(\frac{2\pi x}{L})$ and the diffusion coefficient $D = 0.8$ are shown by solid lines. The estimated values given by the modified Baum Welch algorithm are shown by square and the initial data are shown shown by dashed lines. Here the observation sequence length $T = 1000$ and the space state length (number of state) $L = N = 10$.

where $F(x_i)$ is the correct drift and $\hat{F}_i$ is the estimated one using the modified Baum-Welch algorithm and in the same way as in the eq. (27) the mean square error for the diffusion E_D is computed. Fig. 9 shows the error E_F and E_D as function of the number of states for the modified Baum-Welch algorithm. One can see that the error E_F and E_D decay for the four first number of states $(10, 20, 30, 40)$ and fluctuate around 10^{-5} for the large number of states.

VI. CONCLUSIONS

We have presented analytical and numerical study of hidden diffusion processes in a periodic potential. According to the theory of the HMM we have investigated the approximation of continuous hidden diffusion processes by a discrete models. By using the Baum-Welch algorithm and it's modified version we have obtained an estimation of drift coefficient and diffusion constant.

we have compared the two algorithms for the space homogeneous random walk. It's shown from this comparison that for the modified Baum-Welch algorithm we

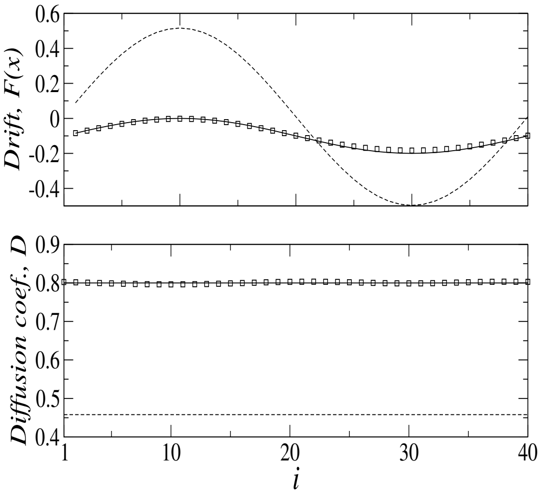

FIG. 8: The drift $F(x)$ (top panel) and the diffusion coefficient D (bottom panel) vs. the number of state. The correct values are shown by solid lines and the estimated values are shown by square and together with the initial data(dashed lines). Same parameters as in Fig. 7 expect the space state length (number of state) $L = N = 40$.

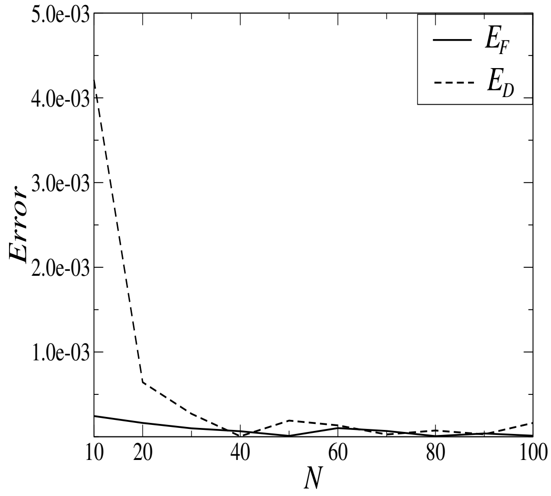

FIG. 9: The mean square error for the drift function and diffusion coefficient vs. number of states. The observation sequence length $T = 1000$

obtain a good estimation of the drift and the diffusion coefficient whereas this estimation is worst for the standard Baum-Welch algorithm.

The advantage of the modified Baum-Welch algorithm is the fixed numbers of parameters to estimate independent of the dimension of the problem compared to the standard Baum-Welch algorithm.

Moreover we have considered a concrete problem of particle subjected to a periodic potential and some constant diffusion. After applying the modified Baum-Welch algorithm we have obtain a quantitative a good estimation of drift parameters and diffusion constant.

Our conclusion is that by using the HMM approach one can obtain quantitative results for more complex problem such the Josephson junctions [33]. By solving numerically the modified Baum-Welch equations we have shown that our methods can be extended to other problems for instance a Fourier expansion is used for the drift and diffusion.

Acknowledgments

The authors thanks J. Timmer for helpful discussions. This work was supported by the DFG-Schwerpunktprogramm 1114 “Mathematical methods for time series analysis and digital image processing”.

APPENDIX A: CALCULATION OF $\Phi_{ij}(t, \vec{y}, \{a_{kl}^{(n)}\})$

Given the observation's vector $\vec{y} = (y_1, y_2, \cdots, y_T)$ the conditional probability P satisfied to

$$\begin{aligned}\Phi_{ij}(t, \vec{y}, \{a_{kl}^{(n)}\}) &= P[x_t = s_i, x_{t+1} = s_j \,|\, \vec{y}, \{a_{kl}^{(n)}\}] \\ &= \frac{P[x_t = s_i, x_{t+1} = s_j, \vec{y} \,|\, \{a_{kl}^{(n)}\}]}{P(\vec{y} \,|\, \{a_{kl}^{(n)}\})} \end{aligned} \quad \text{(A1)}$$

The denominator $P(\vec{y} \,|\, \{a_{kl}^{(n)}\})$ of the eq. (A1) can be computed using the forward-backward algorithm (12) and the nominator can be evaluated as

$$\begin{aligned} P[x_t = s_i, x_{t+1} &= s_j, \vec{y} \,|\, \{a_{kl}^{(n)}\}] = \vec{\pi} P^{(n)}(y_1) \cdots P^{(n)}(y_t) \vec{e}_i \\ &\cdot \; a_{ij}^{(n)}(y_{t+1}) \vec{e}_j P^{(n)}(y_{t+2}) P^{(n)}(y_T) \mathbf{1}' \,, \end{aligned} \quad \text{(A2)}$$

with e_i is the canonical vector and $P^{(n)}(y_t) = a_{ij}^{(n)}(y_t) = a_{ij}^{(n)} b_j(y_t)$ (a_{ij} is the transition probability from site i to j and $b_j(y_t)$ is the output probability for value y_t at site i).

APPENDIX B: RELATION BETWEEN THE DRIFT AND DIFFUSION COEFFICIENT AND THE PARAMETERS OF THE MARKOV MODEL

To establish the connection between the continuous diffusion equation 6 and the space homogenous Markov model two strategies are possible.

Because we are interested in an accurate description of the dynamics of the particle, we compare the expectation values for both processes depending on time.

Starting with eq. 6 we get by multiplying with x:

$$\int_{-\infty}^{+\infty} x\frac{\partial \mathcal{P}}{\partial t}\,dx = \frac{D}{2}\int_{-\infty}^{+\infty} x\frac{\partial^2 \mathcal{P}}{\partial x^2}\,dx - F_0\int_{-\infty}^{+\infty} x\,\frac{\partial \mathcal{P}}{\partial x}\,dx \;. \tag{B1}$$

Integrating by parts on the left-hand side, we get

$$\begin{aligned}\frac{\partial}{\partial t}\int_{-\infty}^{+\infty} x\,\mathcal{P}\,dx &= \frac{D}{2}\left[x\frac{\partial \mathcal{P}}{\partial x}\right]_{-\infty}^{+\infty} - \frac{D^2}{2}\int_{-\infty}^{+\infty}\frac{\partial \mathcal{P}}{\partial x}\,dx \\ &- F_0\left[x\mathcal{P}\right]_{-\infty}^{+\infty} + F_0\int_{-\infty}^{+\infty}\mathcal{P}\,dx \;.\end{aligned} \tag{B2}$$

For the first moment we get after some time t

$$< x(t) >= x_0 + F_0 t \tag{B3}$$

if the particle was initially at the position x_0.

In the same way, starting from the master equation for the probability $\pi_i(t) = P(x = s_i; t)$ of finding the particle at s_i,

$$\pi_i(t + \Delta t) = p_0\pi_{i-1}(t) + q_0\pi_{i+1}(t) + r_0\pi_i(t) \tag{B4}$$

the first moment for the discrete Markov model can be calculated by multiplying with x and summation over all possible states.

$$\begin{aligned}\textstyle\sum_{i=-\infty}^{\infty} s_i\pi_i(t+\Delta t) &= \textstyle\sum_{i=-\infty}^{\infty}[(s_{i-1}+\Delta x)p_0\pi_{i-1}(t) \\ &+ (s_{i+1}-\Delta x)q_0\pi_{i+1}(t) + s_i r_0\pi_i(t)]\end{aligned} \tag{B5}$$

Which gives for the first moment

$$< x_{t+1} >=< x_t > +\Delta x(p_0 - q_0) \;. \tag{B6}$$

After $t/\Delta t$ time steps starting from the initial state x_0, we get

$$< x_t >= x_0 + (p_0 - q_0)\frac{\Delta x}{\Delta t}t \;. \tag{B7}$$

Comparing eq. B3 and B7, we get

$$F_0 = (p_0 - q_0)\frac{\Delta x}{\Delta t} \;. \tag{B8}$$

One gets the same result for deriving eq. B4 directly by introducing difference quotients in eq. 6.

For the analog expression of D one has to consider the second moments $< x^2 >$. The result is

$$D = \left[(p_0 + q_0) - (p_0 - q_0)^2\right]\frac{(\Delta x)^2}{\Delta t} \;. \tag{B9}$$

The direct way starting from eq. 6 gives only the first term of eq. B9. Because $\frac{(\Delta x)^2}{\Delta t}$ should be constant, the second term in eq. B9 is of order Δt. Because we are interested in the movement of some particle it makes sense to approximate the diffusion in terms of the moments, which ensures even in the large time limit some consistency between the continuous and the discrete model.

The problem connected to the approximation becomes even more apparent in case of a non constant drift $F(x)$ but constant diffusion coefficient D. In this case one has to start with local expressions for the drift and diffusion constant

$$F(x) = \lim_{\Delta t\to 0}\frac{1}{\Delta t}\int dx'\,(x - x')\,P(x, t + \Delta t|x', t) \tag{B10}$$

$$D(x) = \lim_{\Delta t\to 0}\frac{1}{\Delta t}\int dx'\,(x - x')^2\,P(x, t + \Delta t|x', t) \tag{B11}$$

.

where P is the conditional probability of finding the particle after some time step Δt at x. These equation can be simply transfered to the time discrete case by assuming a finite Δt and considering some $\zeta_t = x_{t+\Delta t} - x_t$.[18]

One gets the expression in terms of the parameters for the drift and diffusion constant at position s_i

$$F(s_i) = \frac{p_i - q_i}{\Delta x}D_0 \tag{B12}$$

$$D(s_i) = \left[(p_i + q_i) - (p_i - q_i)^2\right]D_0 \tag{B13}$$

with some fixed small time scale diffusion constant $D_0 = (\Delta x)^2/\Delta t$.

For the approximation of a diffusion process with constant diffusion coefficient, one has to ensure that the fluctuations for the Markov model are independent of x as well. For a given $F(s_i)$ and constant D this uniquely defines the parameters p_i and q_i. We get for the parameters, e.g. the teacher model

$$\begin{aligned}p_i &= \frac{D}{2D_0} + \frac{F(s_i)\Delta x}{2D_0}\left(1 + \frac{F(s_i)\Delta x}{D_0}\right) \\ q_i &= \frac{D}{2D_0} - \frac{F(s_i)\Delta x}{2D_0}\left(1 - \frac{F(s_i)\Delta x}{D_0}\right) \;.\end{aligned} \tag{B14}$$

One gets the same result up to order Δx by some straight forward discretization of eq. 1 with constant D.

Because of the local character of the ansatz, these relations are applicable for periodic boundary conditions too. Equations B8 and B9 for constant drift and diffusion are special cases of eq. B12 and B13.

Behind the difficulties in finding the expression for diffusion coefficient lies the problem of finding a limiting diffusion process for a Markov model.[19] For large time scales the approximation in terms of the probability distribution can be quite poor, e.g. for the space homogenous model if $(p_0 - q_0)$ is large.[38]

For finite boundary conditions this problems should be less serious, e.g. the space homogenous model as well as the Fokker-Planck equation with constant coefficients give a uniform distribution for large time scales.

[1] A. Barone and G. Paterno, *Physics and Applications of the Josephson effect*, J. Wiley, (1982).

[2] L. E. Baum and T. Petrie. Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Statist.*, **37**, 1554-1563, 1966.

[3] L. E. Baum, T. Petrie, G. Soules and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann. Math. Statist.*, **41**, 164–171, 1970.

[4] L. E. Baum. An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. in *Inequalities, III (Proc. 3rd Symp., Univ. Calif. Los Angeles, Calif., 1969; dedicated to the memory of Theodore S. Motzkin)*. New York: Academic, 1972.

[5] J. D. Becker, J. Honerkamp, J. Hirsh, E. Schlatter and R. Greger. Analyzing ion channels with hidden Markov models. *Pfğers Arch.*, **426**, 328–332, 1994.

[6] J. B. Boyce and B. A. Huberman. Superionic conductors: Transitions, structures, dynamics. *Phys. Rep.* **51**, 189–265, 1979.

[7] F. Campillo and F. LeGland. Mle for partially observed diffusions - direct maximization vs. the EM algorithm. *Stochastic Process. Appl.*, **33**, 245–274, 1989.

[8] S. Chung, J. B. Moore, L. Xia, L. S. Premkumar and P. W. Gage. Characterization of single channel currents using digital signal processing techniques based on hidden Markov models. *Phil. Trans. R. Soc. Lond. B*, **329**, 265–285, 1990.

[9] H. Daido. Quasi-Entrainment and Slow Relaxation in a Population of Interacting Oscillators. *Phys. Rev. Lett.*, **68**, 1073–1076, 1992.

[10] A. Dembo and O. Zeitouni. Parameter estimation of partially observed continuous-timestochastic processes via the EM algorithm. *Stochastic Process. Appl.*, **23**, 91–113, 1986.

[11] A. Dempster, N. Laird, D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Statist. Soc. Ser. B*, **39**, 1-38 1977.

[12] R. Durbin, S. Eddy, A. Krogh and G. Mitchison. *Biological Sequence Analysis.* Cambridge University Press, 2001.

[13] R. J. Elliot, L. Aggoun and J. B. Moore. *Hidden Markov Models: Estimation and Control.* Springer-Verlag, New York, 1994.

[14] Y. Ephraim and N. Merhav. Hidden Markov Processes. *IEEE Trans. Inform. Theory*, **48**. 1518–1569, 2002.

[15] D. R. Fredkin and J. A. Rice. Maximum likelihood estimation and identification directly from single-channel recording. *Proc. R. Soc. Lond. B*, **249**, 125–132, 1992.

[16] L. Gammaitoni, P. Hänggi, P. Jung and F. Marchesoni. Stochastic resonance. *Rev. Mod. Phys.*, **70**, 223–287, 1998.

[17] X. D. Huang, Y. Ariki, M. A. Jack. *Hidden Markov Models for Speech Recognition.* Edinburgh University Press, Edinburgh, 1990.

[18] Y. M. Kaniovski On Misapllication of Diffusion Approximations in Birth and Death Processes of Boisy Evolution. *Interim Report IR-98-050*, IIASA Laxenburg, 1998.

[19] Y. M. Kaniovski A comparison of diffusion approximations and actual limits in birth and death processes of noisy evolution. *Journal of Evolutionary Economics*, **10**, 545-555, 2000.

[20] R. L. Kautz. Using chaos to generate white noise. *J. Appl. Phys.*, **86**, 5794–5800, 1999.

[21] Y. Kuramoto. *Chemical Oscillations, Waves, and Turbulence.* Springer, Berlin, 1984.

[22] K. K. Likharev, "Dynamics of Josephson junctions and circuits", Gordon and Breach, (1986).

[23] R. S. Lipster and A. N. Shiryayev. *Statistics of Random Processes.* Vol I, Springer 1977.

[24] I. L. MacDonald and W. Zucchini. *Hidden Markov and Other Models for Discrete-valued Time Series.* Chapman-Hall, London, 1997.

[25] S. Michalek and J. Timmer. Estimating rate Constants in hidden Markov models by the EM algorithm. *IEEE Trans. Signal Processing* **47**, 226–228, 1999.

[26] E. Niebur, H. G. Schuster, D. M. Kammen and C. Koch. Oscillator-Phase Coupling for Different Two-Dimensional Network Connectivities. *Phys. Rev. A*, **44**, 6895–6904, 1991.

[27] L. R. Rabiner. A Tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE **77**, 257–286, 1989.

[28] L. R. Rabiner and B. H. Juang. *Fundamentals of Speech Recognition.* Engewood Cliffs, NJ: Prentice-Hall, 1993.

[29] M. Ragwitz and H. Kantz. Indispensable finite time corrections for Fokker-Planck equations from time series data. *Phys. Rev. Lett.* **87**, 254501, 2001.

[30] H. Risken. *The Fokker-Planck Equation.* Springer, Berlin, 1989.

[31] S. Siegert, R. Friedrich, J. Peinke. Analysis of Data Sets of Stochastic Systems. *Physics Letters A.* **243**, 275–280, 1998.

[32] J. Timmer. Parameter estimation in nonlinear stochastic differential equations. *Chaos, Solitons & Fractals***11**, 2571–2578, 2000.

[33] A. V. Ustinov, H. Kohlstedt, and P. Henne. Giant radiation linewidth of multi-fluxon states in long Josephson junctions, *Phys. Rev. Lett.* **77**, 3617–3620, 1996.

[34] M. Wagner, S. Michalek and J. Timmer. Estimating rate constants in aggregated Markov models of ion channel gating with loops and with nearly equal dwell times. *Proc. Roy. Soc. B*, **266**, 1919–1926, 1999.

[35] M. Wagner, J. Timmer. The effects of non-identifiability on testing for detailed balance in aggregated Markov models of ion-channel gating. *Biophys. J.*, **79**, 2918–2924, 2000.

[36] M. Wagner, J. Timmer. Model selection in non-nested Markov models for ion channel gating. *J. Theo. Biol.*,**208**,439–450, 2001.

[37] C. F. Wu. On the convergence properties of the EM algorithm. *Ann. Statist.*,**11**, 95–103, 1983.

[38] In practical applications it may be not possible to increase the sampling rate.