A Bayesian Framework for Modeling Diffusion Processes with Nonlinear Drift based on Nonlinear and Incomplete Observations

Hao Wu*and Frank Noé†

DFG Research Center Matheon, FU Berlin, Arnimallee 6, 14159 Berlin, Germany.

Abstract

Diffusion processes are relevant for a variety of phenomena in the natural sciences, including diffusion of cells or biomolecules within cells, diffusion of molecules on a membrane or surface, diffusion of a molecular conformation within a complex energy landscape. Many experimental tools exist now to track such diffusive motions in single cells or molecules, including high-resolution light microscopy, optical tweezers, fluorescence quenching, and Förster resonance energy transfer (FRET). Experimental observations are most often indirect and incomplete: (1) They do not directly reveal the potential or diffusion constants that govern the diffusion process, (2) they have limited time and space resolution, and (3) the highest-resolution experiments do not track the motion directly but rather probe it stochastically by recording single events, such as photons, whose properties depend on the state of the system under investigation.

Here, we propose a general Bayesian framework to model diffusion processes with nonlinear drift based on incomplete observations as generated by various types of experiments. A maximum penalized likelihood estimator is given as well as a Gibbs sampling method that allows to estimate the trajectories that have caused the measurement, the nonlinear drift or potential function and the noise or diffusion matrices, as well as uncertainty estimates of these properties. The approach is illustrated on numerical simulations of FRET experiments where it is shown that trajectories, potentials and diffusion constants can be efficiently and reliably estimated even in cases with little statistics or non-equilibrium measurement conditions.

*email: hwu@math.fu-berlin.de
†email: frank.noe@fu-berlin.de

1 Introduction

The microscopic dynamics of virtually all liquids and solids at finite temperature are governed by interaction potentials and Brownian motion [1]. As a result, many of these systems can be excellently modeled by diffusion in a potential. Examples include the diffusion of cells in liquids [2], the diffusion of biomolecules within cells [3] and the diffusion of molecules on a membrane or surface. The conformational dynamics of macromolecules can often be described as a diffusion process on an energy landscape in molecular state space [4]. Besides these, there are a large number of macroscopic phenomena that are also well-described by diffusion processes or random walks, even though there might be no obvious physical reason for it, including financial and climate systems [5, 6]. Here, we treat the data-based estimation of stochastic dynamics of the following type:

$$\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x})\mathrm{d}t + \mathbf{g}(\mathbf{x},t)\mathrm{d}t + \mathbf{\Sigma}^{\frac{1}{2}}\mathrm{d}\mathbf{W}, \tag{1}$$

where $\mathbf{x}(t) \in \mathbb{R}^d$ describes the state or position of the diffusing particle at time t, $\mathbf{f}(\mathbf{x}) \in \mathbb{R}^d$ is a deterministic position-dependent, but *a priori* unknown force which may, but not necessarily has to, be due to the gradient of an interaction potential. $\mathbf{g}(\mathbf{x},t) \in \mathbb{R}^d$ is a known external driving force which may arise from the way the particle is probed, and $\mathbf{\Sigma} \in \mathbb{R}^{d\times d}$ is a noise covariance matrix which determines the magnitude of the stochastic force caused by the d-dimensional Brownian motion (Wiener process) $\mathrm{d}\mathbf{W}$. The combined deterministic force $\mathbf{f}+\mathbf{g}$ is also often referred to as drift.

The last decade has seen dramatic improvements in the ability to directly measure trajectories of single particles. The location of single cells and even single molecules can now be tracked in high spatiotemporal resolution microscopy techniques [7, 8, 9, 10, 11, 12]. Time-dependent intramolecular distances and contact formation can be probed using spectroscopic techniques such as single-molecule Förster resonance energy transfer (FRET) [13, 14, 15, 16] or fluorescence quenching [17, 16], and with nanoscopic mechanical devices such as atomic force microscopes or optical tweezers [18, 19]. All these techniques allow to measure observable functions of individual realizations $\mathbf{x}(t)$ of Eq. (1), i.e.:

$$Q(t) \sim p(Q(t) \mid \mathbf{x}(t)) \tag{2}$$

and this observation is generally incomplete or uncertain in some sense: There is some inherent time and space resolution, often dictated by the physics of the measurement process. There might exist significant measurement errors due to background noise or other unintended activation of the sensors used. Finally, in many cases $Q(t)$ is only a low-dimensional projection of $\mathbf{x}(t)$ and sometimes is even a stochastic signal.

In statistics, a host of methods have been developed to robustly and efficiently estimate processes of type (1,2), see [20, 21, 22] and references therein. These methods have been much applied to econometric problems, such as stock market prices. However, there is a lack of methods that can efficiently estimate diffusion processes with unknown $\mathbf{f}$, especially when this force (and the associated potential, if there is any) is arbitrarily nonlinear and high-dimensional. This is a very important aspect of complex diffusion processes in physics, chemistry and biology. On the other hand, a number of methods have been proposed to estimate specific features or specific forms of diffusion processes based on experimental data in these application fields. Just to give some examples, this includes an estimator for correlation times based on one-dimensional diffusion [23], estimation of Ornstein-Uhlenbeck or bistable processes [24] and estimation of diffusion process without potential [25]. When a separation of slow and fast timescales exists, a combination of Ornstein Uhlenbeck processes and Hidden Markov Models can be used [26]. Estimation methods for the unknown $\mathbf{f}$ with discrete-time observations were developed as well [27, 28, 29, 30, 31, 32, 33]. Most of these estimation approaches are based on the framework of maximum likelihood or Bayesian estimation. Some studies have used generalized score functions [34] and variational inference [35] for parameter estimation. [36, 37] has proposed nonparametric estimators for $\mathbf{f}$, which are applicable in the case that its functional form unknown. In general, observations that are noisy or incomplete are still very challenging for these approaches.

Here, we present a Bayesian framework that allows to estimate the probability distributions of trajectories, $\mathbf{x}(t)$, the deterministic force $\mathbf{f}(\mathbf{x})$ (and the potential that has generated it, if there is one), and the noise matrix Σ (and the related diffusion matrix $\mathbf{D}$) of equations of type (1) from such partial and uncertain observations. The key progress over previous estimation methods is that the force $\mathbf{f}$ is modeled in terms of a fuzzy combination of basis functions, which allows arbitrarily nonlinear forces to be approximated with a reasonable number of parameters even in high-dimensional spaces. We derive both an estimation-maximization based maximum penalized likelihood estimator as well as a Gibbs sampling scheme for generating models $(\mathbf{f},\Sigma)$ and trajectories $(\mathbf{x}(t))$ from the posterior distribution. Mean values and uncertainties of these properties can then be straightforwardly estimated. This

framework is widely applicable to model complex dynamical processes based on experimental or other observation data and is shown to be useful for the analysis of single molecule fluorescence experiments.

2 Diffusion model

Let $\mathbf{x}(t) \in \mathbb{R}^d$ describe the time-dependent position of the system which is a solution to the stochastic differential equation (1). In many physical systems, the deterministic force $\mathbf{f}(\mathbf{x})$ is caused by a potential $U(\mathbf{x})$. Moreover, in many of these systems, the fluctuation-dissipation theorem holds, relating diffusion constant and temperature, such that equation (1) has the following form:

$$\mathrm{d}\mathbf{x} = -\frac{1}{kT}\mathbf{D}\nabla U(\mathbf{x})\mathrm{d}t + \mathbf{g}(\mathbf{x},t)\mathrm{d}t + \sqrt{2}\mathbf{D}^{\frac{1}{2}}\mathrm{d}\mathbf{W}, \tag{3}$$

with diffusion matrix $\mathbf{D} = \frac{1}{2}\Sigma$, Boltzmann constant k and temperature T. The force $\mathbf{g}(\mathbf{x},t)$ is some known external forcing, which may be 0, when the system is observed without perturbation, may be constant such as in the presence of an electric field, or time-dependent such as in single-molecule pulling experiments. $\mathbf{x}(t)$ is a continuous time Markov process, and in absence of the driving force $\mathbf{g}(\mathbf{x},t)$, the system has the invariant density given by the Boltzmann density [38]:

$$\pi(\mathbf{x}) \propto \exp\left(-\frac{U(\mathbf{x})}{kT}\right). \tag{4}$$

The form 3 induces some physical constraints on the parameters. In particular, if the components of $\mathbf{x}$ are independent, the diffusion matrix is diagonal $\mathbf{D} = \mathrm{diag}\{D_1, ..., D_m\}$ and $\mathbf{f}(\mathbf{x}) = -\frac{1}{kT}\mathrm{diag}\{d_1, ..., d_m\}\nabla U(\mathbf{x})$ is a conservative vector field. A special case of this is the most common case, namely that diffusion is isotropic, $\mathbf{D} = D\mathbf{I}$ with $D \in \mathbb{R}$ and $\mathbf{I}$ being the identity matrix. However, the present framework also allows for applications which do not obey such constraints. Thus, unless otherwise noted, we will model the system under study based on the general form (1). $\mathbf{f}(\mathbf{x})$ (and $U(\mathbf{x})$) are unknown and can be arbitrarily nonlinear. In order to estimate them, they may be expressed by some parametric model. The present framework allows for models which have the following form:

$$\mathbf{f}(\mathbf{x}) = \mathbf{\Xi}(\mathbf{x})\,\lambda \tag{5}$$

where $\lambda \in \mathbb{R}^m$ is a parameter vector and $\mathbf{\Xi}(\mathbf{x}) \in \mathbb{R}^{d\times m}$ is a matrix of pre-defined basis functions that are combined with the parameters to a general nonlinear $\mathbf{f}$. Computationally convenient models can be constructed from normalized fuzzy membership functions $\mu_l(\mathbf{x})$, e.g. defined by

$$\mu_l(\mathbf{x}) = \frac{\exp\left(-\alpha(\mathbf{x}-\mathbf{x}_l)^2\right)}{\sum_k \exp\left(-\alpha(\mathbf{x}-\mathbf{x}_k)^2\right)}, \tag{6}$$

where α is a size parameter and $\mathbf{x}_l$ are pre-defined support points in the relevant ranges of $\mathbf{x}$. There are many ways of combining these parameters to obtain models of the form (5). A straightforward model is to express $\mathbf{f}(\mathbf{x})$ in terms of L locally linear force models, which are superimposed *via* membership functions $\mu_l(\mathbf{x})$ into a general nonlinear force [39]:

$$\mathbf{f}(\mathbf{x}) = \sum_{l=1}^{L} \mu_l(\mathbf{x})\left(a_l\mathbf{x} + b_l\right), \tag{7}$$

Defining the vectors $\lambda = (a_1, \ldots, a_L, b_1, \ldots, b_L)^{\mathrm{T}}$ and $\mathbf{\Xi}(\mathbf{x}) = [\mathbf{x}\mu_1(\mathbf{x}), \ldots, \mathbf{x}\mu_L(\mathbf{x}), \mu_1(\mathbf{x}), \ldots, \mu_L(\mathbf{x})]$ allows (7) to be written as Eq. (5). The fuzzy model described in (7) has been proved to be a very good representation for continuous nonlinear functions due to its universal approximation capability [40]. Note that a $\mathbf{f}(\mathbf{x})$ of this form is not guaranteed to be conservative, and thus a generating potential $U(\mathbf{x})$ may not exist. An alternative approach that formulates the model in terms of a potential and then derives $\mathbf{f}(\mathbf{x})$ from its gradient is given in the Appendix.

Thus, the parameters $\theta = (\lambda, \mathbf{\Sigma})$ completely describe the equation of motion and need to be estimated from the observed measurement signal.

2.1 Probability model

Based on the above model we can now derive the probability for a trajectory $\mathbf{x}(t)$ given a set of parameters θ and from that the probability to observe the recorded photon trajectories. This step will allow us to estimate the unknown parameters and the hidden trajectories $\mathbf{x}(t)$ from the data.

First, the equation of motion (1) is discretized using the Euler scheme with time steps τ, which may correspond to the experimental time resolution. The discrete dynamics within one trajectory are denoted by $\mathbf{x}_n$, $n = 0...N$:

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \tau \mathbf{f}\left(\mathbf{x}_n\right) + \tau \mathbf{g}(\mathbf{x}_n) + \sqrt{\tau}\mathbf{W}_n \tag{8}$$

where $\mathbf{W}_n \sim \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}\right)$ is d-dimensional white noise with covariance matrix Σ. In other words, the probability of observing a certain $\mathbf{x}_{n+1}$ given the previous $\mathbf{x}_n$ is given by the previous step plus the displacement due to the deterministic force caused by the potential, and a random displacement resulting from the noise $\mathbf{\Sigma}$:

$$\mathbb{P}(\mathbf{x}_{n+1} \mid \mathbf{x}_n) \quad = \quad \frac{1}{(2\pi)^{\frac{d}{2}} |\tau\mathbf{\Sigma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2\tau}\mathbf{s}^T \Sigma^{-1} \mathbf{s}\right), \tag{9}$$

with

$$\mathbf{s} = \mathbf{x}_{n+1} - \mathbf{x}_n - \tau\mathbf{f}(\mathbf{x}) - \tau\mathbf{g}(\mathbf{x})$$

3 Model Estimation from Observation

Using Bayesian inference, the joint probability of a set of parameters θ and the effective distance trajectories x is given by:

$$p(\theta, \mathbf{x} \mid Q) \propto p\left(\theta\right)\ p(\mathbf{x} \mid \theta)\, p(Q \mid \mathbf{x}) \tag{10}$$

where $p(\theta)$ is the prior distribution of parameters, $p\left(\mathbf{x} \mid \theta\right)$ is the probability that a particular series of hidden trajectories will be generated by parameter set θ while $p(Q \mid \mathbf{x})$ is the probability of observing the measured signal from these hidden trajectories. Since this probability distribution cannot be evaluated analytically, an estimation procedure is proposed to generate samples of $(\theta, \mathbf{x})$ distributed from $p(\theta, \mathbf{x} \mid Q)$ using Markov chain Monte Carlo (MCMC). This is done using a Gibbs sampling scheme which, in brief, consists of the following iteration:

1. Generate a new sample of $\mathbf{x}$ while keeping the parameters θ constant. This is done by sampling each $\mathbf{x}_n^{(k)}$ while keeping the remaining $\mathbf{x}$ constant and cycling through all $\mathbf{x}_n^{(k)}$ in this way.
2. Generate a new sample of θ for fixed trajectories $\mathbf{x}$.

3.1 Estimating the trajectory for given parameters

Consider the distribution of one trajectory $\mathbf{x}$ for fixed parameters θ obtained from Eq. (10):

$$p(\mathbf{x} \mid \theta, Q) \propto p(Q \mid \mathbf{x})\, p(\mathbf{x} \mid \theta)$$

We generate a sample of $\mathbf{x}$ by a Gibbs scheme that provides a sample of each $\mathbf{x}_n$ while keeping the other time steps fixed. Due to Markovianity of the equation of motion we have:

$$\begin{aligned}
p(\mathbf{x}_n \mid \mathbf{x}_{0\ldots n-1, n+1\ldots N}, \theta, Q) \quad &\propto \quad p(\mathbf{x}_n \mid \mathbf{x}_{n-1}, \mathbf{x}_{n+1}, \theta, Q) \\
&\propto \quad p(Q_n \mid \mathbf{x}_n)\, p(\mathbf{x}_n \mid \mathbf{x}_{n-1}, \mathbf{x}_{n+1}, \theta) \\
&\propto \quad p(Q_n \mid \mathbf{x}_n)\, p(\mathbf{x}_n, \mathbf{x}_{n-1}, \mathbf{x}_{n+1}, \theta) \\
&\propto \quad p(Q_n \mid \mathbf{x}_n)\, p(\mathbf{x}_{n+1} \mid \mathbf{x}_n, \theta)\, p(\mathbf{x}_n \mid \mathbf{x}_{n-1}, \theta)
\end{aligned}$$

where $p(\mathbf{x}_{n+1} \mid \mathbf{x}_n, \theta)$ and $p(\mathbf{x}_n \mid \mathbf{x}_{n-1}, \theta)$ can be evaluated using equation (9), while the output probability $p(Q_n \mid \mathbf{x}_n)$ needs to be calculated depending on the specific application (e.g. see Eq. (22) for FRET). Thus $p(\mathbf{x}_n \mid \theta, Q, \mathbf{x}_{0\ldots n-1}^{(j)}, \mathbf{x}_{n+1\ldots N}^{(j-1)})$ can be sampled with a Metropolis-Hastings sampling step [41], where a value for $\mathbf{x}_n$ is proposed with a proposal probability $p(\mathbf{x}_n \to \mathbf{x}_n^*)$ and then accepted with the following acceptance probability:

$$p_{acc}\left(\mathbf{x}_n^*\right) = \min\left\{1, \frac{p(\mathbf{x}_n^* \to \mathbf{x}_n)\, p(Q_n \mid \mathbf{x}_n^*)\, p(\mathbf{x}_n^* \mid \mathbf{x}_{n-1}) p(\mathbf{x}_{n+1} \mid \mathbf{x}_n^*)}{p(\mathbf{x}_n \to \mathbf{x}_n^*)\, p(Q_n \mid \mathbf{x}_n)\, p(\mathbf{x}_n \mid \mathbf{x}_{n-1}) p(\mathbf{x}_{n+1} \mid \mathbf{x}_n)}\right\} \tag{11}$$

in principle any proposal step $p(\mathbf{x}_n \to \mathbf{x}_n^*)$ can be used. A simple and efficient approach is to directly use the equation of motion to propose a step via Eq. (9), which simplifies the acceptance probability to:

$$p_{acc}(\mathbf{x}_n^*) = \min\left\{1, \frac{p(Q_t \mid \mathbf{x}_t^*)}{p(Q_t \mid \mathbf{x}_t)} \frac{p(\mathbf{x}_n^* \mid \mathbf{x}_{n-1}) p(\mathbf{x}_{n+1} \mid \mathbf{x}_n^*)}{p(\mathbf{x}_n \mid \mathbf{x}_{n-1}) p(\mathbf{x}_{n+1} \mid \mathbf{x}_n)}\right\} \tag{12}$$

which can be evaluated with Eq. (9). An alternative efficient proposal step is described in the Appendix.

3.2 Parameter estimation for given trajectory

For estimating the parameters we now need to consider the fact that one may have multiple trajectories $\mathbf{x}^1, \ldots, \mathbf{x}^K$ which are independent observations of the same system, or of different systems with the same dynamics. From (10), the distribution of parameters for fixed trajectories is given by:

$$p\left(\theta \mid \mathbf{x}^1, \ldots, \mathbf{x}^K\right) \propto p(\theta) \prod_{k=1}^{K} p(\mathbf{x}^k \mid \theta)$$

the prior probability $p(\theta) = p(\lambda, \Sigma)$ is significant when very little observation data is available and must restrict the parameters Σ, λ to meaningful values in this situation. When a reasonable amount of observation is available, the posterior probability distribution will be dominated by the likelihood. For the present case, the prior distribution is given by:

$$\begin{cases} \Sigma \sim \mathcal{W}^{-1}\left(\Psi_0, m_0\right) \\ \lambda \sim \mathcal{N}\left(0, \Sigma_0\right) \end{cases} \tag{13}$$

where $\mathcal{W}^{-1}\left(\Psi_0, m_0\right)$ denotes the inverse Wishart distribution with inverse scale matrix Ψ_0 and degrees of freedom $m_0 > d-1$. From a physical point of view, this prior keeps the parameters in their allowed ranges (especially $\Sigma > 0$ and excluding extremely small and large values of θ).

The likelihood is given by

$$p(\mathbf{x}^k \mid \theta) \propto |\Sigma|^{-\frac{N}{2}} \exp\left(-\frac{1}{2\tau} \sum_{n=0}^{N-1} \mathbf{v}(\mathbf{x}_{n+1}, \mathbf{x}_n)^{\mathrm{T}} \Sigma^{-1} \mathbf{v}(\mathbf{x}_{n+1}, \mathbf{x}_n)\right)$$

with

$$\mathbf{v}(\mathbf{x}_{n+1}, \mathbf{x}_n) = \mathbf{x}_{n+1}^k - \mathbf{x}_n^k - \tau \mathbf{g}(\mathbf{x}_n^k) - \tau \Xi\left(\mathbf{x}_n^k\right) \lambda$$

Using some elementary transformations (see Appendix for complete derivation), it turns out that, $p\left(\theta | \mathbf{x}^1, \ldots, \mathbf{x}^K\right)$ can be sampled as:

$$\begin{cases} \Sigma \mid \mathbf{x}^1, \ldots, \mathbf{x}^K, \lambda \sim \mathcal{W}^{-1}\left(\Psi_0 + \mathbf{C}_s, m_0 + NK\right) \\ \lambda \mid \mathbf{x}^1, \ldots, \mathbf{x}^K, \Sigma \sim \mathcal{N}\left(\tilde{\lambda}, \left(\Sigma_0^{-1} + \mathbf{A}_s\right)^{-1}\right) \end{cases} \tag{14}$$

with the functionals:

$$\mathbf{A}_s \stackrel{\text{def}}{=} \tau \sum_{k=1}^{K} \sum_{n=0}^{N-1} \Xi\left(\mathbf{x}_n^k\right)^{\mathrm{T}} \Sigma^{-1} \Xi\left(\mathbf{x}_n^k\right) \tag{15}$$

$$\mathbf{B}_s \stackrel{\text{def}}{=} 2 \sum_{k=1}^{K} \sum_{n=0}^{N-1} \left(\mathbf{x}_{n+1}^k - \mathbf{x}_n^k - \tau \mathbf{g}(\mathbf{x}_n)\right)^{\mathbf{T}} \Sigma^{-1} \Xi\left(\mathbf{x}_n^k\right) \tag{16}$$

$$\tilde{\lambda} \stackrel{\text{def}}{=} -\frac{1}{2}\left(\Sigma_0^{-1} + \mathbf{A}_s\right)^{-1} \mathbf{B}_s \tag{17}$$

$$\mathbf{C}_s \stackrel{\text{def}}{=} \frac{1}{\tau} \sum_{k=1}^{K} \sum_{n=0}^{N-1} \left(\mathbf{x}_{n+1}^k - \mathbf{x}_n^k - \tau g(\mathbf{x}_n^k) + \tau \Xi\left(\mathbf{x}_n^k\right) \lambda\right) \left(\mathbf{x}_{n+1}^k - \mathbf{x}_n^k - \tau \mathbf{g}(\mathbf{x}_n^k) + \tau \Xi\left(\mathbf{x}_n^k\right) \lambda\right)^{\mathbf{T}} \tag{18}$$

3.3 Sampling the joint distribution of trajectories and parameters

Having schemes available that allow the parameters to be sampled for a given trajectory and the trajectory to be sampled for given parameters, we can now construct a Gibbs-sampling scheme, where either sets of variables are sampled while keeping the others fixed. Such a combined sampling scheme provides a sample of the joint probability distribution of trajectories and parameters for a given observation, and the resulting joint sample can be used to extract the sample means and statistical uncertainties of the trajectory and the parameters. Algorithm 1 summarizes the Gibbs sampling

scheme for the general case of multiple observation trajectories. An alternative to sampling the full distribution is to compute the maximum probability (or maximum likelihood) solution which only yields a single trajectory and parameter set. Such an approach is given in the Appendix.

Algorithm 1 Gibbs Sampler for $p\left(\mathbf{x}_{0:N}^1, \ldots, \mathbf{x}_{0:N}^K, \theta | Q\right)$

(**Initialization**)
sample $\mathbf{x}_{0:N}^{1,(0)}, \ldots, \mathbf{x}_{0:N}^{K,(0)}$ and $\theta^{(0)}$ from an arbitrary initial distribution
(**Iteration**)
for $j = 1$ to M
  for $k = 1$ to K
    for $n = 0$ to N
      sample $\mathbf{x}_n^{k,(j)} \sim p\left(\mathbf{x}_n^k \mid \mathbf{x}_{n-1}^{k(j)}, \mathbf{x}_{n+1}^{k(j-1)}, \theta^{(j-1)}, Q\right)$
    end for
  end for
  sample $\Sigma^{(j)} \sim p\left(\Sigma \mid \mathbf{x}_{0:N}^{1,(j)}, \ldots, \mathbf{x}_{0:N}^{K,(j)}, \lambda^{(j-1)}\right)$
  sample $\lambda^{(j)} \sim p\left(\lambda \mid \mathbf{x}_{0:N}^{1,(j)}, \ldots, \mathbf{x}_{0:N}^{K,(j)}, \Sigma^{(j)}\right)$
end for

With the procedure described above, each iteration j generates samples $\mathbf{x}^{k,(j)}$ of all K full trajectories and a sample $\theta^{(j)}$ of the parameters. In order to escape from the bias induced by the initial draw, the sampler should be run with a burn-in phase until the mean values and standard deviations of the properties of the $\mathbf{x}_n$ and θ have stabilized and show no more drift. These initial samples must be neglected from the analysis. From the ensemble of M samples, the properties of interest can now be estimated. In particular, the expectation values and standard deviations (uncertainties) of the effective distance trajectories $\mathbf{x}_n^k$ and the diffusion constant $\mathbf{D} = \frac{1}{2}\Sigma$ can be directly computed. If $\mathbf{D} = \mathrm{diag}\{D_1, ..., D_m\}$ is diagonal, then there also exists a corresponding potential energy that is unique up to an additive constant. In order to obtain the potential in each sampling step, one numerically integrates the force:

$$\frac{U^{(j)}(\mathbf{x})}{kT} \propto -\mathrm{diag}\{D_1^{-1}, ..., D_m^{-1}\} \int_0^{\mathbf{x}} \theta^{(j)\mathrm{T}} \xi\left(\mathbf{x}'\right) d\mathbf{x}' \tag{19}$$

Thus, expectation values and uncertainties can also be straightforwardly computed for the potential of mean force and the stationary distribution that is obtained by Eq. (4).

4 Application to Single-Molecule Experiments

We now show a few examples for different settings in which the estimation framework presented above is useful. All of them are microscopy techniques that are widely used in Physics, Chemistry and Biology.

4.1 Single-Molecule Förster Resonance Energy Transfer

Förster resonance electron transfer (FRET) is an experimental technique used to measure nanometer distances between pairs of chemical groups. It is commonly employed to track binding or folding processes in macromolecules [13, 14, 15]. In a FRET experiment, one attaches two chemical groups, one donor and one acceptor dye, to the molecule(s) to be measured. The donor is then specifically excited with a laser and subsequently relaxes either by emitting a donor photon (usually green), or by transferring the energy to a nearby acceptor dye which then emits an acceptor photon (usually red). The probability of an emitted photon being an acceptor photon is called transfer efficiency and related to the inter-dye distance r by the Förster law:

$$E = \frac{R_0^6}{r^6 + R_0^6}. \tag{20}$$

Here, R_0 is the Förster radius; a number that depends on the chemical composition of the dyes, the permittivity of the medium between them and on their orientation. E is an interesting "coordinate" of the system which characterizes its conformational state (although a conversion of E to physical distances and angles is difficult).

FRET experiments can now be conducted with single molecules, which provides trajectory information useful for the present model estimation. In single molecule experiments, one cannot directly

obtain a time-continuous E as in each time step one only obtains a few red, green, or (very often) no photon at all. In typical experimental analyses this is addressed by introducing a time window τ and computing an averaged $E_\tau(t)$. However, the results of the estimation will depend on the choice of τ and very often this leads to important dynamical information to be missed. Here, the model estimation can be straightforwardly made in terms of the photon trajectories by treating the recorded photons as an observation of the dynamics.

We propose a one-dimensional model in terms of the effective normalized distance x:

$$x(t) \stackrel{\text{def}}{=} \left(\frac{1}{E(t)} - 1\right)^{\frac{1}{6}} = \frac{r(t)}{R_0}, \tag{21}$$

which is proportional to the physical inter-dye distance via $r(t) = R_0 x(t)$ if the Förster radius R_0 is approximately constant for the process under investigation. Even if that is not the case, it is expected that $x(t)$ and $r(t)$ are well correlated. Although the equation of motion could be formulated in $E(t)$ instead of $x(t)$, x is likely a better modeling variable since it allows changes in the potential to be described at distances much smaller or larger than R_0 which would occur near the 0 or 1 boundary of E. FRET experiments can in principle include more than two dyes and thus be sensitive to more than one distance, which then calls for a multidimensional model [42].

FRET experiments are passive, tracking the dynamics that are brought about by Brownian motion. Thus, $g(x,t) = 0$ and Eq. ((1)) becomes a one-dimensional Smoluchowski equation:

$$dx = -\frac{D}{kT}\nabla U(x)\mathrm{d}t + \sqrt{2D}dW.$$

At each time step n, a number of acceptor and donor photons, $Q_n = \{A_n, D_n\}$, may be observed. The probability of this event is (up to a constant that depends on the laser intensity and is not relevant for the analysis) given by:

$$p(Q_n \mid x_n) \propto \left(\frac{1}{x\left(t\right)^6 + 1}\right)^{A_n} \left(\frac{x\left(t\right)^6}{x\left(t\right)^6 + 1}\right)^{D_n}. \tag{22}$$

All the the estimation equations above then are one-dimensional and lend themselves to a simple and efficient estimation procedure.

Note a special case of the method proposed here is the one where the potential is identically zero and thus diffusion is free. This is identical to assuming that $\{x\left(t\right)\}$ is a realization of Wiener process with effective diffusion coefficient D_w, *i.e.* $x\left(t+s\right)|x\left(t\right) \sim \mathcal{N}\left(x\left(t\right), 2D_w s\right)$. Such a variant along with a corresponding maximum likelihood estimator has been proposed in [25]. This method is contained in our method as a special case that is obtained by setting $\mathbf{f} \equiv 0$ in (1).

4.2 Single Molecule Pulling with Optical Traps

Optical traps, or tweezers, are a measurement and manipulation tool now widely used in nanotechnology and biophysics [12, 19]. Here, one uses optical beads of sub-micrometer size that bend laser light and are held in the laser focus due to restoring forces caused by the photon momenta. The bead can be moved around by repositioning the laser beam. At the same time, the trap can also be used to track the position of the bead when this is displaced by a mechanical force. For example, one can use two beads and connect them by a macromolecule than than change its conformation, e.g. by unfolding and folding. One can then track intramolecular distance changes when holding the trap on a constant position, or induce unfolding by pulling on the trap and then letting loose in order to watch the molecule refold. The model coordinate here is the inter-bead distance, $x(t)$, and is again one-dimensional. The diffusion equation relevant for this case is given by:

$$\mathrm{d}x = -\frac{D}{kT}\nabla U(x)\mathrm{d}t + \alpha(x - x_0(t))^2 + \sqrt{2D}\mathrm{d}W.$$

where $g(x,t) = \alpha(x - x_0(t))^2$ is a model for the restoring trap force and $x_0(t)$ is the reference position of the trap set by the experimenter over the measuring time t. Since x is observed directly, the fraction $p(Q_n \mid x_n^*)/p(Q_n \mid x_n)$ drops from the acceptance probability (12).

The present framework can be used to estimate the potential $U(x)$ despite the fact that experiment can be arbitrarily driven trough $g(x,t)$.

4.3 Single Particle Tracking

Single particle tracking is a fluorescence technique in which single molecules are labeled with a fluorescent dye and tracked with a fluorescence microscope [10, 11, 43]. The experiment may be set up so as to watch the interaction of labeled particles with other molecules on a surface or membrane, by attaching this surface on a glass surface, into which the excitation laser light is scattered at a total reflection angle. As a result, the laser beam will not leave the glass surface, but there is an evanescent field just thick enough to have a significant probability to excite molecules that are directly adjacent to the surface. This will allow to collect a number 2-dimensional trajectories of molecules that track their diffusion above the surface. In order setting, this leads to a 2- dimensional model with no forcing g,

$$\mathrm{d}\mathbf{x} = -\frac{D}{kT}\nabla U(\mathbf{x})\mathrm{d}t + \sqrt{2D}\mathrm{d}\mathbf{W}.$$

where we have assumed isotropic diffusion. For modeling $U(\mathbf{x})$, a dense grid of basis functions in the 2d-area of interest may be used.

5 Numerical Examples

We now illustrate the usefulness of the Bayesian framework to estimate trajectories, potentials and diffusion constants from FRET measurements. In order to validate the method, no real experimental data is used here, but artificial model data with experimentally realistic parameters.

In order to rate the performance of our method, its estimation results are compared to the standard time window method commonly used in experimental analyses [44]. In this method, the FRET efficiency and inter-dye distance are assumed to be approximately constant within a time window of width T_w. Then $E(t)$ and $x(t)$ are estimated as

$$\hat{E}(t) = \frac{|T_A(t)|}{|T_A(t)| + |T_D(t)|} \tag{23}$$

and

$$x(t) = \left(\frac{1}{\hat{E}(t)} - 1\right)^{\frac{1}{6}} \tag{24}$$

where $T_A(t)$ and $T_D(t)$ count the number of acceptor and donor photons in the time window $[t - T_w/2, t + T_w/2]$, respectively. Here we adopt the "leave-one-out cross-validation" technique [45] to select the value of T_w. In the time-window method, the stationary distributions $\pi(x)$ are conventionally estimated by simply binning $\{x(t)\}$ under the assumption that $\{x(t)\}$ are samples of $\pi(x)$. Here we use the kernel density approximations [46] to get continuous $\hat{\pi}(x)$ from $\{x(t)\}$ and compute the approximation of potential functions by (4).

We address two numerically simulated FRET experiments with artificial potential functions. For modeling the nonlinear force $f(x)$ we used Gaussian-shaped membership functions as suggested in Eq. (6) with $\alpha = 1$ and positions $x_l = \{0, 0.5, 1, 1.5, 2\}$. The estimation procedure uses the following parameters: $M = 10000$, $J = 20$, $\Sigma_0 = 10^4$, $\Psi_0 = m_0 = 1$. And the initial sample of trajectories is provided by the time-window method.

5.1 Potential/invariant distribution estimation

We first consider a potential function that exhibits metastability, defined by

$$V(\bar{r}) = \begin{cases} -111.01\bar{r}^3 + 178.63\bar{r}^2 - 82.27\bar{r} + 8.98, & \bar{r} < 0.75 \\ 182.8915\bar{r}^3 - 482.64\bar{r}^2 + 413.69\bar{r} - 115.01, & 0.75 \le \bar{r} < 1 \\ -153.36\bar{r}^3 + 526.11\bar{r}^2 - 595.06\bar{r} + 221.24 & 1 \le \bar{r} < 1.25 \\ 84.94\bar{r}^3 - 367.53\bar{r}^2 + 521.98\bar{r} - 244.20, & 1.25 < \bar{r} \end{cases} \tag{25}$$

See Fig.(1) (right, black line) for an illustration. The noise intensity is set to $\Sigma = [1.4^2]$. 30 trajectories of length 3 s are run using the Heun scheme [47] with simulation time step $\Delta t = 50\,\mu s$. The initial points are generated from the invariant density

$$x(0) \sim \pi(x) \propto \exp\left(-\frac{2V(x)}{\Sigma^2}\right) \tag{26}$$

The probability of recording a single photon in each time step is 0.2. This setup mimics the situation in FRET experiments of immobilized molecules [48].

Fig. (1) shows the invariant distribution $\pi(x)$ (right column) and the corresponding potential $U(x)$ (left column) of the system (true values in black lines) and compares the corresponding estimates from the Bayesian estimator proposed here (panels a and b, mean shown as solid blue line, one-standard-error confidence intervals as dashed blue line). Panels c and d show the time-window based estimates for short, intermediate and long window lengths. By direct counting, the time window estimate would have a spurious metastable state at $x = 0$, resulting from the fact that for many time windows the distance cannot be estimated due to lack of donor photons. In order to avoid this, time windows with $x = 0$ were removed before generating the estimate shown in Fig. (1). A consequence of this is that for short time windows the distribution and the potential cannot be estimated at small values of x. Both the time-window estimate with an intermediate window length and the Bayesian estimate capture the essential features of the system, i.e. the existence of the three metastable states and their relative depths / probabilities.

The Bayesian estimator has the advantage that it also provides a rigorous error estimate, and the true invariant density and potential are indeed correctly estimated by the Bayesian mean within this uncertainty. The time window estimates miss some features, in part due to lack of photons in short windows, in part due to smoothing out effects at long time windows. Clearly, for $T_W \to \infty$, the time window method will estimate $\pi(x)$ as a δ-distribution located at the mean distance. This bias leads to a more than tenfold increase of the estimation error in the invariant density of the time window estimate compared to the Bayesian estimate (see Table (1)). Finally, the true noise intensity of $\sqrt{\Sigma} = 1.4$ is well approximated by the Bayesian estimate of $\sqrt{\Sigma} = 1.8372 \pm 0.1130$.

Panels e and f show the results for the Bayesian estimator for the case where the potential is flat, i.e. the deterministic force is zero. This is a "free diffusion" estimator of the type proposed in [49]. It is seen that this model performs much better than window-based estimators. However, at small and large distances, where the potential increases significantly, this estimator is significantly off.

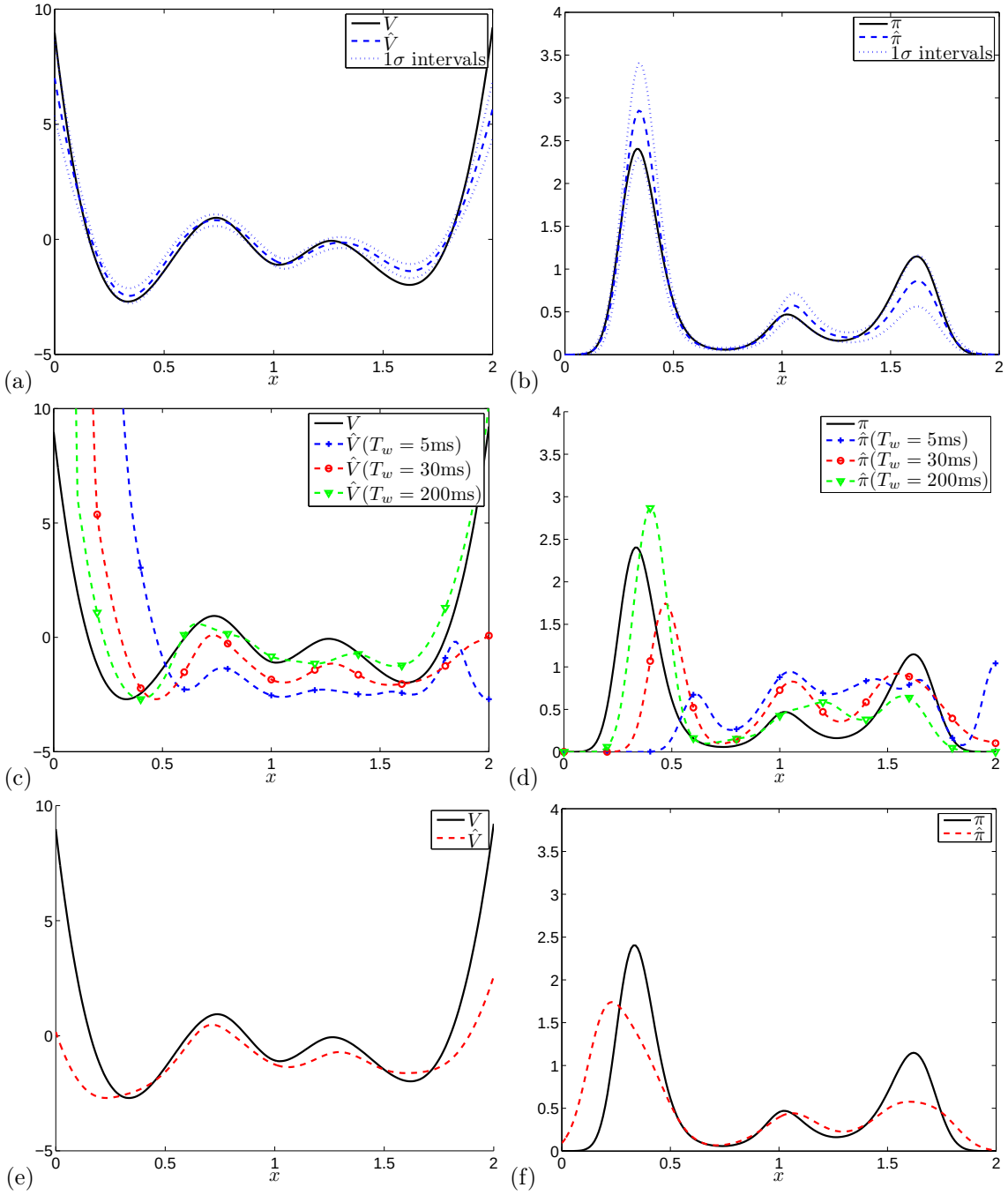

Figure 1: Invariant distribution and potential function estimation results from few long trajectories comparing the true values (solid line) with estimations from different methods. (a,b) Bayesian estimator, (c,d) time-window estimate using different time windows, (e,f) free diffusion estimator.

Table 1: Quantitative comparison of errors of estimating the trajectory and the stationary distribution from for few long trajectories using different methods (Bayesian estimator, time-window estimator with different window widths, Bayesian estimator with a flat potential/free diffusion). The error of the trajectories was defined as: $\mathrm{MSE}_x = \frac{1}{T}\int_0^T \left(\hat{x}\left(t\right) - \bar{r}\left(t\right)\right)^2 \mathrm{d}t$, and the error of the stationary distribution as: $D_{KL}\left(\pi \parallel \hat{\pi}\right) = \int \pi\left(x\right) \log\left(\pi\left(x\right)/\hat{\pi}\left(x\right)\right) \mathrm{d}x$.

	Bayesian	$T_w = 5$ ms	$T_w = 30$ ms	$T_w = 200$ ms	Free diffusion
Error in trajectories $x(t)$	0.0066	0.0865	0.0544	0.0262	0.0138
Error in stationary distribution $\pi(x)$	0.0196	5.9601	1.0862	0.2379	0.2160

5.2 Trajectory estimation

Fig. 2(black) shows one of the 30 trajectory realizations (black line) along with estimates from the Bayesian method (B), the time-window (TW) estimates and the Bayesian method with free diffusion.

This system poses a hard estimation problem, because the main metastable state at $x \approx 0.3$ has a FRET efficiency close to 100%, i.e. very few donor photons will be emitted in that state and with any time window that is not very long, the time-window method will often estimate a distance of 0. This is seen in the time window estimation of the trajectory in panels b and c, which frequently drops to $x = 0$ for small time windows. At the same time, any time window that is not very short will smooth over the trajectory and loose many fast features of the dynamics. This is illustrated in panels b and d, where long time windows are seen to smooth out the transitions between metastable states 2 and 3, arriving at an averaged distance estimate of ≈ 1.2, a value that is actually close to a potential maximum rather than a stable state, thus providing a deceptive picture of the dynamics. In conclusion, no single time window length is adequate for all data.

The Bayesian estimator (panel a, mean as blue solid line, 1-standard-error confidence intervals as blue dashed line) does a much better job of estimating the trajectory. In the statistically poorly determined first metastable state, the estimate cannot track all fluctuations of the real trajectory, simply because they are not supported by any of the recorded data, but the mean and error estimates of x are in the correct range. In metastable states 2 and 3, the estimate follows the real trajectory and even detects the fast re-crossings missed out by the time window estimator. The maximum likelihood estimate gives a very similar estimate to the mean of the sampled Bayesian estimator and is not shown for clarity. For a more quantitative result consult Table (1) which shows that the mean square error of the estimation is about 5 times better for the mean of the Bayesian estimate than the time window estimate.

When dismissing the deterministic force from the potential, we obtain the free diffusion estimator (panel e). As for the estimation of the stationary distribution it is much better than any time-window based estimator, but is biased in estimating the fluctuations at very small or very large FRET efficiencies.

5.3 Single versus multiple trajectories

There are two common approaches to conduct FRET experiments: (1) immobilizing a few molecules and measuring their FRET signal for a long time (illustrated by the simulations above), and (2) letting them diffuse freely at a very low concentration and simply waiting until a molecule enters the measurement spot. The second method typically generates a similar total amount of data, but scattered across many more and much shorter trajectories. This situation is simulated here by using the trajectory data from above and cutting it into 3600 pieces of 0.025 s each (free diffusion trajectories are usually even shorter. 25 ms length can be obtained with microfluidic devices).

Fig. (3) compares the estimation results of the time-window and Bayesian estimators. The Bayesian estimate of the invariant density and the potential is now significantly better than the time window estimate, owing to the fact that the time window estimate looses a lot of information as a time segment of length $T_w/2$ is discarded from the beginning and end of each trajectory. Fig. 4 illustrates the quality of the trajectory estimation on a trajectory that transitions from metastable state 1 to 2. It is apparent that the Bayesian estimate is much smoother and closer to the real trajectory than the time window estimate. Moreover, the blue dashed line shows for comparison a Bayesian estimate that was constructed from this single trajectory only. The Bayesian estimate constructed from the complete data (solid blue line) is better, i.e. taking into account the other trajectories helps each individual estimate through the improved estimation of the joint dynamical model consisting of the potential and diffusion constant.

A main advantage of applying the present framework to data of this type is that the diffusion model obtained can be directly queried for long-time information that cannot directly be computed from the trajectories. To illustrate this, we have computed the transition rates between the three metastable states from the diffusion model, defined by the inverse mean passage times between the potential energy minima. These rates cannot be computed from the trajectory data directly, because the individual trajectories are usually too short to connect two metastable state minima. The results are shown in Table (2), showing that the transition rates are well estimated.

5.4 Equilibrium estimates from non-equilibrium measurement conditions

A particular advantage of estimating a dynamic model compared to directly evaluating the observed trajectory data is that it is in principle able to provide the equilibrium dynamical information of a

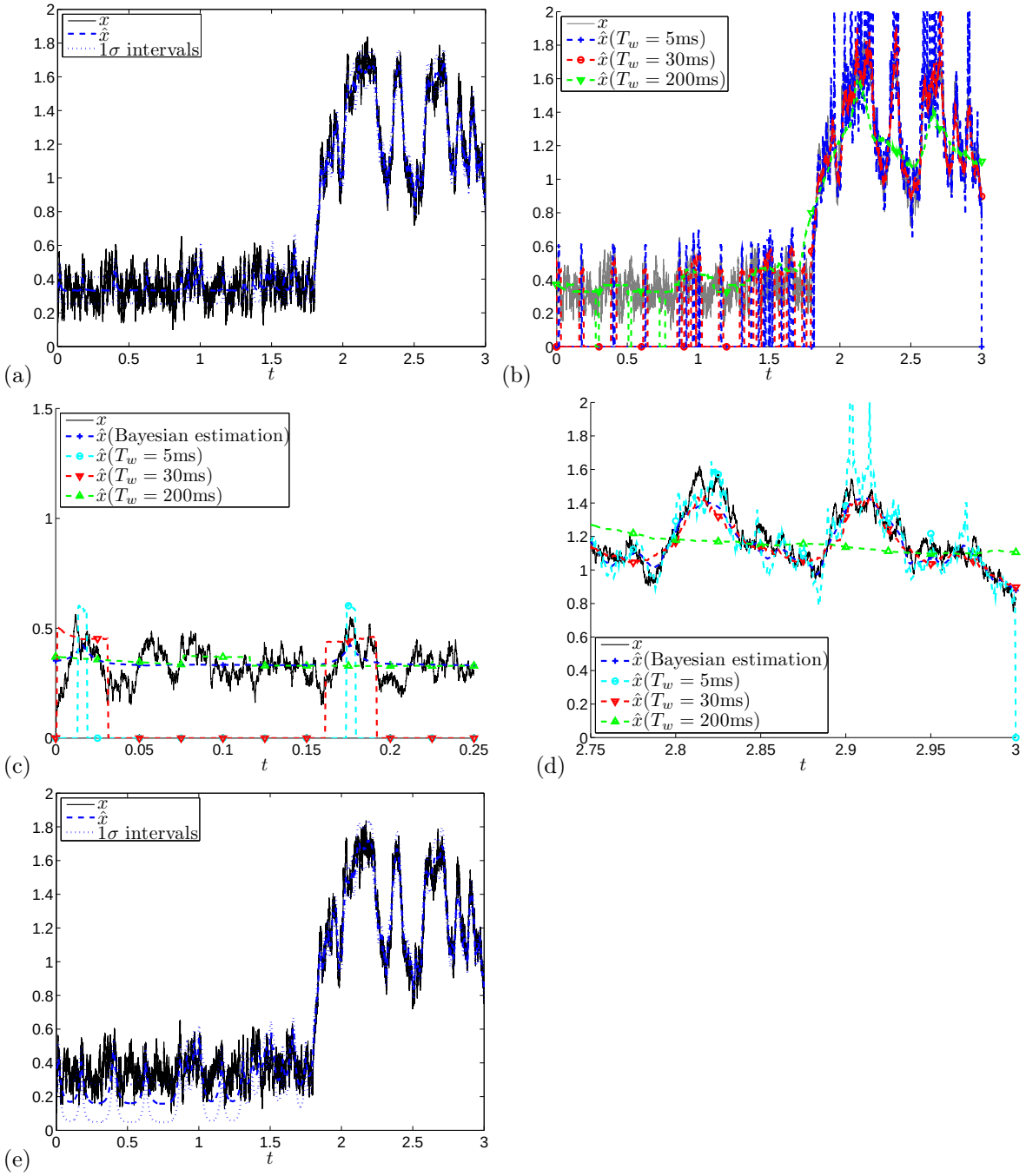

Figure 2: Trajectory estimation results from few long simulations comparing the real trajectory (solid line) with different estimation approached. (a) Bayesian estimation for the mean and the statistical uncertainty. (b) Estimation results using the standard time window binning approach for different window sizes. (c,d) Close comparison on two parts of the trajectory between the Bayesian estimate and time-window esimates. (e) Results of the Bayesian estimator when the potential is forced to be flat (free diffusion).

Table 2: Bayesian estimates of transition rates for the three-state system from many short trajectories

Transition	True Rate		Estimated Rate	
	mean	standard error	mean	standard error
$1 \to 2$	$1.5719\,\text{s}^{-1}$	$0.0306\,\text{s}^{-1}$	$1.1659\,\text{s}^{-1}$	$0.0217\,\text{s}^{-1}$
$2 \to 1$	$3.2010\,\text{s}^{-1}$	$0.0550\,\text{s}^{-1}$	$2.4656\,\text{s}^{-1}$	$0.0401\,\text{s}^{-1}$
$2 \to 3$	$6.5849\,\text{s}^{-1}$	$0.0800\,\text{s}^{-1}$	$5.0941\,\text{s}^{-1}$	$0.0611\,\text{s}^{-1}$
$3 \to 2$	$4.2004\,\text{s}^{-1}$	$0.0575\,\text{s}^{-1}$	$4.3455\,\text{s}^{-1}$	$0.0538\,\text{s}^{-1}$

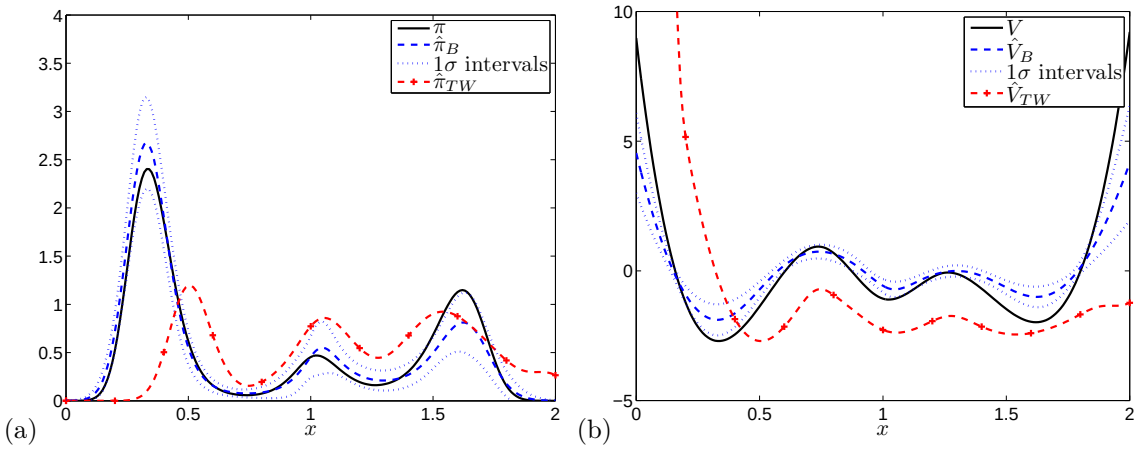

Figure 3: Stationary distribution and potential function results of artificial data from many short trajectories, comparing true values (solid line) with the time-window estimator using 20 ms window length and the Bayesian estimator

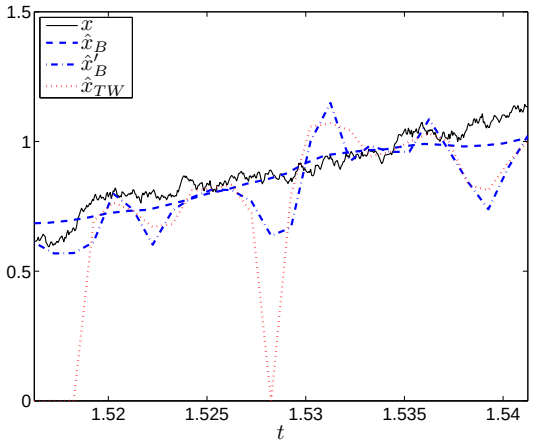

Figure 4: Trajectory estimation results of artificial data of a single out of many short trajectories, comparing the true trajectory (solid line) with the time-window estimator using 20 ms window length, the Bayesian estimator using all data, and the Bayesian estimator using this single trajectory only.

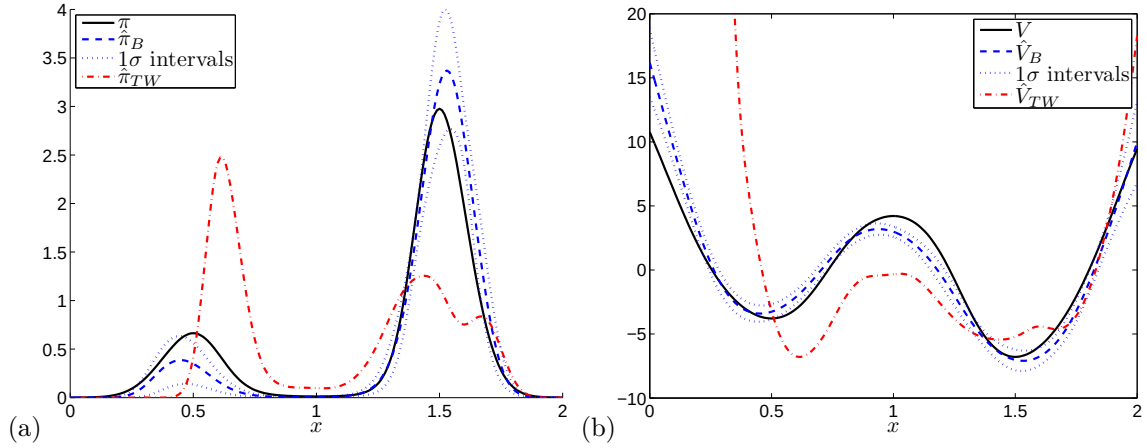

Figure 5: Stationary distribution and potential function results of artificial data from few long trajectories all started in the left potential energy minimum, comparing true values (solid line) with the time-window estimator and the Bayesian estimator

system even if the experiment was conducted under non-equilibrium conditions. For this, let us now investigate a potential with two metastable states defined by:

$$U(x) = \frac{\sum_{i=1}^{5} (|x - x_i| + 0.001)^{-2} z_i}{\sum_{i=1}^{5} (|x - x_i| + 0.001)^{-2}} \tag{27}$$

where $(x_1, \ldots, x_5) = (-0.3, 0.5, 1.0, 1.5, 2.3)$ and $(z_1, \ldots, z_5) = (21, 4, 8, -1, 20)$.This model has 2 potential wells centered at $\bar{r} = 0.5$ and 1.5, with the second well being deeper than the first one. The invariant density and potential are illustrated in Fig. 5 (black lines). Again, 30 trajectories of 3 s length were launched using the same simulation settings as above, but this time the initial positions were not generated from the invariant density but all started in the left metastable state at $x = 0.05$. While this precise setup is somewhat artificial, non-equilibrium starting conditions can indeed be found in experiments, for example if environmental parameters that influence the dynamical behavior, such as the Magnesium concentration, are varied faster than the global relaxation time of the system under study [50], in order to drive the system into visiting unlikely states.

Most of the trajectories are too short to cross the potential barrier and reach the second well during the observation time. This is visible from the time window estimate shown in Fig. 5 (red line) that is based on directly counting the number of occurrences of each value x. In this estimate, the stability of the first metastable state is strongly overestimated due to the predominance of trajectories in this state. The Bayesian estimate (solid blue line), on the other hand excellently approximates the true invariant density and potential within the error (dashed blue line).

6 Conclusion

We have introduced a Bayesian framework for estimating realizations and parameters of stochastic differential equations with nonlinear force terms from possibly incomplete, nonlinear and stochastic observations. The method has been demonstrated for single-molecule Förster energy resonance data and has proved to be useful even in cases where the estimation was difficult e.g. due to little statistics.

This framework will be useful for analyzing and interpreting data from various types of experiments or observations. A particular strength of estimating a diffusion model from experimental data over analyzing this data directly is that information can be reconstructed that is not directly computable from the data. We have demonstrated how to reconstruct very slow transition rates from very short trajectory data. Another property that is accessible with the model are transition path times in fluorescence-probed molecules that cannot be computed directly from the data due to the limited photon resolution [51].

Acknowledgements

The authors acknowledge funding from Deutsche Forschungsgemeinschaft (DFG) through research center Matheon, SFB 449 and grant NO 825/2.

Appendix

Modeling the nonlinear drift via the potential

A special case of modeling the nonlinear drift $\mathbf{f}(\mathbf{x})$ is in terms of a generating potential $U(\mathbf{x})$ with quadratic terms. This also generates a mix of locally linear force models as proposed in the main text, but generates only conservative vector fields $\mathbf{f}(\mathbf{x})$. For this, consider a potential that is a mix of potential minima and maxima,

$$U(\mathbf{x}) = \sum_{l=1}^{L} \mu_l(\mathbf{x}) \left(\frac{1}{2} (\mathbf{x} - \mathbf{x}_l)^{\mathrm{T}} \mathbf{A}_l (\mathbf{x} - \mathbf{x}_l) + b_l \right)$$

where $\mathbf{A}_l$ can be positive definite (minimum) or negative definite (maximum), and define the corresponding force:

$$\mathbf{f}(\mathbf{x}) = -\nabla V(\mathbf{x}) = -\sum_{l=1}^{L} \mu_l(\mathbf{x}) \, \mathbf{A}_l(\mathbf{x} - \mathbf{x}_l) - \sum_{l=1}^{L} \frac{1}{2} \nabla \mu_l(\mathbf{x}) \, (\mathbf{x} - \mathbf{x}_l)^{\mathrm{T}} \mathbf{A}_l (\mathbf{x} - \mathbf{x}_l) - \sum_{l=1}^{L} b_l \nabla \mu_l(\mathbf{x}).$$

This equation involves the definitions $\lambda = \left(a_1^{\mathrm{T}}, \ldots, a_L^{\mathrm{T}}, b_1, ..., b_L\right)^{\mathrm{T}}$,

$$\mathbf{\Xi}(\mathbf{x}) = \begin{bmatrix} -\mu_1(\mathbf{x}) I_d \otimes (\mathbf{x} - \mathbf{x}_1) - \frac{1}{2} \nabla \mu_1(\mathbf{x}) \left(I_d \otimes (\mathbf{x} - \mathbf{x}_1) \right) (\mathbf{x} - \mathbf{x}_1) \\ \vdots \\ -\mu_L(\mathbf{x}) I_d \otimes (\mathbf{x} - \mathbf{x}_L) - \frac{1}{2} \nabla \mu_L(\mathbf{x}) \left(I_d \otimes (\mathbf{x} - \mathbf{x}_L) \right) (\mathbf{x} - \mathbf{x}_L) \\ \nabla \mu_1(\mathbf{x}) \\ \vdots \\ \nabla \mu_L(\mathbf{x}) \end{bmatrix}^{\mathrm{T}}$$

$I_d \in \mathbb{R}^{d \times d}$ denotes the identity matrix, $a_l \in \mathbb{R}^{d^2}$ and the $(i-1)\,d + j$-th entry of a_l equals to the (i,j)-th entry of A_l. $\otimes$ here indicates the Kronecker product. A more direct discretization that is only based on support points of some potential values is obtained by using $a_l = 0$ for all l.

Metropolis-Hastings Step for x_n

Starting from the Euler discretization of the Smoluchowski equation, we are interested in the mid-point problem: $p(\mathbf{x}_n \mid \mathbf{x}_{n-1}, \mathbf{x}_{n+1})$ and sample $\mathbf{x}_n$ from it:

$$\begin{aligned} p(\mathbf{x}_n \mid \mathbf{x}_{n-1}, \mathbf{x}_{n+1}) &= \frac{p(\mathbf{x}_n, \mathbf{x}_{n-1}, \mathbf{x}_{n+1})}{p(\mathbf{x}_{n-1}, \mathbf{x}_{n+1})} \\ &\propto \frac{p(\mathbf{x}_n \mid \mathbf{x}_{n-1}) p(\mathbf{x}_{n+1} \mid \mathbf{x}_n)}{\int_{x_t} p(\mathbf{x}_n \mid \mathbf{x}_{n-1}) p(\mathbf{x}_{n+1} \mid \mathbf{x}_n)} \\ &\propto p(\mathbf{x}_n \mid \mathbf{x}_{n-1}) p(\mathbf{x}_{n+1} \mid \mathbf{x}_n) \\ &= \mathcal{N}(\mathbf{x}_{n-1} - \tau \mathbf{f}(\mathbf{x}_{n-1}) - \tau \mathbf{g}(\mathbf{x}_{n-1}), \Sigma) \mid_{\mathbf{x}_n} \\ &\quad \times \mathcal{N}(\mathbf{x}_n - \tau \mathbf{f}(\mathbf{x}) - \tau \mathbf{g}(\mathbf{x}), \Sigma) \mid_{\mathbf{x}_{n+1}} \end{aligned}$$

Now we can consider a MCMC step for the point $\mathbf{x}_n$:

$$p_{acc}(\mathbf{x}_n^*) = \min \left\{ 1, \frac{p(\mathbf{x}_n^* \to \mathbf{x}_n) \, p(Q_n \mid \mathbf{x}_n^*) \, p(\mathbf{x}_n^* \mid \mathbf{x}_{n-1}) p(\mathbf{x}_{n+1} \mid \mathbf{x}_n^*)}{p(\mathbf{x}_n \to \mathbf{x}_n^*) \, p(Q_n \mid \mathbf{x}_n) \, p(\mathbf{x}_n \mid \mathbf{x}_{n-1}) p(\mathbf{x}_{n+1} \mid \mathbf{x}_n)} \right\}$$

Now we simply *choose* the proposal step to be given by the system dynamics, i.e. $p(\mathbf{x}_n \to \mathbf{x}_n^*) = p(\mathbf{x}_n^* \mid \mathbf{x}_{n-1})$ and $p(\mathbf{x}_n^* \to \mathbf{x}_n) = p(\mathbf{x}_{n-1} \mid \mathbf{x}_n^*)$

Metropolis-Hastings Step for x_n based on a free diffusion model

An alternative to the above approach is to assume for a moment that $V = \text{const}$. In this case, we can write:

$$\begin{aligned}\mathbb{P}(\mathbf{x}_{n+1} \mid \mathbf{x}_n)\mathbb{P}(\mathbf{x}_n \mid \mathbf{x}_{n-1}) &= \mathcal{N}(\mathbf{x}_{n+1}|\mathbf{x}_n, \Sigma)\mathcal{N}(\mathbf{x}_n|\mathbf{x}_{n-1}, \Sigma) \\ &\propto \exp(\frac{1}{2}(\mathbf{x}_n - \mathbf{x}_{n+1})\Sigma^{-1}(\mathbf{x}_n - \mathbf{x}_{n+1})^{\mathrm{T}}) \\ &\times \exp(\frac{1}{2}(\mathbf{x}_n - \mathbf{x}_{n-1})\Sigma^{-1}(\mathbf{x}_n - \mathbf{x}_{n-1})^{\mathrm{T}}) \\ &\propto \mathcal{N}(\mathbf{x}_n|\frac{1}{2}(\mathbf{x}_{n+1} + \mathbf{x}_{n-1}), \frac{1}{2}\Sigma)\end{aligned}$$

where $N(\mathbf{x}|\mu, \Sigma)$ denotes the probability density function of Gaussian distribution with mean value μ and covariance matrix Σ, and use this as a proposal step $p(\mathbf{x}_n \to \mathbf{x}_n^*)$.

$$p_{acc}(\mathbf{x}_n^*) = \min\left\{1, \frac{\mathcal{N}(\frac{1}{2}(\mathbf{x}_{n+1} + \mathbf{x}_{n-1}), \frac{1}{2}\Sigma)\mid_{\mathbf{x}_n}\ p(Q_n \mid \mathbf{x}_n^*)\, p(\mathbf{x}_n^* \mid \mathbf{x}_{n-1})p(\mathbf{x}_{n+1} \mid \mathbf{x}_n^*)}{\mathcal{N}(\frac{1}{2}(\mathbf{x}_{n+1} + \mathbf{x}_{n-1}), \frac{1}{2}\Sigma)\mid_{\mathbf{x}_n^*}\ p(Q_n \mid \mathbf{x}_n)\, p(\mathbf{x}_n \mid \mathbf{x}_{n-1})p(\mathbf{x}_{n+1} \mid \mathbf{x}_n)}\right\}$$

Posterior Probability of Parameters θ

The likelihood of θ for a given trajectory $\mathbf{x}_{0\ldots N}$ is given by:

$$p(\mathbf{x}_{0:N}|\theta) \propto |\Sigma|^{-\frac{N}{2}} \exp\left(-\frac{1}{2\tau}\sum_{n=0}^{N-1}\left(\mathbf{x}_{n+1} - \mathbf{x}_n - \tau\mathbf{g}(\mathbf{x}_n) + \tau\Xi(\mathbf{x}_n)\lambda\right)^{\mathrm{T}}\Sigma^{-1}\left(\mathbf{x}_{n+1} - \mathbf{x}_n - \tau\mathbf{g}(\mathbf{x}_n) + \tau\Xi(\mathbf{x}_n)\lambda\right)\right)$$

If Σ is known,

$$\begin{aligned}p\left(\lambda|\mathbf{x}^1, \ldots, \mathbf{x}^K, \Sigma\right) &\propto p(\lambda)\, p\left(\mathbf{x}^1, \ldots, \mathbf{x}^K|\lambda, \Sigma\right) \\ &\propto \exp\left(-\frac{1}{2}\lambda^T\Sigma_0^{-1}\lambda\right)\exp\left(-\frac{1}{2}\left(\lambda^T\mathbf{A}_s\lambda + \mathbf{B}_s\lambda\right)\right) \\ &\propto \exp\left(-\frac{1}{2}\left(\lambda^T\left(\Sigma_0^{-1} + \mathbf{A}_s\right)\lambda + \mathbf{B}_s\lambda\right)\right) \\ &\propto \exp\left(-\frac{1}{2}\left(\left(\lambda - \tilde{\lambda}\right)^{\mathrm{T}}\left(\Sigma_0^{-1} + \mathbf{A}_s\right)\left(\lambda - \tilde{\lambda}\right)\right)\right) \qquad (28)\end{aligned}$$

where we have defined the functionals:

$$\mathbf{A}_s \stackrel{\text{def}}{=} \tau\sum_{k=1}^{K}\sum_{n=0}^{N-1}\Xi\left(\mathbf{x}_n^k\right)^{\mathrm{T}}\Sigma^{-1}\Xi\left(\mathbf{x}_n^k\right) \qquad (29)$$

$$\mathbf{B}_s \stackrel{\text{def}}{=} 2\sum_{k=1}^{K}\sum_{n=0}^{N-1}\left(\mathbf{x}_{n+1}^k - \mathbf{x}_n^k - \tau\mathbf{g}(\mathbf{x}_n)\right)^{\mathrm{T}}\Sigma^{-1}\Xi\left(\mathbf{x}_n^k\right) \qquad (30)$$

$$\tilde{\lambda} \stackrel{\text{def}}{=} -\frac{1}{2}\left(\Sigma_0^{-1} + \mathbf{A}_s\right)^{-1}\mathbf{B}_s^{\mathrm{T}} \qquad (31)$$

Thus λ can be sampled as

$$\lambda \mid \Sigma, \mathbf{x}^1, \ldots, \mathbf{x}^K \sim \mathcal{N}\left(\tilde{\lambda}, \left(\Sigma_0^{-1} + \mathbf{A}_s\right)^{-1}\right) \qquad (32)$$

If λ is known,

$$\begin{aligned}p\left(\Sigma|\mathbf{x}^1, \ldots, \mathbf{x}^K, \lambda\right) &\propto p(\Sigma)\, p\left(\mathbf{x}^1, \ldots, \mathbf{x}^K|\lambda, \Sigma\right) \\ &\propto |\Sigma|^{-\frac{m_0+d+1}{2}}\exp\left(-\frac{1}{2}\mathrm{trace}\left(\Psi_0\Sigma^{-1}\right)\right)|\Sigma|^{-\frac{NK}{2}}\exp\left(-\frac{1}{2}\mathrm{trace}\left(\mathbf{C}_s\Sigma^{-1}\right)\right) \\ &= |\Sigma|^{-\frac{m_0+d+NK+1}{2}}\exp\left(-\frac{1}{2}\mathrm{trace}\left(\left(\Psi_0 + \mathbf{C}_s\right)\Sigma^{-1}\right)\right) \qquad (33)\end{aligned}$$

where

$$\mathbf{C}_s \stackrel{\text{def}}{=} \frac{1}{\tau}\sum_{k=1}^{K}\sum_{n=0}^{N-1}\left(\mathbf{x}_{n+1}^k - \mathbf{x}_n^k - \tau\mathbf{g}(\mathbf{x}_n^k) + \tau\Xi\left(\mathbf{x}_n^k\right)\lambda\right)\left(\mathbf{x}_{n+1}^k - \mathbf{x}_n^k - \tau\mathbf{g}(\mathbf{x}_n^k) + \tau\Xi\left(\mathbf{x}_n^k\right)\lambda\right)^{\mathrm{T}} \qquad (34)$$

Thus Σ can be sampled as

$$\Sigma \mid \lambda, \mathbf{x}^1, \ldots, \mathbf{x}^K \sim \mathcal{W}^{-1}\left(\Psi_0 + \mathbf{C}_s, m_0 + NK\right) \qquad (35)$$

Maximum Probability / maximum likelihood estimation

In this subsection we consider the problem of estimating the parameters θ through maximizing the penalized log-likelihood [52]

$$L_p(\theta) = \log p(Q|\theta) + D(\theta) = \sum_{k=1}^{K} \log p\left(Q^k|\theta\right) + D(\theta) \tag{36}$$

Here a penalized term $D(\theta)$ is added to the log-likelihood function as a regularizer. Note that

$$p\left(Q^k|\theta\right) = \int p\left(Q^k|\mathbf{x}_{0:N}^k\right) p\left(\mathbf{x}_{0:N}^k|\theta\right) \mathrm{d}\mathbf{x}_{0:N}^k \tag{37}$$

Therefore, if we chose the logarithm of the prior for θ as the penalty function, then the maximum penalized likelihood (MPL) estimate is equivalent to the maximum a posterior (MAP) estimate and may efficiently be computed with the expectation maximization (EM) algorithm [53], which consists of the iterative application of the following steps:

1. **E-step**: Compute the functional

$$\begin{aligned} Q_{MPL}(\theta|\theta_j) &= \mathbb{E}\left[\sum_{k=1}^{K} \log p\left(\mathbf{x}_{0:N}^k|\theta\right) + \log p(\theta)\,|\theta_j, Q\right] \\ &= \sum_{k=1}^{K} \mathbb{E}\left[\log p\left(\mathbf{x}_{0:N}^k|\theta\right)|\theta_j, Q\right] + \log p(\theta) \end{aligned} \tag{38}$$

2. **M-step**: Update

$$\theta_{j+1} = \arg\max_{\theta} Q_{MPL}(\theta|\theta_j) \tag{39}$$

We cannot solve (38) analytically, but it is possible to get a suboptimal solution by

$$\lambda_{j+1} = \arg\max_{\lambda} Q_{MPL}(\lambda, \Sigma_j|\theta_j) \tag{40}$$

$$\Sigma_{j+1} = \arg\max_{\Sigma} Q_{MPL}(\lambda_{j+1}, \Sigma|\theta_j) \tag{41}$$

Using Eq. ((28) and (33)), we can conclude that

$$Q_{MPL}(\lambda, \Sigma_j|\theta_j) = -\frac{1}{2}\left(\lambda^T\left(\Sigma_0^{-1} + \mathbf{A}_j\right)\lambda + \mathbf{B}_j\lambda\right) + C_\lambda \tag{42}$$

$$Q_{MPL}(\lambda_{j+1}, \Sigma|\theta_j) = -\frac{m_0 + d + NK + 1}{2}\log|\Sigma| - \frac{1}{2}\mathrm{trace}\left(\left(\Psi_0 + \mathbf{C}_j\right)\Sigma^{-1}\right) + C_\Sigma \tag{43}$$

where

$$\mathbf{A}_j \overset{\text{def}}{=} \mathbb{E}\left[\tau\sum_{k=1}^{K}\sum_{n=0}^{N-1}\Xi\left(\mathbf{x}_n^k\right)^{\mathrm{T}}\Sigma_j^{-1}\Xi\left(\mathbf{x}_n^k\right)|\theta_j, Q\right] \tag{44}$$

$$\mathbf{B}_j \overset{\text{def}}{=} \mathbb{E}\left[2\sum_{k=1}^{K}\sum_{n=0}^{N-1}\left(\mathbf{x}_{n+1}^k - \mathbf{x}_n^k - \tau\mathbf{g}(\mathbf{x}_n)\right)^{\mathbf{T}}\Sigma_j^{-1}\Xi\left(\mathbf{x}_n^k\right)|\theta_j, Q\right] \tag{45}$$

$$\begin{aligned} \mathbf{C}_j \overset{\text{def}}{=} \mathbb{E}\Bigg[&\frac{1}{\tau}\sum_{k=1}^{K}\sum_{n=0}^{N-1}\left(\mathbf{x}_{n+1}^k - \mathbf{x}_n^k - \tau g(\mathbf{x}_n^k) + \tau\Xi\left(\mathbf{x}_n^k\right)\lambda_{j+1}\right) \qquad (46) \\ &\times\left(\mathbf{x}_{n+1}^k - \mathbf{x}_n^k - \tau\mathbf{g}(\mathbf{x}_n^k) + \tau\Xi\left(\mathbf{x}_n^k\right)\lambda_{j+1}\right)^{\mathrm{T}}|\theta_j, Q\Bigg] \qquad (47) \end{aligned}$$

C_λ is independent of λ and C_Σ is independent of Σ. By setting $\partial Q_{MPL}(\lambda, \Sigma_j|\theta_j)/\partial\lambda = 0$ we can get the solution of (40) as following:

$$\lambda_{j+1} = -\frac{1}{2}\left(\Sigma_0^{-1} + \mathbf{A}_j\right)^{-1}\mathbf{B}_j^{\mathrm{T}} \tag{48}$$

And Σ_{j+1} should be the mode of the distribution $\mathcal{W}^{-1}(\Psi_0 + \mathbf{C}_j, m_0 + NK)$, so

$$\Sigma_{j+1} = \frac{1}{m_0 + d + NK + 1}\left(\Psi_0 + \mathbf{C}_j\right) \tag{49}$$

And the conditional expectations in (48) and (49) can be approximated by sampling the path variables $\mathbf{x}_n$ as described in Subsection 3.1. The complete EM algorithm then becomes what is described by Algorithm 2.

Algorithm 2 EM Estimator for θ

(**Initialization**)
set the values of θ_1 arbitrarily
(**Iteration**)
for $j = 2$ to J
  (**E-step**)
  for $k = 1$ to K
    perform the Gibbs sampler to get the Monte Carlo estimate of $p\left(\mathbf{x}_{0:N}^k | \theta_j, Q^k\right)$
  end for
  calculate $\mathbf{A}_j$, $\mathbf{B}_j$, $\mathbf{C}_j$ by the Monte Carlo estimates
  (**M-step**)
compute θ_{j+1} by (48) and (49)
end for

Remark 1 *Obviously, the maximum likelihood (ML) estimate is a special case of MPL estimate with a flat prior* $p\left(\bar{\theta}\right) \propto 1$. *And for this case, the update equations are*

$$\lambda_{j+1} = -\frac{1}{2}\mathbf{A}_j^{-1}\mathbf{B}_j^{\mathrm{T}} \tag{50}$$

$$\Sigma_{j+1} = \frac{1}{NK}\mathbf{C}_j \tag{51}$$

But in practice, the estimated $\mathbf{A}_j$ *may be nearly singular, which will make the EM algorithm computationally unfeasible. For this reason, a prior such as the one introduced above is essential in practice.*

References

[1] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*. Elsevier, Amsterdam, 4th ed., 2006.

[2] K. Anguige and C. Schmeiser, "A one-dimensional model of cell diffusion and aggregation, incorporating volume filling and cell-to-cell adhesion.," *Journal of mathematical biology*, vol. 58, pp. 395–427, Mar. 2009.

[3] A. Spaar, C. Dammer, R. R. Gabdoulline, R. C. Wade, and V. Helms, "Diffusional encounter of barnase and barstar," *Biophys J*, vol. 90, pp. 1913–1924, 2006.

[4] R. B. Best and G. Hummer, "Coordinate-dependent diffusion in protein folding.," *Proc. Natl. Acad. Sci. USA*, vol. 107, pp. 1088–1093, Jan. 2010.

[5] D. L. McLeish and A. W. Kolkiewicz, *Fitting Diffusion Models in Finance*, vol. 32. Institute of Mathematical Statistics, 1997.

[6] J. D. Pelletier, "A Stochastic Diffusion Model of Climate Change," *arXiv:ao-sci/9510001*, Oct. 1995.

[7] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, "Imaging Intracellular Fluorescent Proteins at Nanometer Resolution," *Science*, vol. 313, pp. 1642–1645, Sept. 2006.

[8] S. W. Hell and J. Wichmann, "Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy," *Opt. Lett.*, vol. 19, pp. 780–782, June 1994.

[9] W. E. Moerner and M. Orrit, "Illuminating Single Molecules in Condensed Matter," *Science*, vol. 283, pp. 1670–1676, Mar. 1999.

[10] T. Schmidt, G. J. Schütz, W. Baumgartner, H. J. Gruber, and H. Schindler, "Imaging of single molecule diffusion.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, pp. 2926–2929, Apr. 1996.

[11] H. Uji-i, S. M. Melnikov, A. Deres, G. Bergamini, F. De Schryver, A. Herrmann, K. Müllen, J. Enderlein, and J. Hofkens, "Visualizing spatial and temporal heterogeneity of single molecule rotational diffusion in a glassy polymer by defocused wide-field imaging," *Polymer*, vol. 47, pp. 2511–2518, Mar. 2006.

[12] A. Ashkin, J. M. Dziedzic, and T. Yamane, "Optical trapping and manipulation of single cells using infrared laser beams," *Nature*, vol. 330, pp. 769–771, Dec. 1987.

[13] B. Schuler, "Single-Molecule Fluorescence Spectroscopy of Protein Folding," *ChemPhysChem*, vol. 6, no. 7, pp. 1206–1220, 2005.

[14] B. Schuler, E. A. Lipman, and W. A. Eaton, "Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy," *Nature*, vol. 419, pp. 743–747, Oct. 2002.

[15] H. D. Kim, G. U. Nienhaus, T. Ha, J. W. Orr, J. R. Williamson, and S. Chu, "Mg2+-dependent conformational change of RNA studied by fluorescence correlation and FRET on immobilized single molecules," *Procl. Natl. Acad. Sci. USA*, vol. 99, pp. 4284–4289, Apr. 2002.

[16] X. Michalet, S. Weiss, and M. Jäger, "Single-Molecule Fluorescence Studies of Protein Folding and Conformational Dynamics," *Chem. Rev.*, vol. 106, pp. 1785–1813, 2006.

[17] H. Neuweiler, M. Löllmann, S. Doose, and M. Sauer, "Dynamics of Unfolded Polypeptide Chains in Crowded Environment Studied by Fluorescence Correlation Spectroscopy," *J. Mol. Biol.*, vol. 365, pp. 856–869, 2007.

[18] G. Binnig, C. F. Quate, and Ch, "Atomic Force Microscope," *Phys. Rev. Lett.*, vol. 56, pp. 930–933, Mar. 1986.

[19] A. Ashkin, J. M. Dziedzic, J. E. Bjorkholm, and S. Chu, "Observation of a single-beam gradient force optical trap for dielectric particles," *Opt. Lett.*, vol. 11, pp. 288–290, May 1986.

[20] O. Elerian, S. Chib, and N. Shephard, "Likelihood Inference for Discretely Observed Nonlinear Diffusions," *Econometrica*, vol. 69, no. 4, pp. 959–993, 2001.

[21] A. Beskos, O. Papaspiliopoulos, G. O. Roberts, and P. Fearnhead, "Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion)," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, pp. 333–382, June 2006.

[22] H. Sorensen, "Parametric Inference for Diffusion Processes Observed at Discrete Points in Time: a Survey," *Internat. Statist. Rev.*, vol. 72, pp. 337–354, 2004.

[23] I. V. Gopich, D. Nettels, B. Schuler, and A. Szabo, "Protein dynamics from single-molecule fluorescence intensity correlation functions," *The Journal of Chemical Physics*, vol. 131, no. 9, pp. 095102+, 2009.

[24] S. C. Kou, X. Sunney Xie, and J. S. Liu, "Bayesian analysis of single-molecule experimental data," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 54, pp. 469–506, June 2005.

[25] G. Schröder and H. Grubmüller, "Maximum likelihood trajectories from single molecule fluorescence resonance energy transfer experiments," *The Journal of Chemical Physics*, vol. 119, p. 9920, 2003.

[26] I. Horenko, E. Dittmer, A. Fischer, and Ch, "Automated Model Reduction for Complex Systems exhibiting Metastability," *Multiscale Model. Sim.*, vol. 5, pp. 802–827, 2006.

[27] G. Roberts and O. Stramer, "On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm," *Biometrika*, vol. 88, no. 3, p. 603, 2001.

[28] G. Durham and A. Gallant, "Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes," *Journal of Business & Economic Statistics*, vol. 20, no. 3, pp. 297–316, 2002.

[29] S. Donnet and A. Samson, "Parametric Estimation for Diffusion Processes from Discrete-time and Noisy Observations," tech. rep., INRIA, 2005.

[30] P. Dellaportas, N. Friel, and G. Roberts, "Bayesian model selection for partially observed diffusion models," *Biometrika*, vol. 93, no. 4, p. 809, 2006.

[31] A. Beskos, O. Papaspiliopoulos, G. Roberts, and P. Fearnhead, "Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes," *Journal of the Royal Statistical Society*, vol. 68, no. 3, pp. 333–382, 2006.

[32] K. Kalogeropoulos, "Likelihood-based inference for a class of multivariate diffusions with unobserved paths," *Journal of Statistical Planning and Inference*, vol. 137, no. 10, pp. 3092–3102, 2007.

[33] A. Golightly and D. Wilkinson, "Bayesian inference for nonlinear multivariate diffusion models observed with error," *Computational Statistics and Data Analysis*, vol. 52, no. 3, pp. 1674–1693, 2008.

[34] B. Bibby, M. Jacobsen, and M. Sørensen, "Estimating functions for discretely sampled diffusion-type models," *Handbook of financial econometrics*, 2004.

[35] C. Archambeau, M. Opper, Y. Shen, D. Cornford, and J. Shawe-Taylor, "Variational inference for diffusion processes," *Advances in Neural Information Processing Systems*, vol. 20, pp. 17–24, 2007.

[36] L. Panzar and H. van Zanten, "Nonparametric Bayesian inference for ergodic diffusions," *Journal of Statistical Planning and Inference*, vol. 139, no. 12, pp. 4204–4210, 2009.

[37] Y. Pokern, O. Papaspiliopoulos, G. Roberts, and A. Stuart, "Nonparametric Bayesian drift estimation for one-dimensional diffusion processes,"

[38] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*. Amsterdam: Elsevier, 1989.

[39] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 15, no. 1, pp. 116–132, 1985.

[40] H. Ying, "General Takagi-Sugeno fuzzy systems are universal approximators," in *Fuzzy Systems Proceedings, IEEE International Conference on Fuzzy Systems in IEEE World Congress on Computational Intelligence*, vol. 1, 1998.

[41] W. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC, 1996.

[42] J. Ross, P. Buschkamp, D. Fetting, A. Donnermeyer, C. M. Roth, and P. Tinnefeld, "Multicolor Single-Molecule Spectroscopy with Alternating Laser Excitation for the Investigation of Interactions and Dynamics," *The Journal of Physical Chemistry B*, vol. 111, pp. 321–326, Jan. 2007.

[43] T.-Y. Kim, H. Uji-i, M. Moeller, B. Muls, J. Hofkens, and U. Alexiev, "Monitoring the Interaction of a Single G-Protein Key Binding Site with Rhodopsin Disk Membranes upon Light Activation," *Biochemistry*, vol. 48, pp. 3801–3803, May 2009.

[44] X. Zhuang, L. Bartley, H. Babcock, R. Russell, T. Ha, D. Herschlag, and S. Chu, "A single-molecule study of RNA catalysis and folding," *Science*, vol. 288, no. 5473, pp. 2048–2051, 2000.

[45] R. Picard and R. Cook, "Cross-validation of regression models," *Journal of the American Statistical Association*, pp. 575–583, 1984.

[46] A. Bowman and A. Azzalini, *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations*. Oxford University Press, USA, 1997.

[47] J. Garcıá and F. Lazaro, "Langevin-dynamics study of the dynamical properties of small magnetic particles," *Physics Review B*, vol. 58, p. 14937, 1998.

[48] A. Y. Kobitski, A. Nierth, M. Helm, A. Jäschke, and G. U. Nienhaus, "Mg2+ dependent folding of a Diels-Alderase ribozyme probed by single-molecule FRET analysis," *Nucleic Acids Res.*, vol. 35, pp. 2047–2059, 2007.

[49] G. F. Schröder and H. Grubmüller, "Maximum likelihood trajectories from single molecule fluorescence resonance energy transfer experiments," *J. Chem. Phys.*, vol. 119, pp. 9920–9924, 2003.

[50] X. Qu, G. J. Smith, K. T. Lee, T. R. Sosnick, T. Pan, and N. F. Scherer, "Single-molecule nonequilibrium periodic Mg2+-concentration jump experiments reveal details of the early folding pathways of a large RNA," *Proceedings of the National Academy of Sciences*, vol. 105, pp. 6602–6607, May 2008.

[51] H. S. Chung, J. M. Louis, and W. A. Eaton, "Experimental determination of upper bound for transition path times in protein folding from single-molecule photon-by-photon trajectories," *Proceedings of the National Academy of Sciences*, vol. 106, pp. 11837–11844, July 2009.

[52] D. Ormoneit and V. Tresp, "Averaging, maximum penalized likelihood and Bayesian estimation forimproving Gaussian mixture probability density estimates," *IEEE Transactions on Neural Networks*, vol. 9, no. 4, pp. 639–650, 1998.

[53] A. Dempster, N. Laird, D. Rubin, *et al.*, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society*, vol. 39, no. 1, pp. 1–38, 1977.