Variational Inference for Continuous-Time Switching Dynamical Systems

Anonymous Author(s)
Affiliation
Address
email

Abstract

Switching dynamical systems provide a powerful, interpretable modeling framework for inference in time-series data in, e.g., the natural sciences or engineering applications. Since many areas, such as biology or discrete-event systems, are naturally described in continuous time, we present a model based on an Markov jump process modulating a subordinated diffusion process. We provide the exact evolution equations for the prior and posterior marginal densities, the direct solutions of which are however computationally intractable. Therefore, we develop a new continuous-time variational inference algorithm, combining a Gaussian process approximation on the diffusion level with posterior inference for Markov jump processes. By minimizing the path-wise Kullback-Leibler divergence we obtain (i) Bayesian latent state estimates for arbitrary points on the real axis and (ii) point estimates of unknown system parameters, utilizing variational expectation maximization. We extensively evaluate our algorithm under the model assumption and for real-world examples.

1 Introduction

Many natural and engineered dynamical systems can be understood in terms of continuous-discrete hybrid models, in which a given system switches between discrete modes exhibiting continuous dynamics. Examples include neuro-mechanical models of locomotion in neuroscience [1], state-dependent volatility dynamics in financial analysis [2] or electric power systems [3], and phenotype differentiation in systems biology [4].

In a discrete-time setting, a widely used class of stochastic hybrid models are switching linear dynamical system (SLDS) [5], which have received considerable attention in recent years [6, 7, 8, 9]. Since real-world physical and biological systems naturally evolve in continuous time, a discrete-time description of such systems is however limiting. In biological experiments and discrete-event systems in engineering [10], for instance, one typically (i) is interested in the system behavior at any given point in time and (ii) can not easily determine an appropriate time discretization. To overcome these limitations, a continuous-time analog to SLDS models has been put forward termed switching stochastic differential equations (SSDEs) [11], which augment a set of diffusion processes with an underlying Markov jump process (MJP). Hybrid models of this kind have a long tradition in statistics [12] and have been analyzed in particular in applications to biological systems [13].

Diffusion processes and MJPs have been treated extensively in the literature [14, 15]. For each process class individually, exact expressions for the posterior paths given some set of observations can be obtained [16]. However, for diffusion processes in particular, these expressions quickly become intractable as they entail solving multi-dimensional partial differential equations (PDEs). This is aggravated if the diffusion is coupled to an underlying jump process, because both processes then have to be solved jointly.

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.

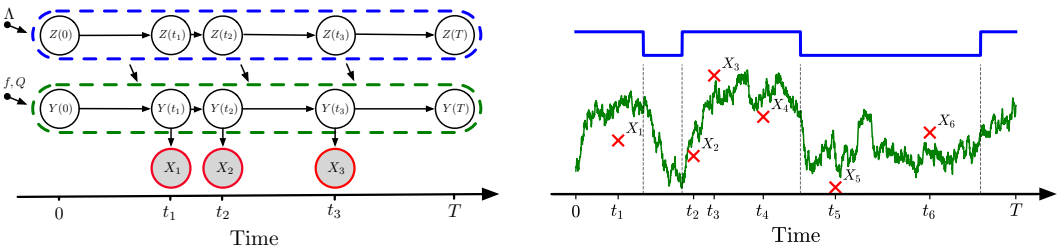

Figure 1: Sketch of the three layers of the hybrid process model. A two-state MJP $Z(t)$ (blue) evolves freely in the time interval $t \in [0, T]$, see Eq. (1) . The MJP controls the dynamics of the SSDE $Y(t) \mid Z(t)$ (green), see Eq. (2). From these continuous dynamics, only noisy observations $X_1, X_2, \dots$ (red) are available for inference at irregularly-spaced time points $t_1, t_2, \dots$. Left: Graphical model. Right: Sample path (vertical dashed lines indicate the Z-transitions).

One way to circumvent this issue are Monte Carlo approaches, which have been devised for both diffusion [17] and jump processes [18]. Sampling methods may however suffer from slow convergence, and, as we shall show, still face computational intractabilities for hybrid systems. An established approach avoiding these problems are variational inference (VI) methods, which approximate the exact posteriors by optimization [19]. For inference and parameter learning in diffusion processes, continuous-time VI frameworks have been developed utilizing, e.g., Gaussian processes (GPs) [20, 21], and general exponential family distributions [22]. Similar methods have also been devised for inference in MJPs [23, 24]. To the best of our knowledge, an inference framework for continuous-time hybrid processes is however lacking. We draw on these previous works and present a generalized VI framework for hybrid systems which recovers existing diffusion and MJP approximations as special cases. We specifically focus on meta-stable systems which remain in distinct, qualitatively different regimes over extended periods of time, which are of special interest, e.g., in computational structural biology [25]. An implementation of our proposed method is publicly available[1].

2 Mathematical Background

2.1 The Model

In this work, we consider three joint stochastic processes $\{Z(t)\}_{t\geq 0}$, $\{Y(t)\}_{t\geq 0}$, and $\{X_i\}_{i\in\mathbb{N}}$. A continuous-time adaptation of a probabilistic graphical model and a realization of the processes is shown in Fig. 1.

The Switching Process. The discrete-valued process $Z(t) \in \mathcal{Z} \subseteq \mathbb{N}$ is given as a latent Markov jump process (MJP) freely evolving in time t. An MJP is a continuous-time Markov process [26] on a countable state space $\mathcal{Z}$ and is completely characterized by an initial probability $p_0(z) := \mathbb{P}(Z(0) = z)$, $\forall z \in \mathcal{Z}$ and a transition rate function

$$\Lambda(z, z', t) = \lim_{h \searrow 0} \frac{\mathbb{P}(Z(t+h) = z' \mid Z(t) = z)}{h} \tag{1}$$

for $z' \in \mathcal{Z} \setminus z$, with the *exit rate* $\Lambda(z, t) := \sum_{z' \in \mathcal{Z}\setminus z} \Lambda(z, z', t)$.

The Subordinated Diffusion Process. The freely evolving MJP controls the dynamics of a continuous-valued process $Y(t) \in \mathcal{Y} \subseteq \mathbb{R}^n$, which is given as a latent switching stochastic differential equation (SSDE) in an Itô sense, [11]

$$\mathrm{d}Y(t) = f(Y(t), Z(t), t)\,\mathrm{d}t + Q(Y(t), Z(t), t)\,\mathrm{d}W(t), \tag{2}$$

where $f : \mathcal{Y} \times \mathcal{Z} \times \mathbb{R}_{\geq 0} \to \mathcal{Y}$ is an arbitrary drift function, the dispersion $Q : \mathcal{Y} \times \mathcal{Z} \times \mathbb{R}_{\geq 0} \to \mathbb{R}^{n\times m}$ determines the noise characteristics of the system and $W(t) \in \mathbb{R}^m$ is a standard Brownian motion. We define the system noise covariance $D(Y(t), Z(t), t)$ as $D := QQ^\top$. Note that the only difference to a conventional stochastic differential equation (SDE) is the $Z(t)$-dependence of the drift function and dispersion; hence, Eq. (2) can be understood as a collection of individual SDEs between which the systems switches via $Z(t)$. For an accessible introduction to SDEs, see [27].

[1] A link to the implementation will be inserted here in the camera-ready version of the paper.

The Hybrid Process. In the following, we refer to the continuous value $Y(t)$ as the *state* and to the discrete value $Z(t)$ as the *mode* of a hybrid process. The hybrid process $\{Z(t), Y(t)\}_{t\geq 0}$ is fully characterized by its time-point wise marginal density $p(y, z, t) := \partial_{y_1} \cdots \partial_{y_n} \mathbb{P}(Y(t) \leq y, Z(t) = z)$, where "$\leq$" has to be interpreted element-wise. This density is given as the solution to the hybrid master equation (HME) [28, 29]

$$\partial_t p(y, z, t) = \mathcal{A} p(y, z, t), \tag{3}$$

with initial condition $p(y, z, 0) = p_0(z) p_0(y, z)$, where $p_0(y, z) := \partial_{y_1} \cdots \partial_{y_n} \mathbb{P}(Y(0) \leq y \mid Z(0) = z)$ denotes the initial density of the Y-process and $p(z, 0) = p_0(z)\ \forall z \in \mathcal{Z}$ the initial probability mass function of the Z-process. The operator $\mathcal{A}(\cdot) = \mathcal{F}(\cdot) + \mathcal{T}(\cdot)$ is given via

$$\mathcal{F}\phi(y, z, t) := -\sum_{i=1}^{n} \partial_{y_i} \{f_i(y, z, t)\phi(y, z, t)\} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \partial_{y_i} \partial_{y_j} \{D_{ij}(y, z, t)\phi(y, z, t)\}$$

$$\mathcal{T}\phi(y, z, t) := \sum_{z' \in \mathcal{Z} \setminus z} \Lambda(z', z, t)\phi(y, z', t) - \Lambda(z, t)\phi(y, z, t),$$

for an arbitrary test function $\phi : \mathbb{R}^n \times \mathcal{Z} \times \mathbb{R}_{\geq 0} \to \mathbb{R}$. We provide a detailed derivation in Appendix A.1. As the discrete process $Z(t)$ is independent of $Y(t)$, we further obtain the dynamics of the marginal distribution $p(z, t) := \mathbb{P}(Z(t) = z)$ by integrating over the continuous variable Y in the HME. This recovers the traditional master equation [26]

$$\frac{\mathrm{d}}{\mathrm{d}t} p(z, t) = \sum_{z' \in \mathcal{Z} \setminus z} \Lambda(z', z, t) p(z', t) - \Lambda(z, t) p(z, t), \tag{4}$$

with $p(z, 0) = p_0(z)\ \forall z \in \mathcal{Z}$; for details, see Appendix A.1.1.

Note that a general, analytical solution to the PDE (3) does not exist. Numerical solvers utilizing schemes such as the finite differences or finite element method suffer from the curse of dimensionality and can in principle only be applied to very low-dimensional state spaces [30] . Even in low dimensions however, solvers need to be adapted to the problem at hand and may struggle due to, e.g., slow step-size adaptation. On the other hand, sampling trajectories from a hybrid process $\{Y(t), Z(t)\}$ is straightforward: One can draw the process $Z(t)$ by utilizing the Doob-Gillespie algorithm [31, 32]. Given this trajectory, the SSDE $Y(t) \mid Z(t)$ can be simulated using, e.g., an Euler-Maruyama or stochastic Runge-Kutta method [33].

The Observation Process. Lastly, we denote with $\{X_i\}_{i\in\mathbb{N}}$ the countable set of observed data points at times $\{t_i\}_{i\in\mathbb{N}}$. The observations $X_i \in \mathcal{X}^l$ are generated as $X_i \sim p(x_i \mid y_i)$, where $p(x_i \mid y_i) := \partial_{x_{i1}} \cdots \partial_{x_{il}} \mathbb{P}(X_i \leq x_i \mid Y(t_i) = y_i)$, $i \in \mathbb{N}$, by conditioning on the diffusion process $Y(t)$. The observation space $\mathcal{X}$ can be either discrete, $\mathcal{X} \subseteq \mathbb{N}$, or continuous, $\mathcal{X} \subseteq \mathbb{R}$. Note that in our model, a continuous-time description for the latent processes is assumed, while the observations are recorded at discrete time points. It therefore belongs to the class of continuous-discrete models, which have a long history in the filtering community [27, 34, 35]. This type of description is of great practical relevance as data is often recorded at discrete time points while the system of interest in fact evolves continuously in time, see, e.g., [10].

2.2 Exact State Inference

We now show how the exact posterior inference problem is solved in principle; the detailed derivations can be found in Appendix A.2.

The inference problem consists of finding the posterior hybrid process $\{Z(t), Y(t) \mid x_{[1,N]}\}$, where we condition on a finite set $x_{[1,N]} = \{x_1, \ldots, x_N\}$ of N observations obtained at time points $\{t_1, \ldots, t_N\}$ in the interval $[0, T]$. The posterior process is fully specified by its marginal density $p(y, z, t \mid x_{[1,N]}) := \partial_{y_1} \cdots \partial_{y_n} \mathbb{P}(Y(t) \leq y, Z(t) = z \mid X_1 = x_1, \ldots, X_N = x_N)$, which is known as the *smoothing distribution*. The smoothing distribution is given by

$$p(y, z, t \mid x_{[1,N]}) = C^{-1}(t)\alpha(y, z, t)\beta(y, z, t), \tag{5}$$

with the filtering density $\alpha(y, z, t) := \partial_{y_1} \cdots \partial_{y_n} \mathbb{P}(Y(t) \leq y, Z(t) = z \mid X_1 = x_1, \ldots, X_k = x_k)$, the backward density $\beta(y, z, t) := \prod_{i=1}^{l} \prod_{j=k+1}^{N} \partial_{x_{j_i}} \mathbb{P}(X_{k+1} \leq x_{k+1}, \ldots, X_N \leq x_N \mid Y(t) =$

$y, Z(t) = z)$ and a time-dependent normalizer $C(t) = \sum_z \int \alpha(y,z,t)\beta(y,z,t)\,\mathrm{d}y$, where $k = \max(k' \in \mathbb{N} \mid t_{k'} \leq t)$. The components α, β and C are continuous-time analogs to the quantities of the forward-backward algorithm for discrete-time hidden Markov models (HMMs) [5].

It is a standard result for continuous-discrete filtering problems [27] that the filtering distribution between observation time points follows the prior dynamics, $\partial_t \alpha(y,z,t) = \mathcal{A}\alpha(y,z,t)$, with initial condition $\alpha(y,z,0) = p_0(z)p_0(y,z)$. At the observation times, it is reset as $\alpha(y,z,t_i) = \tilde{C}_i^{-1}\alpha(y,z,t_i^-)p(x_i \mid y)$, with the normalizer $\tilde{C}_i = \sum_z \int \alpha(y,z,t_i^-)p(x_i \mid y)\,\mathrm{d}y$ and $\alpha(y,z,t_i^-) := \lim_{h\searrow 0}\alpha(y,z,t_i - h)$. Similarly, the backward distribution between observations is given as the solution to another PDE [16]

$$\partial_t \beta(y,z,t) = -\mathcal{A}^\dagger \beta(y,z,t), \tag{6}$$

with end point condition $\beta(y,z,T) = 1$ and adjoint operator $\mathcal{A}^\dagger$, see Appendix A.1. The reset conditions at observation times are given as $\beta(y,z,t_i^-) = \beta(y,z,t_i)p(x_i \mid y)$.

By calculating the time derivative of Eq. (5), it can be shown (see Appendix A.2.3) that the smoothing distribution itself follows a HME

$$\partial_t p(y,z,t \mid x_{[1,N]}) = \tilde{\mathcal{A}} p(y,z,t \mid x_{[1,N]}), \tag{7}$$

with initial condition $p(y,z,0 \mid x_{[1,N]}) \propto p_0(z)p_0(y,z)\beta(y,z,0)$. The operator $\tilde{\mathcal{A}}$ contains the posterior drift function $\tilde{f}_i(y,z,t) = f_i(y,z,t) + \sum_{j=1}^n D_{ij}(y,z,t)\partial_{y_j}\{\log\beta(y,z,t)\}$, the dispersion matrix $\tilde{D}(y,z,t) = D(y,z,t)$ and the posterior rate function $\tilde{\Lambda}(z',z,t) = \Lambda(z',z,t)\frac{\beta(y,z,t)}{\beta(y,z',t)}$.

3 Approximate Inference

Since the smoothing distribution is governed by the HME (7), which depends on the solution of Eq. (6), the exact inference problem amounts to solving two PDEs, which is computationally intractable already for toy systems. Similarly, a naïve posterior sampling scheme would still require solving the backward PDE Eq. (6) and hence suffers from the same issue. To address this challenge, we adopt a VI approach: we aim to find an approximate path measure $\mathbb{Q}_{Y,Z}$ that minimizes the path-wise Kullback-Leibler (KL) divergence

$$\mathrm{KL}\left(\mathbb{Q}_{Y,Z} \,||\, \mathbb{P}_{Y,Z|X}\right) = \mathsf{E}_{\mathbb{Q}_{Y,Z}}\left[\log\frac{\mathrm{d}\mathbb{Q}_{Y,Z}}{\mathrm{d}\mathbb{P}_{Y,Z|X}}\right], \tag{8}$$

where $\frac{\mathrm{d}\mathbb{Q}_{Y,Z}}{\mathrm{d}\mathbb{P}_{Y,Z|X}}$ is the Radon-Nikodym derivative between $\mathbb{Q}_{Y,Z}$ and the exact posterior measure $\mathbb{P}_{Y,Z|X}$ over paths $Y_{[0,T]} := \{Y(t)\}_{t\in[0,T]}$ and $Z_{[0,T]} := \{Z(t)\}_{t\in[0,T]}$. For details on the path-wise KL divergence between stochastic processes, see, e.g., [22, 36, 37]. It is a standard result for VI methods that Eq. (8) can be recast as [19, 38]

$$\mathrm{KL}\left(\mathbb{Q}_{Y,Z} \,||\, \mathbb{P}_{Y,Z|X}\right) = \mathrm{KL}\left(\mathbb{Q}_{Y,Z} \,||\, \mathbb{P}_{Y,Z}\right) - \mathsf{E}_{\mathbb{Q}_{Y,Z}}[\ln p(x_{[1,N]} \mid y_{[0,T]})] + \log p(x_{[1,N]}), \tag{9}$$

with the expected log-likelihood $\mathsf{E}_{\mathbb{Q}_{Y,Z}}[\ln p(x_{[1,N]} \mid y_{[0,T]})] = \sum_{i=1}^N \mathsf{E}_{\mathbb{Q}_{Y,Z}}[\ln p(x_i \mid y_i)]$. The minimization problem over Eq. (8) can then be cast as a maximization problem over the evidence lower bound (ELBO) [19]

$$\mathcal{L}[\mathbb{Q}_{Y,Z}] = \mathsf{E}_{\mathbb{Q}_{Y,Z}}[\ln p(x_{[1,N]} \mid y_{[0,T]})] - \mathrm{KL}\left(\mathbb{Q}_{Y,Z} \,||\, \mathbb{P}_{Y,Z}\right), \tag{10}$$

which does not include the computationally intractable marginal log-likelihood $\log p(x_{[1,N]})$ and can hence be evaluated.

Optimizing Eq. (10) requires explicitly computing the path-wise KL divergence $\mathrm{KL}\left(\mathbb{Q}_{Y,Z} \,||\, \mathbb{P}_{Y,Z}\right)$. For two hybrid processes of the same class obeying Eq. (3), this expression can formally be derived using Girsanov's theorem for diffusion processes [39] and MJPs [40]. A more intuitive derivation can however be carried out using a limiting argument, similar to [23]. Assuming a constant, state- and mode-independent dispersion D as done also in [21] and a drift $g(y,z,t)$ and rate function $\tilde{\Lambda}(z,z',t)$ pertaining to the variational measure $\mathbb{Q}_{Y,Z}$, the path-wise KL divergence is obtained as

$$\begin{aligned}\mathrm{KL}\left(\mathbb{Q}_{Y,Z} \,||\, \mathbb{P}_{Y,Z}\right) = \mathrm{KL}\left(\mathbb{Q}^0_{Y,Z} \,||\, \mathbb{P}^0_{Y,Z}\right) + \frac{1}{2}\int_0^T \mathsf{E}\Bigg[\|(g(y,z,t) - f(y,z,t)\|^2_{D^{-1}} \\ + \sum_{z'\in\mathcal{Z}\setminus z}\left\{\tilde{\Lambda}(z,z',t)\left(\ln\tilde{\Lambda}(z,z',t) - \ln\Lambda(z,z',t)\right)\right\} - \left(\tilde{\Lambda}(z,t) - \Lambda(z,t)\right)\Bigg]\,\mathrm{d}t,\end{aligned} \tag{11}$$

with the weighted norm $\|x\|_A^2 := x^\top A x$, the KL of the initial distributions, $\mathrm{KL}\left(\mathbb{Q}^0_{Y,Z} \mid\mid \mathbb{P}^0_{Y,Z}\right)$, and the expectation is carried out with respect to the variational time-point marginal $q(y,z,t) := \partial_{y_1}\cdots\partial_{y_n}\mathbb{Q}(Y(t)\leq y, Z(t)=z)$. We note that extensions to mode- and time-dependent $D = D(z,t)$ [39] or state-dependent $D = D(y)$ [22] are also possible. Additionally, in the absence of any coupling between $Z(t)$ and $Y(t)$, i.e., $f(y,z,t) = f(y,t)$, Eq. (11) reduces to the sum of the known individual path-wise KL divergences for diffusion processes and MJPs. The detailed derivation can be found in Appendix A.3.1.

3.1 The Constrained Objective

Since Eq. (10) is a mere reformulation, it is still optimized by the true, intractable posterior distribution $\mathbb{P}_{Y,Z|X}$. To arrive at computationally tractable expressions, we restrict the class $\mathcal{Q}$ of admissible variational processes. Making a structured mean-field ansatz, we approximate the exact joint posterior density $p(y,z,t \mid x_{[1,N]})$ as

$$\begin{aligned} p(y,z,t \mid x_{[1,N]}) &= p(z,t \mid x_{[1,N]}) \cdot p(y,t \mid z, x_{[1,N]}) \\ &\approx q_Z(z,t)\cdot q_Y(y,t\mid z) = q(y,z,t), \end{aligned} \tag{12}$$

with $p(y,t \mid z, x_{[1,N]}) := \partial_{y_1}\cdots\partial_{y_n}\mathbb{P}(Y(t)\leq y \mid Z(t)=z, x_{[1,N]})$. We approximate the exact conditional $p(y,t\mid z, x_{[1,N]})$, which in general does not have a simple parametric form, by one fixed parametric expression per mode $q_Y(y,t\mid z) := \partial_{y_1}\cdots\partial_{y_n}\mathbb{Q}(Y(t)\leq y \mid \tilde{Z}=z)$ via the introduction of the time-independent random variable $\tilde{Z}$. This results in a point-wise mixture distribution $q(y,z,t)$ with weights $q_Z(z,t) := \mathbb{Q}(Z(t)=z)$ and mixture densities $q_Y(y,t\mid z)$. Note that this is similar in spirit to amortized inference techniques [41], because we utilize the same parametric form for all times t.

To ensure the mixture distribution structure Eq. (12) to hold at every time point t, we impose separate constraints on the dynamics of $q_Z(z,t)$ and $q_Y(y,t\mid z)$. Firstly, we require the marginal $q_Z(z,t)$ to obey a master equation

$$\frac{\mathrm{d}}{\mathrm{d}t}q_Z(z,t) = \sum_{z'\in\mathcal{Z}\setminus z}\tilde{\Lambda}(z',z,t)q_Z(z',t) - \tilde{\Lambda}(z,t)q_Z(z,t),\ \forall z\in\mathcal{Z}. \tag{13}$$

This reproduces the structure of the exact posterior marginal $p(z,t\mid x_{[1,N]})$, which can be seen by integrating out the continuous variable in the HME (7), c.f. Eq. (4). Secondly, we constrain the variational factor $q_Y(y,t\mid z)$ to follow a Fokker-Planck equation (FPE) [15] with linear variational drift $g(y,z,t) = A(z,t)y + b(z,t)$ for every mode z individually

$$\partial_t q_Y(y,t\mid z) = -\sum_{i=1}^n \partial_{y_i}\left\{g_i(y,z,t)q_Y(y,t\mid z)\right\} + \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \partial_{y_i}\partial_{y_j}\{D_{ij}q_Y(y,t\mid z)\}. \tag{14}$$

Equation (14) describes the marginal density of a classical SDE, which, under linear drift, is equivalent to a GP [42]. This PDE is hence solved by a time-dependent Gaussian distribution $q_Y(y,t\mid z) = \mathcal{N}(y\mid \mu(z,t), \Sigma(z,t))$ [27], where the dynamics of the parameters is described by two ordinary differential equations (ODEs)

$$\dot{\mu}(z,t) = A(z,t)\mu(z,t) + b(z,t), \quad \dot{\Sigma}(z,t) = A(z,t)\Sigma(z,t) + \Sigma(z,t)A^\top(z,t) + D,\ \forall z\in\mathcal{Z}. \tag{15}$$

Our approach hence amounts to a mixture of GPs: this approximation will be accurate whenever the distribution $q_Z(z,t)$ is peaked at one $z\in\mathcal{Z}$. In this case, the HME separates into a FPE and a master equation for the processes $Y(t)\mid\tilde{Z}$ and $Z(t)$, respectively; see Appendix A.3.2 for details. Accordingly, the approximation error over the whole interval $[0,T]$ will be small if the original system dynamics are linear in each mode and the modes are well discernible. Since we are interested specifically in meta-stable systems which, by definition, transition between qualitatively different regimes and exhibit a separation of time scales between the intra-mode diffusive dynamics and the inter-mode transitions, we expect these criteria to be met reasonably well for our systems of interest.

The constraints (13) and (15) can be included into the objective Eq. (10) via Lagrange multiplier functions, yielding an augmented objective, the *Lagrangian*. We define the multipliers $\lambda(z,t)$, $\Psi(z,t), \nu(z,t)$ for the variational mean $\mu(z,t)$ and covariance $\Sigma(z,t)$ and variational rates $\tilde{\Lambda}(z,z',t)$,

respectively. Writing the ELBO as $\mathcal{L}[\mathbb{Q}_{Y,Z}] = \int_0^T \ell_{\mathbb{Q}}(t)\,\mathrm{d}t$, the full Lagrangian L to be maximized reads

$$L = \int_0^T \ell_{\mathbb{Q}}(t) + \sum_{z\in\mathcal{Z}} \Big[\lambda^\top(z,t)\left(\dot{\mu}(z,t) - \left(A(z,t)\mu(z,t) + b(z,t)\right)\right) + \mathrm{tr}\,\Big\{\Psi^\top(z,t)\Big(\dot{\Sigma}(z,t) - \big(A(z,t)\Sigma(z,t) + \Sigma(z,t)A^\top(z,t) + D\big)\Big)\Big\} + \nu(z,t)\left(\dot{q}_Z(z,t) - \sum_{z'\in\mathcal{Z}} \tilde{\Lambda}_{z'z}(t) q_Z(z',t)\right)\Big]\,\mathrm{d}t, \tag{16}$$

where we used the shorthand $\tilde{\Lambda}_{z'z}(t) := \tilde{\Lambda}(z,' z,t)$ for $z' \neq z$ and $\tilde{\Lambda}_{zz}(t) := -\tilde{\Lambda}(z,t)$ else. Note that the dependency of L on the variational measure $\mathbb{Q}_{Y,Z}$ is fully captured by the variational factors, $L = L[\mathbb{Q}_{Y,Z}] = L[q_Z, \mu, \Sigma]$.

3.2 Optimizing the Variational Distributions

The optimization problem consists in finding the optimal variational factors q_Z^*, μ^*, Σ^* and parameters $A^*, b^*, \tilde{\Lambda}^*, \phi^*$ maximizing Eq. (16), where ϕ summarizes the variational initial conditions. The optimal variational factors have to fulfil the respective constraint equations (13) and (15) as well as the Euler-Lagrange (EL) equation $\frac{\mathrm{d}}{\mathrm{d}t}\partial_{\dot{q}}\ell = \partial_q \ell$, where $L = \int_0^T \ell(t)\,\mathrm{d}t$ [43]. The latter is giving rise to ODEs for the Lagrange multiplier functions: Firstly, the EL equation with respect to $q_Z(z,t)$ yields

$$\dot{\nu}(z,t) = \partial_{q_Z(z,t)}\ell_{\mathbb{Q}} - \sum_{z'\in\mathcal{Z}\setminus z} \tilde{\Lambda}(z,z',t)\nu(z',t) + \tilde{\Lambda}(z,t)\nu(z,t). \tag{17}$$

Secondly, the EL equations hold separately for both Gaussian parameters $\mu(z,t), \Sigma(z,t)$. We obtain

$$\begin{aligned} \dot{\lambda}(z,t) &= \partial_{\mu(z,t)}\ell_{\mathbb{Q}} - A^\top(z,t)\lambda(z,t), \\ \dot{\Psi}(z,t) &= \partial_{\Sigma(z,t)}\ell_{\mathbb{Q}} - A^\top(z,t)\Psi(z,t) - \Psi(z,t)A(z,t). \end{aligned} \tag{18}$$

For a detailed derivation, including the explicit expressions for the gradients $\partial\ell_{\mathbb{Q}}$ for linear prior models $f(y,z,t) = A_p(z,t)y + b_p(z,t)$, see Appendix A.3.3.

The full optimization problem requires the Lagrange multiplier ODEs (17) and (18) and the constraint Eqs. (13) and (15) to be solved jointly as a boundary-value problem with terminal conditions $\nu(\cdot,T), \lambda(\cdot,T), \Psi(\cdot,T) = 0$ and initial conditions on the distribution parameters [43]. The variational parameters have to be optimized simultaneously. A standard approach to this problem is an iterative forward-backward sweeping algorithm [44, 45]: (i) solve the Lagrange multiplier ODEs Eqs. (17) and (18) backward in time, starting from the terminal conditions $\nu, \lambda, \Psi = 0$. Next, (ii) update the variational parameters acting on the constraints, in our case $A, b, \tilde{\Lambda}, \phi$. Here, we employ a simple gradient ascent scheme: for each $u(t) \in \{A(z,t), b(z,t), \tilde{\Lambda}(z,z',t), \phi\}$, we update

$$u(t) \leftarrow u(t) + \kappa(t)\cdot\partial_{u(t)}\ell, \tag{19}$$

where we use a back-tracking line search [46] for the step size $\kappa(t)$, see Appendix A.3.4. Then, (iii) solve the constraint equations (13) and (15) forward in time, starting from initial conditions $\phi = \{q_Z(z,0) = q_Z^0(z), \mu(z,0) = \mu^0(z), \Sigma(z,0) = \Sigma^0(z)\}$. Finally, (iv) repeat until convergence.

Note that our results generalize the findings of [20, 23]: for $f(y,z,t) = f(y,t)$, $g(y,z,t) = g(y,t)$, i.e., in the absence of coupling between the Y- and Z-processes, the MJP and diffusion contributions to the prior KL Eq. (11) separate and the derivative $\partial_{q_Z(z,t)}\ell_{\mathbb{Q}}$ reduces to the result of [23] between observations. Furthermore, for the special case $|\mathcal{Z}| = 1$, our result recovers the conventional GP approximation [20].

3.3 Parameter Learning

To learn the model parameters, that is, the prior transition rate matrix Λ, the dispersion D, the prior initial conditions, the parameters of the drift function $f(y,z,t)$ and the parameters of the observation likelihood $p(x_i \mid y_i)$, we employ a variational expectation maximization (VEM) scheme [5]. After converging onto variational distributions $q_Z(z,t)$ and $q_Y(y,t \mid z)$, we perform a simple gradient ascent scheme on the Lagrangian L, Eq. (16), see Appendix A.3.5. The complete optimization scheme is summarized in Algorithm 1, where we subsume all model parameters under Θ for conciseness.

4 Experiments

4.1 Model Validation on Ground-Truth Data

We validate our method on synthetic data generated from a 1D, two-mode hybrid system with observations corrupted by Gaussian noise, $p(x_i \mid y_i) = \mathcal{N}(x_i \mid y_i, \Sigma_{\text{obs}})$, where the observation times are drawn from a Poisson point process [14]. Both mode dynamics are given by linear drift functions

$$f(y, z, t) = \alpha_z(\beta_z - y), \tag{20}$$

with set points β_z and dynamics $\alpha_z > 0$, $z \in \mathcal{Z} = \{1, 2\}$. For $|\mathcal{Z}| = 1$, this would recover the well-known Ornstein-Uhlenbeck process [27].

input : observation data $\{t_i, x_i\}_{i=1,\dots,N}$

Initialize q_Z, μ, Σ, A, b, $\tilde{\Lambda}$, Θ
while $\mathcal{L}$ *not converged* **do**
  while $\mathcal{L}$ *not converged* **do**
    Compute multiplier functions λ, Ψ, ν via Eqs. (17) and (18)
    Update variational parameters A, b, $\tilde{\Lambda}$ via Eq. (19)
    Compute variational factors μ, Σ, q_Z via Eqs. (13) and (15)
    Update lower bound $\mathcal{L}$
  end
  Update prior parameters Θ via gradient ascent
  Update lower bound $\mathcal{L}$
end

Algorithm 1: VI for hybrid processes

As shown in Fig. 2 A, the inferred posterior distributions $q_Y(y, t) = \sum_{z \in \mathcal{Z}} q_Y(y, t \mid z)$ and $q_Z(z, t)$ both faithfully reconstruct the respective latent ground-truth trajectories. This is also reflected by the maximum a-posteriori (MAP) paths $y^{\text{MAP}}(t), z^{\text{MAP}}(t) = \arg\max_{y,z} q(y, z, t)$. In regions around mode transitions, one can observe artifacts from the variational approximation as a mixture of GPs: in the time interval Δt_{trans} between the last observation before and the first observation after the ground-truth mode transition at $t \approx 12.5$, the marginal $q_Y(y, t)$ does not exhibit a smooth transition across $Y = 0$, but splits the probability density between the independently evolving Gaussian distributions $q_Y(y, t \mid z)$. This is not surprising, as we have argued (c.f. Section 3.1) that our approximation will be accurate in regions where $q_Z(z, t) \approx 1$, which is the case at the beginning and the end of the interval Δt_1, but not in between. Since the relaxation onto the mode set points β_z is fast compared to the mode remain times, these transition regions are short, yielding a high approximation quality. As we pursue a generative modeling approach, we can verify this by sampling full trajectories from the variational posterior. The empirical distribution over paths closely resembles the latent continuous trajectory. We provide a plot of the sampling distribution in Appendix B.1. Furthermore, the model parameters are identified with high accuracy: the learned set points, for instance, $\beta_1 = 0.72, \beta_2 = -0.41$, where the ground-truth values are ± 1. The exhaustive list of both the learned and ground truth model parameters is also provided in Appendix B.1.

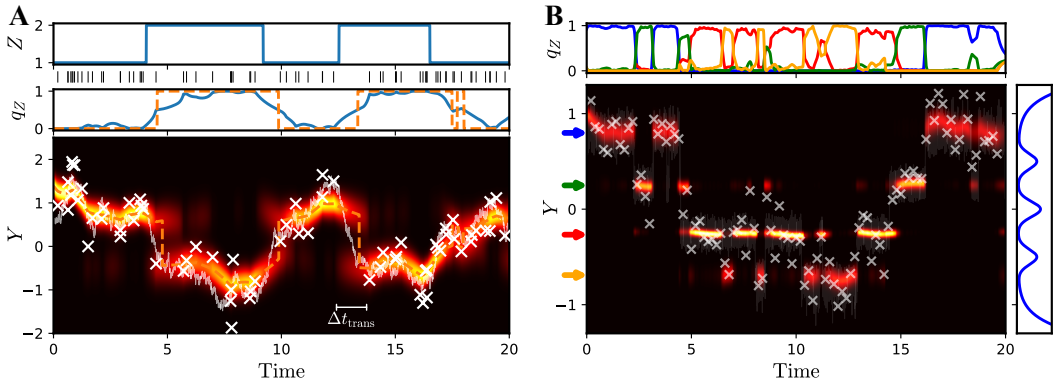

Figure 2: **A**: Model validation on ground-truth data from a 1D, two-mode hybrid process. Top: True discrete path $z(t)$ with observation times (vertical lines below). Middle: inferred marginal $q_Z(2, t)$ and $z^{\text{MAP}}(t)$ (dashed). Bottom: true SSDE path (gray line), observations (crosses), the inferred marginal $q_Y(y, t)$ and $y^{\text{MAP}}(t)$ (dashed). Brighter colors indicate higher probability density. **B**: Diffusion in a 1D four-well potential (shown right). Top: inferred marginals $q_Z(z, t)$. Bottom: true SDE and observations with the inferred $q_Y(y, t)$ and learned set points β_z (arrows).

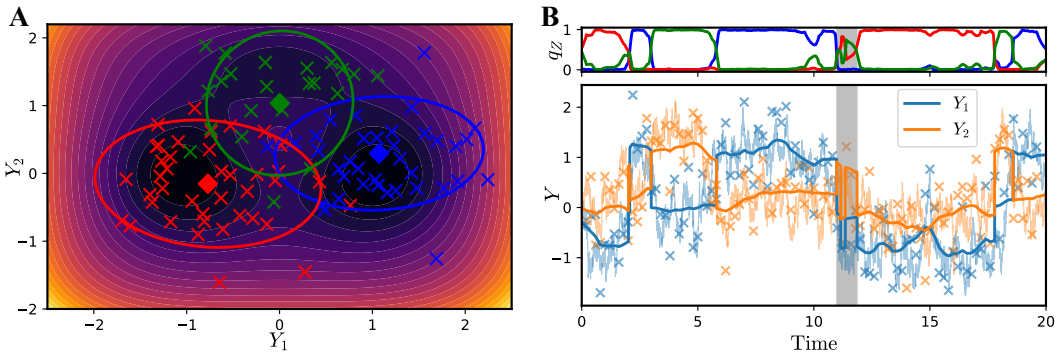

Figure 3: Diffusion in a three-well potential. **A**: Potential landscape with the inferred set points β_z (diamonds) and dispersions $D(z)$ (ellipses, 3σ-region) with observations (crosses); colors according to $z^{\mathrm{MAP}}(t_i)$ for each observation x_i. **B**: Top: inferred marginals $q_Z(z,t)$. Bottom: components of $y^{\mathrm{MAP}}(t)$ (thick lines), the ground-truth path (thin lines) and the observations (crosses). Shaded region: transition region with high ambiguity.

4.2 Diffusions in Multi-Well Potentials

In many real-world scenarios, discrete modes that drive continuous dynamics do not exist a priori, but continuous dynamics often exhibit regimes that are qualitatively different from one another. Transitions between different regimes typically occur on vastly longer time scales than the relaxation dynamics within each regime, as is observed, e.g., for the folding dynamics of complex biomolecules [47]. For such meta-stable systems in particular, probabilistic models explicitly assuming the existence of discrete modes can aid understanding and enable targeted interventions on the system. To demonstrate the capability of our model to yield sensible representations of distinct dynamic regimes, we apply it to latent Itô diffusions driven by 1D and 2D benchmark potentials widely used in computational biology [48, 49, 50, 51] We model these system via hybrid processes with linear drift, c.f. Eq. (20), and Gaussian observation noise in both the 1D and 2D case, as before. We assume a mode-dependent dispersion, $D = D(z)$. The observation time points are regularly spaced and we fix the observation covariance Σ_{obs}.

In both the 1D and 2D case, the mode reconstructions accurately capture the global transitions between distinct potential minima, see Fig. 2 B and Fig. 3 B. We note that in the 1D example (Fig. 2 B), the true latent continuous trajectory exhibits a particularly clear separation of time scales between the inter- and intra-well dynamics, which is reflected in sharp transitions in the mode reconstruction and accordingly particularly accurate results. On the other hand, as shown in Fig. 3, in the 2D case, more pronounced transition regions exist, where it is not possible to unambiguously assign the state at a given time t to one of the three minima. The posterior marginals $q_Z(z,t)$ sensibly capture this uncertainty, which is also reflected in a high quality mode-assignment of the observed data points $x_{[1,N]}$ as shown in Fig. 3 A. Furthermore, the asymmetry of the learned mode-dependent dispersions accurately reflects the topology of the underlying potential. An overview over all parameters is given in Appendix B.2.

4.3 Switching Ion Channel Data

We apply our method to a structural molecular biology problem: we aim to identify the switching behavior of the viral ion channel $\mathrm{Kcv}_{\mathrm{MT325}}$ exhibiting three different channel conformations [52]. Different conformations yield different ion permeabilities and hence different conductivities which can be directly detected by applying a voltage across the cell membrane and measuring the trans-membrane current. We model the conformation dynamics as discrete process Z and the current, passing through an analog (continuous-time) filter and incurring amplifier noise, as continuous process Y. The observations X_i are sampled at a fixed rate and subject to quantization errors from an analog-to-digital converter. The observation noise is a known property of the used setup; we hence fix Σ_{obs}. We reconstruct a highly plausible switching behavior and filter out individual outliers, as depicted in Fig. 4; see Appendix B.3 for a list of all learned parameters and experimental details. We note that very similar problem setups can be found, e.g., in nanopore sequencing technologies [53].

4.4 Learning Complex Latent Continuous Dynamics

After having demonstrated the applicability of our model to hybrid systems with different time scales for the discrete and continuous dynamics, we lastly show that it also works well when this criterion is not met. In areas such as automation and robotics, hybrid models are ubiquitously used to, e.g., encode highly complex continuous movements via a discrete set of movement primitives [54], which are non-stationary processes. To demonstrate that our method is able to reconstruct such complex latent continuous dynamics, we employ a 2D version of Eq. (20) where the mode dynamics $f(y, z, t) = f(y, z)$ are given as two counter-rotating vector fields, see Fig. 5 B. We fix the observation covariance, as we can assume its value to be known and small compared to the system volatility for applications such as robotics.

As shown in Fig. 5 A and C, the mode and state reconstructions accurately recover the true paths. Accordingly, also the underlying mode dynamics are correctly learned, exhibiting the counter-rotating behavior of the ground-truth model. We provide a list of all parameters in Appendix B.4.

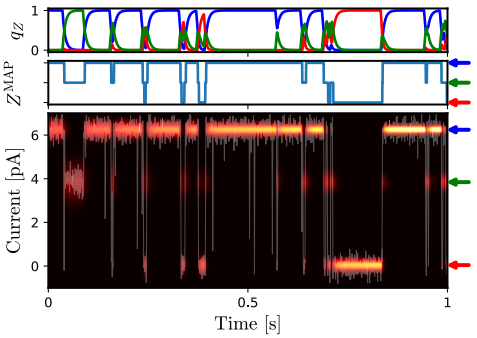

Figure 4: Switching behavior of viral ion channel $\mathrm{Kcv}_{\mathrm{MT325}}$. Top: marginals $q_Z(z, t)$ corresponding to three channel conformations. Middle: $z^{\mathrm{MAP}}(t)$. Bottom: data (white line), marginals $q_Y(y, t)$ and learned set points β_z (arrows). Brighter colors indicate higher probability density.

5 Conclusion

We presented, to the best of our knowledge, the first variational inference framework for continuous time hybrid process models: since the exact filtering and smoothing distributions are computationally intractable, we proposed a variational approximation to the exact model. The key assumption is that the true discrete posterior is peaked at any $z \in \mathcal{Z}$ for extended periods of time, allowing for a straightforward, easily interpretable mixture of GPs to be used as approximation. We have evaluated our framework on various benchmark tasks including real-world biological data and demonstrated its ability to faithfully reconstruct complex latent dynamics and to learn the unknown system parameters, in particular in applications to meta-stable systems. While we implemented parameter learning via point estimates, we aim to extend this to a fully Bayesian framework in the future, enabling the integration of prior domain knowledge and the associated uncertainty about the system at hand. As many natural and engineered systems can be described as hybrid systems, we think that extending the toolbox for inference will be of great utility for the analysis and control of such systems.

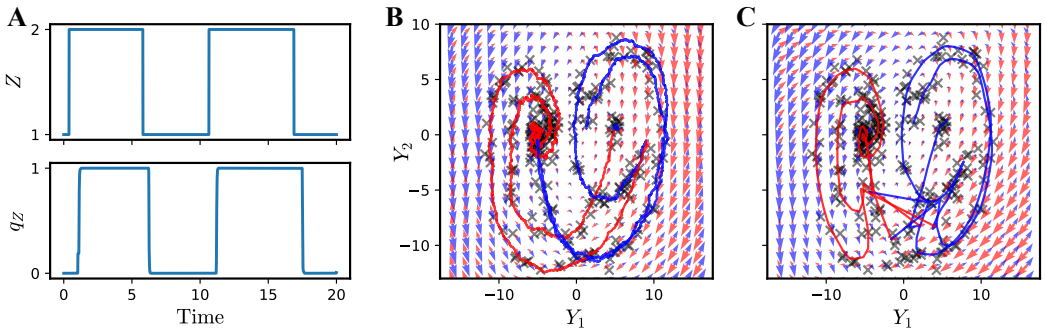

Figure 5: Inference of complex structured continuous dynamics. **A**: True mode path $z(t)$ (top) and inferred marginal $q_Z(2, t)$ (bottom). **B**: True state path $y(t)$ (coloring according to associated mode; $Z = 1$: blue, $Z = 2$: red), observations (crosses) and true mode dynamics $f(y, z)$ (arrows) in the phase plane. **C**: Reconstructed MAP path $y^{\mathrm{MAP}}(t)$ (coloring according to $z^{\mathrm{MAP}}(t)$) and reconstructed mode dynamics (arrows).

References

[1] J. Proctor, R. P. Kukillaya, and P. Holmes. "A phase-reduced neuro-mechanical model for insect locomotion: feed-forward stability and proprioceptive feedback". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 368.1930 (2010), pp. 5087–5104.

[2] Carlos M. Carvalho and Hedibert F. Lopes. "Simulation-based sequential analysis of Markov switching stochastic volatility models". In: *Computational Statistics & Data Analysis* 51.9 (2007), pp. 4526–4542.

[3] Alan S. Willsky. "A survey of design methods for failure detection in dynamic systems". In: *Automatica* 12.6 (1976), pp. 601–611.

[4] Paul C. Bressloff. "Stochastic switching in biology: from genotype to phenotype". In: *Journal of Physics A: Mathematical and Theoretical* 50.13 (2017), p. 133001.

[5] David Barber. *Bayesian reasoning and machine learning*. Cambridge University Press, 2012.

[6] Emily Fox, Erik Sudderth, Michael Jordan, and Alan Willsky. "Nonparametric Bayesian learning of switching linear dynamical systems". In: *Advances in Neural Information Processing Systems* 21 (2008), pp. 457–464.

[7] Matthew J. Johnson, David Duvenaud, Alexander B. Wiltschko, Sandeep R. Datta, and Ryan P Adams. "Composing graphical models with neural networks for structured representations and fast inference". In: *Advances in Neural Information Processing Systems 29* (2016), pp. 2954–2962.

[8] Scott W. Linderman, Matthew J. Johnson, Andrew C. Miller, Ryan P. Adams, David M. Blei, and Liam Paninski. "Bayesian learning and inference in recurrent switching linear dynamical systems". In: *20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017*. 2017.

[9] Joshua Glaser, Matthew Whiteway, John P. Cunningham, Liam Paninski, and Scott Linderman. "Recurrent Switching Dynamical Systems Models for Multiple Interacting Neural Populations". In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin. Vol. 33. Curran Associates, Inc., 2020.

[10] Christos G. Cassandras and Stephane Lafortune. *Introduction to discrete event systems*. Springer Science & Business Media, 2009.

[11] Xuerong Mao and Chenggui Yuan. *Stochastic differential equations with Markovian switching*. Imperial college press, 2006.

[12] Mark H. A. Davis. *Markov models & optimization*. Routledge, 2018.

[13] Paul C. Bressloff. "Switching diffusions and stochastic resetting". In: *Journal of Physics A: Mathematical and Theoretical* 53.27 (2020), p. 275003.

[14] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes: characterization and convergence*. John Wiley & Sons, 2009.

[15] Bernt Øksendal. *Stochastic differential equations*. Springer, 2003.

[16] Brian D. O. Anderson and Ian B. Rhodes. "Smoothing algorithms for nonlinear finite-dimensional systems". In: *Stochastics: An International Journal of Probability and Stochastic Processes* 9.1-2 (1983), pp. 139–165.

[17] Hans-Christian Ruiz and Hilbert J. Kappen. "Particle smoothing for hidden diffusion processes: Adaptive path integral smoother". In: *IEEE Transactions on Signal Processing* 65.12 (2017), pp. 3191–3203.

[18] Vinayak Rao and Yee Whye Teh. "Fast MCMC Sampling for Markov Jump Processes and Extensions." In: *Journal of Machine Learning Research* 14.11 (2013).

[19] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. "Variational inference: A review for statisticians". In: *Journal of the American statistical Association* 112.518 (2017), pp. 859–877.

[20] Cédric Archambeau, Dan Cornford, Manfred Opper, and John Shawe-Taylor. "Gaussian process approximations of stochastic differential equations". In: *Journal of Machine Learning Research* 1 (2007), pp. 1–16.

[21] Cédric Archambeau, Manfred Opper, Yuan Shen, Dan Cornford, and John Shawe-Taylor. "Variational Inference for Diffusion Processes". In: *Advances in Neural Information Processing Systems* 20 (2007), pp. 17–24.

[22] Tobias Sutter, Arnab Ganguly, and Heinz Koeppl. "A Variational Approach to Path Estimation and Parameter Inference of Hidden Diffusion Processes". In: *Journal of Machine Learning Research* 17.190 (2016), pp. 1–37.

[23] Manfred Opper and Guido Sanguinetti. "Variational inference for Markov jump processes". In: *Advances in neural information processing systems* 20 (2007), pp. 1105–1112.

[24] Christian Wildner and Heinz Koeppl. "Moment-based variational inference for Markov jump processes". In: *International Conference on Machine Learning*. PMLR. 2019, pp. 6766–6775.

[25] E. Weinan and E. Vanden-Eijnden. "Metastability, conformation dynamics, and transition pathways in complex systems". In: *Multiscale modelling and simulation*. Springer, 2004, pp. 35–68.

[26] J. R. Norris. *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997.

[27] Simo Särkkä and Arno Solin. *Applied stochastic differential equations*. Cambridge University Press, 2019.

[28] R. Pawula. "Generalizations and extensions of the Fokker-Planck-Kolmogorov equations". In: *IEEE Transactions on Information Theory* 13.1 (1967), pp. 33–41.

[29] Derya Altıntan and Heinz Koeppl. "Hybrid master equation for jump-diffusion approximation of biomolecular reaction networks". In: *BIT Numerical Mathematics* (2019), pp. 1–34.

[30] Christian Grossmann, Hans-Görg Roos, and Martin Stynes. *Numerical treatment of partial differential equations*. Springer, 2007.

[31] Joseph L Doob. "Markoff chains–denumerable case". In: *Transactions of the American Mathematical Society* 58.3 (1945), pp. 455–473.

[32] Daniel T. Gillespie. "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions". In: *Journal of computational physics* 22.4 (1976), pp. 403–434.

[33] Peter E. Kloeden and Eckhard Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, 1992.

[34] Peter S. Maybeck. *Stochastic models, estimation, and control*. Academic press, 1982.

[35] Frederick E. Daum. "Exact finite dimensional nonlinear filters for continuous time processes with discrete time measurements". In: *The 23rd IEEE Conference on Decision and Control*. IEEE. 1984, pp. 16–22.

[36] Sanjoy K. Mitter and Nigel J. Newton. "A variational approach to nonlinear estimation". In: *SIAM journal on control and optimization* 42.5 (2003), pp. 1813–1833.

[37] Alexander G. de G. Matthews, James Hensman, Richard Turner, and Zoubin Ghahramani. "On sparse variational methods and the Kullback-Leibler divergence between stochastic processes". In: *Artificial Intelligence and Statistics*. PMLR. 2016, pp. 231–239.

[38] Ramon Van Handel. "Filtering, stability, and robustness". PhD thesis. California Institute of Technology, 2007.

[39] Tommi Sottinen and Simo Särkkä. "Application of Girsanov theorem to particle filtering of discretely observed continuous-time non-linear systems". In: *Bayesian Analysis* 3.3 (2008), pp. 555–584.

[40] Claude Kipnis and Claudio Landim. *Scaling limits of interacting particle systems*. Vol. 320. Springer Science & Business Media, 1998.

[41] C. Zhang, J. Butepage, H. Kjellstrom, and S. Mandt. "Advances in Variational Inference". In: *IEEE Transactions on Pattern Analysis & Machine Intelligence* 41.08 (2019), pp. 2008–2026. ISSN: 1939-3539.

[42] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005. ISBN: 026218253X.

[43] Daniel Liberzon. *Calculus of variations and optimal control theory: a concise introduction*. Princeton university press, 2011.

[44] Robert F. Stengel. *Optimal control and estimation*. Courier Corporation, 1994.

[45] Michael McAsey, Libin Mou, and Weimin Han. "Convergence of the forward-backward sweep method in optimal control". In: *Computational Optimization and Applications* 53.1 (2012), pp. 207–226.

[46] Dimitri P. Bertsekas. "Nonlinear programming". In: *Journal of the Operational Research Society* 48.3 (1997), pp. 334–334.

[47] Jiri Sponer, Giovanni Bussi, Miroslav Krepl, Pavel Banáš, Sandro Bottaro, Richard A Cunha, Alejandro Gil-Ley, Giovanni Pinamonti, Simón Poblete, Petr Jurečka, et al. "RNA structural dynamics as captured by molecular simulations: a comprehensive overview". In: *Chemical reviews* 118.8 (2018), pp. 4177–4338.

[48] Jan-Hendrik Prinz, Hao Wu, Marco Sarich, Bettina Keller, Martin Senne, Martin Held, John D Chodera, Christof Schütte, and Frank Noé. "Markov models of molecular kinetics: Generation and validation". In: *The Journal of chemical physics* 134.17 (2011), p. 174105.

[49] Hao Wu, Andreas Mardt, Luca Pasquali, and Frank Noe. "Deep Generative Markov State Models". In: *Advances in Neural Information Processing Systems* 31 (2018), pp. 3975–3984.

[50] Frank Noé, Hao Wu, Jan-Hendrik Prinz, and Nuria Plattner. "Projected and hidden Markov models for calculating kinetics and metastable states of complex molecules". In: *The Journal of chemical physics* 139.18 (2013).

[51] Hao Wu, Feliks Nüske, Fabian Paul, Stefan Klus, Péter Koltai, and Frank Noé. "Variational Koopman models: slow collective variables and molecular kinetics from short off-equilibrium simulations". In: *The Journal of chemical physics* 146.15 (2017), p. 154104.

[52] Sabrina Gazzarrini, Ming Kang, Svetlana Epimashko, James L. Van Etten, Jack Dainty, Gerhard Thiel, and Anna Moroni. "Chlorella virus MT325 encodes water and potassium channels that interact synergistically". In: *Proceedings of the National Academy of Sciences* 103.14 (2006), pp. 5355–5360.

[53] Miten Jain, Sergey Koren, Karen H Miga, Josh Quick, Arthur C Rand, Thomas A Sasani, John R Tyson, Andrew D Beggs, Alexander T Dilthey, Ian T Fiddes, et al. "Nanopore sequencing and assembly of a human genome with ultra-long reads". In: *Nature biotechnology* 36.4 (2018), pp. 338–345.

[54] Debora Clever, Monika Harant, Katja Mombaur, Maximilien Naveau, Olivier Stasse, and Dominik Endres. "COCoMoPL: A novel approach for humanoid walking generation combining optimal control, movement primitives and learning and its transfer to the real robot HRP-2". In: *IEEE Robotics and Automation Letters* 2.2 (2017), pp. 977–984.

[55] Cedric Archambeau and Manfred Opper. "Approximate inference for continuous-time Markov processes". In: *Bayesian Time Series Models* (2011), pp. 125–140.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] We provide the contributions we announce in the abstract.
   (b) Did you describe the limitations of your work? [Yes] See Sections 3 to 5.
   (c) Did you discuss any potential negative societal impacts of your work? [N/A] This is theoretical work.
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes] See Section 3.
   (b) Did you include complete proofs of all theoretical results? [Yes] See appendix.
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No] We will include a link to the implementation and all synthetic data in the camera-ready version of the paper. We will not include the data from biological experiments since they are proprietary.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See appendix.

(c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A] We plot full distributions.

(d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [No] Not relevant since the code is not computationally heavy.

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...

(a) If your work uses existing assets, did you cite the creators? [No] We will include a citation in the camera-ready version of the paper; we omitted it to ensure double-blindness.

(b) Did you mention the license of the assets? [N/A]

(c) Did you include any new assets either in the supplemental material or as a URL? [N/A]

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [No] We will mention the lab in the acknowledgements in the camera-ready version to ensure double-blindness.

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [No] It does not.

5. If you used crowdsourcing or conducted research with human subjects...

(a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]

(b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]