A Bayesian Nonparametric Model for Spectral Estimation of Metastable Systems

Hao Wu
Department of Mathematics and Computer Science
Free University of Berlin
Arnimallee 6, 14195 Berlin, Germany

Abstract

The identification of eigenvalues and eigenfunctions from simulation or experimental data is a fundamental and important problem for analysis of metastable systems, because the dominant spectral components usually contain a lot of essential information of the metastable dynamics on slow timescales. It has been shown that the dynamics of a strongly metastable system can be equivalently described as a hidden Markov model (HMM) under some technical assumptions and the spectral estimation can be performed through HMM learning. However, the spectral estimation with unknown number of dominant spectra is still a challenge in the framework of traditional HMMs, and the infinite HMMs developed based on stick-breaking processes cannot satisfactorily solved this problem either. In this paper, we analyze the difficulties of spectral estimation for infinite HMMs, and present a new nonparametric model called stick-breaking half-weighted model (SB-HWM) to address this problem. The SB-HWM defines a sparse prior of eigenvalues and can be applied to Bayesian inference of dominant eigenpairs of metastable systems in a nonparametric manner. We demonstrate by simulations the advantages of applying SB-HWM to spectral estimation.

1 INTRODUCTION

In a variety of scientific areas, we are confronted with the task of analyzing and modeling a complex system which can be described as a Markov process $\{x_t\}$ with time evolution equation

$$\begin{aligned}\rho_{t+\tau}(x) &= \mathcal{P}(\tau)\,\rho_t(x) \\ &\triangleq \int_\Omega p(x_{t+\tau}=x|x_t=x')\,\rho_t(x')\,\mathrm{d}x \quad (1)\end{aligned}$$

where x_t denotes the system state at time t, Ω is the state space, ρ_t represents the probability density function of x_t, and $\mathcal{P}$ represents the Markov propagator. For many real-world physical and chemical systems, e.g., conformational transitions in macromolecules (Noé and Fischer, 2008), autocatalytic chemical reactions (Biancalani et al., 2012) and climate changes (Berglund and Gentz, 2002), it is common and natural to further assume that $\{x_t\}$ is a time-reversible and metastable process. The reversibility means that $p(x_t = x', x_{t+\tau} = x) = p(x_t = x, x_{t+\tau} = x')$ and generally arises from the time symmetries of classical mechanics, thermodynamics and quantum mechanics, and the metastability of a dynamical system means that the state space of the system can be decomposed into a set of macrostates called *metastable states* so that the local equilibrium within a metastable state can be reached quickly and the transitions between different metastable states can only be observed on slow timescales. A large number of recent studies in statistical physics indicate that the dominant spectral components (or called dominant eigenpairs, i.e., the largest eigenvalues and the associated eigenfunctions) of the Markov propagator is a key to understand and characterize such a process, because they can provide a lot of essential and useful information for the computation of ensemble averages and correlation functions (Noé et al., 2011), detection of spatial structures of metastable states (Deuflhard and Weber, 2005), choice of reaction coordinates (Rohrdanz et al., 2011; Perez-Hernandez et al., 2013), and construction of low-dimensional approximate models (Kube and Weber, 2007; Noé and Nüske, 2013).

However, directly solving the eigenvalue problem of the Markov propagator is generally impossible except for some extremely simple cases (e.g., Ornstein-Uhlenbeck process), and the dominant spectral components can only be estimated from simulation or experimental data through statistical inference and numerical computation. The most

popular and successful method for the spectral estimation is the Markov state model (MSM) method (Prinz et al., 2011; Djurdjevac et al., 2010), which discretizes the state space into a set of discrete bins and calculates the dominant spectral components in a finite element manner by assuming transitions between the bins are Markovian. Obviously, the main difficulty of this method is the choice of the discretization. On the one hand the Markov assumption will be severely violated if the discretization is too coarse, and on the other hand too many bins may cause the problem of "curse of dimensionality" in the estimation of transition probabilities. A more general method is the variational method (Noé and Nüske, 2013; Nüske et al., 2013), which allows one to perform the spectral estimation by using "soft bins" defined by a set of smooth basis functions instead of the "crisp bins" used in MSMs. Numerical experiments show that the variational method can achieve more accurate estimation than the MSM method with the same number of bins. However, there is no systematic algorithm for the choice of basis functions, and the basis function set can only be determined by trial and error in practice. Moreover, in some literature, the diffusion maps is used to identify dominant spectral components in a nonparametric manner by treating each sample point as a discrete bin (Rohrdanz et al., 2011; Ferguson et al., 2011), but this method is applicable only if the Markov propagator is defined by a Brownian dynamics.

In (Noé et al., 2013; Prinz et al., 2014), a novel framework call projected Markov model (PMM) is proposed for spectral analysis of metastable processes without the Markov assumption on discrete bins. Within this framework, it is shown that if a metastable Markov process contains only m nonzero eigenvalues then the corresponding coarse-grained dynamics on the space of discrete bins is equivalent to an m-state hidden Markov model (HMM), and the equivalence is independent of the choice of the discretization. Then HMM learning methods can be utilized to identify dominant eigenvalues and projected eigenfunctions efficiently and effectively even in the case that the investigated system is only experimentally observable and some important dimensions of the system state cannot be directly observed. (See more details in Subsection 2.1.) The main disadvantage of the PMM approach is that the estimation performance strongly depends on the choice of m, and a small change of the value of m may lead to a great error on the estimation of spectral components because of the orthogonality of eigenfunctions (Noé et al., 2013).

The aim of this paper is to propose an infinite HMM based method to solve the spectral estimation problem of metastable processes with unknown dominant spectra. Infinite HMMs (Teh et al., 2006; Teh and Jordan, 2010; Paisley and Carin, 2009; Fox et al., 2011) are a generalization of classical HMMs, which contains infinite hidden states and provide a powerful tool for nonparametric dynamical modeling of sequential data. In contrast with classical HMMs, infinite HMMs encourage sparse utilization of infinite state sets through defining suitable prior models, and can be used to infer both model parameters and state numbers from observation data in a pure Bayesian manner. However, our investigation (see Section 3) shows that a sparse prior on hidden states cannot guarantee that the eigenvalue set also has a sparse structure, and the spectral estimation is an "ill-posed" problem for the existing infinite HMMs. In this paper, we construct a new infinite HMM named stick-breaking half-weighted model (SB-HWM), which has a sparse prior distribution on eigenvalues and tends to approximate the underlying dynamics of an unknown system with a small number of dominant spectral components. Moreover, we develop a sampling inference algorithm for applying SB-HWMs to Bayesian nonparametric inference of spectral components.

The rest of the paper is organized as follows. In Section 2, we present the relevant mathematical background on PMMs and infinite HMMs, then Section 3 outlines the Bayesian nonparametric framework for solving the spectral estimation problem and explains the reason why the existing infinite HMMs cannot be directly applied. In Section 4 we introduce the SB-HWM and its sampling inference algorithm. Section 5 demonstrates through simulations the effectiveness of the proposed model and algorithm.

2 BACKGROUND

2.1 COARSE-GRAINED DYNAMICS AND PROJECTED MARKOV MODELS

Let $\{x_t\}$ be a Markov process with propagator $\mathcal{P}$ and state space Ω as in (1) and $\{y_t\}$ is the corresponding observation process obtained from the spatial coarse-graining

$$\Pr\left(y_t = k | x_t = x\right) = \chi_k\left(x\right), \quad k \in \mathcal{O} \tag{2}$$

where $\mathcal{O} = \{1, \ldots, K\}$ denotes the discrete observation space and $\chi_k\left(x\right)$ denotes the observation probability function for the observed value k. Often, the coarse-graining is employed by the Galerkin discretization and $\{\chi_k\left(x\right)\}$ is a set of indicator functions with each k representing a finite element space $\{x | x \in \Omega, \chi_k\left(x\right) = 1\}$. But in some practical cases, e.g. where $\{y_t\}$ obtained from noisy measurements, each $\chi_k\left(x\right)$ is a continuous probability density function and characterizes a soft finite element space.

It is obvious that (1) and (2) is in fact an HMM, but it is infeasible to reconstruct $\mathcal{P}$ from $\{y_t\}$ by direct statistical inference because of the continuity of Ω and the complexity of the dynamics of $\{x_t\}$ in general cases. In order to overcome this difficulty, the PMM (Noé et al., 2013) provides a low-dimensional approximation of the coarse-grained dynamics based on the following metastability assumption:

Assumption 1. *$\{x_t\}$ is ergodic and reversible w.r.t. the*

unique stationary distribution $\mu(x)$, and there is a τ' such that $\mathcal{P}(\tau')$ has only m eigenvalues which are not close to 0.

Note that this assumption holds for most practical metastable systems and m is usually a small number depends on the number of metastable states in Ω.[1] Under this assumption, we can conclude that $\mathcal{P}(\tau)$ is a compact and self-adjoint operator w.r.t. the inner product inner product $\langle\cdot,\cdot\rangle_{\mu^{-1}}$ defined by

$$\langle u_1, u_2\rangle_{\mu^{-1}} = \int \frac{u_1(x)\, u_2(x)}{\mu(x)} \mathrm{d}x \tag{3}$$

and the dynamics of $\{x_t\}$ can be decomposed as

$$\rho_{t+\tau} = \sum_{i=1}^{m} \lambda_i(\tau) \langle \rho_t, \phi_i\rangle_{\mu^{-1}} \phi_i + \mathcal{P}_{\text{fast}}(\tau)\, \rho_t \tag{4}$$

with

$$\lambda_i(\tau) = \exp(-\kappa_i \tau) \tag{5}$$

Here $\lambda_i(\tau)$ denotes the i-th largest magnitude eigenvalue of $\mathcal{P}(\tau)$ with eigenfunction ϕ_i and decay rate $\kappa_i \geq 0$ ($\kappa_1 = 0 < \kappa_2$ and $\phi_1 = \mu$ due to the ergodicity). The operator $\mathcal{P}_{\text{fast}}(\tau)$ consists of spectral components of $\mathcal{P}(\tau)$ which decay to zero quickly and $\|\mathcal{P}_{\text{fast}}(\tau)\| \approx 0$ for $\tau \geq \tau'$. Omitting the second term on the r.h.s. of (4), the correlation matrix $\mathbf{C}(n\tau) = [c_{ij}(n\tau)] = [\Pr(y_t = i, y_{t+n\tau} = j)]$ of $\{y_t\}$ can be decomposed as

$$\mathbf{C}(n\tau) = \mathbf{Q}\mathbf{\Lambda}(\tau)^n \mathbf{Q}^\intercal \tag{6}$$

where $\mathbf{\Lambda}(\tau) = \operatorname{diag}(\lambda_1(\tau), \ldots, \lambda_m(\tau))$ contains the dominant eigenvalues of $\mathcal{P}(\tau)$, and the i-th column of $\mathbf{Q} \in \mathbb{R}^{K\times m}$ is the i-th projected eigenfunction

$$\mathbf{q}_i = (\int \chi_1(x)\, \phi_i(x)\, \mathrm{d}x, \ldots, \int \chi_K(x)\, \phi_i(x)\, \mathrm{d}x)^\intercal \tag{7}$$

Therefore, we can characterize the coarse-grained dynamics of $\{y_t\}$ by low-dimensional PMM variables $\{\mathbf{Q}, \mathbf{\Lambda}(\tau)\}$ on a large timescale $\tau \geq \tau'$.

It is important to point out that we can also get a similar approximation of $\mathbf{C}(n\tau)$ by using a m-state HMM. Assume that $\{y_t\}$ are observations of an HMM with hidden states $\{s_t\}$, state set $\{1, \ldots, m\}$, transition matrix $\mathbf{A} = [a_{ij}] = [\Pr(s_{t+\tau} = j | s_t = i)]$ and observation matrix $\mathbf{B} = [b_{ij}] = [\Pr(y_t = j | s_t = i)]$, then $\mathbf{C}(n\tau)$ can be expressed as

$$\begin{aligned} \mathbf{C}(n\tau) &= \mathbf{B}^\intercal \operatorname{diag}(\boldsymbol{\pi})\, \mathbf{A}^n \mathbf{B} \\ &= (\mathbf{B}^\intercal \mathbf{L})\, \tilde{\mathbf{\Lambda}}^n\, (\mathbf{B}^\intercal \mathbf{L})^\intercal \end{aligned} \tag{8}$$

under the condition[2] that $\mathbf{A}$ is a reversible transition matrix w.r.t. the stationary distribution $\boldsymbol{\pi}$, where $\tilde{\mathbf{\Lambda}}$ is a diagonal matrix containing eigenvalues of $\mathbf{A}$, $\mathbf{L}$ consists of left eigenvectors of $\mathbf{A}$ with $\mathbf{L}^\intercal \mathbf{A} = \tilde{\mathbf{\Lambda}} \mathbf{L}^\intercal$ and $\mathbf{L}^\intercal \operatorname{diag}(\boldsymbol{\pi})^{-1} \mathbf{L} = \mathbf{I}$, and $\mathbf{I}$ denotes the identity matrix. Based on the similarity between (6) and (8), the PMM theory provides the following conclusion: *Under the metastability assumption (Assumption 1) with $\|\mathcal{P}_{\text{fast}}(\tau)\| = 0$ and some technical assumptions, the dynamics of $\{y_t\}$ is equivalent to a m-state HMM with a reversible transition matrix.* Thus, if a suitable m is given, we can utilize HMM learning algorithms to efficiently estimate the dominant eigenvalues and projected eigenfunctions of $\mathcal{P}(\tau)$ from $\{y_t\}$ with $\mathbf{Q} = \mathbf{B}^\intercal \mathbf{L}$ and $\mathbf{\Lambda}(\tau) = \tilde{\mathbf{\Lambda}}$. However, the choice of m is still an unsatisfactorily solved problem for the PMM method, and the numerical experiments in (Noé et al., 2013) show that the estimation results of the PMM method is very sensitive to the value of m.

2.2 STICK-BREAKING PROCESSES AND INFINITE HIDDEN MARKOV MODELS

Roughly speaking, a stick-breaking process (SBP) (Ishwaran and James, 2001) is a prior for discrete distributions, and the realization of an SBP with parameters α', α and base distribution G_0 can be expressed by the following probability density function:

$$G = \sum_{i=1}^{\infty} w_i \delta_{\theta_i} \tag{9}$$

where δ_θ denotes the Dirac point measure concentrated on θ, θ_i is the i-th component of the discrete distribution with $\theta_i \overset{\text{iid}}{\sim} G_0$, and w_i denotes the corresponding weight which are drawn by $w_i = V_i \prod_{j=1}^{i-1}(1 - V_j)$ and $V_i \overset{\text{iid}}{\sim} \operatorname{Beta}(\alpha', \alpha)$. Obviously, the SBP model is a generalization of the finite-dimensional Dirichlet distribution (Gelman et al., 2003), and allows one to easily construct discrete distributions with infinite components. For convenience of computation and notation, in this paper we only consider a special class[3] of SBPs with $\alpha' = 1$, and denote by $\operatorname{DP}(\alpha, G_0)$ and $\operatorname{GEM}(\alpha)$ the prior distributions of G and $\{w_i\}$ defined in (9).

The SBP model provides a powerful and flexible tool for nonparametric estimation of multi-modal mixture models, and can be applied to building HMMs with infinite

[1] Generally speaking, a stochastic system with m metastable states only has m eigenvalues which are significantly larger than zero on a large timescale. It is worth pointing out that this assumption of sparse spectrum is a very important basis in the research of metastability (see, e.g., (Deuflhard and Weber, 2005; Djurdjevac et al., 2010; Noé and Nüske, 2013; Prinz et al., 2014))), and a large number of studies have shown the validity of this assumption for common physical processes which exhibits metastability.

[2] This condition means $\operatorname{diag}(\boldsymbol{\pi})\mathbf{A}$ is a symmetric matrix, which is a sufficient condition for reversibility of $\{y_t\}$.

[3] An SBP with $\alpha' = 1$ is equivalent to a Dirichlet process (Ferguson, 1973).

states for sequential statistical modeling. The most commonly used infinite HMM is the HDP-HMM (Teh et al., 2006), which constructs prior distributions of the infinite-dimensional transition matrix $\mathbf{A} = [a_{ij}]$ and observation matrix $\mathbf{B} = [b_{ij}]$ by organizing multiple SBPs in a hierarchical structure as

$$\begin{array}{rcl} G_0 = \sum_{k=1}^{\infty} \beta_k \delta_{\mathbf{b}_k} & \sim & \mathrm{DP}\,(\gamma, H) \\ G_i = \sum_{j=1}^{\infty} a_{ij} \delta_{\mathbf{b}_i} & \overset{\mathrm{iid}}{\sim} & \mathrm{DP}\,(\alpha, G_0) \end{array} \tag{10}$$

where $\mathbf{b}_i$ denotes the i-th row of $\mathbf{B}$ and represents the observation probability distribution of the i-th state, H represents the prior distribution of each $\mathbf{b}_i$ and is usually a Dirichlet distribution, and α, γ are hyperparameters. In (Fox et al., 2011), a modified HDP-HMM called "sticky HDP-HMM" is proposed to encourage large self-transition probabilities and and avoids "unphysically" fast switching between different states, which can be expressed as

$$\begin{array}{rcl} G_0 = \sum_{k=1}^{\infty} \beta_k \delta_{\mathbf{b}_k} & \sim & \mathrm{DP}\,(\gamma, H) \\ G_i = \sum_{j=1}^{\infty} a_{ij} \delta_{\mathbf{b}_i} & \overset{\mathrm{ind}}{\sim} & \mathrm{DP}\left(\alpha + \kappa, G_0^{(i)}\right) \end{array} \tag{11}$$

with

$$G_0^{(i)} = \frac{\alpha G_0 + \kappa \delta_{\mathbf{b}_i}}{\alpha + \kappa} \tag{12}$$

where $\kappa > 0$ is the sticky factor and $\lim_{\kappa \to \infty} a_{ii} = 1$. Furthermore, it is worthwhile to point out that for most of the SBP based infinite HMMs, including HDP-HMM, sticky HDP-HMM and stick-breaking HMM proposed in (Paisley and Carin, 2009), the infinite-dimensional prior distributions can be approximated by high- but finite-dimensional ones for convenience of implementing sampling inference.

3 BAYESIAN NONPARAMETRIC FRAMEWORK FOR SPECTRAL ESTIMATION

The main purpose of this paper is to develop a Bayesian nonparametric framework for spectral estimation of metastable Markov processes with unknown number m of dominant eigenpairs. In the rest of paper, unless otherwise stated, the lagtime τ is set to be fixed, and $\{x_t\}$ and $\{y_t\}$ are separately defined as $\{x_{n\tau}\}_{n=0}^{N}$ and $\{y_{n\tau}\}_{n=0}^{N}$.

Suppose that $\{x_t\}$ is a metastable process with the available observation process $\{y_t\}$ as described in Subsection 2.1 and $\{x_t\}$ satisfies Assumption 1. Based on the discussion in Section 2, the Bayesian estimation of the i-th largest eigenvalue λ_i and the corresponding projected eigenfunction $\mathbf{q}_i$ of the Markov propagator $\mathcal{P}(\tau)$ of $\{x_t\}$ can be achieved by the following steps with m not given a priori: First, the dynamics of $\{y_t\}$ is described by an infinite HMM consisting of a infinite-dimensional transition matrix $\mathbf{A}$ and observation matrix $\mathbf{B}$ with prior $p(\mathbf{A}, \mathbf{B})$. Second, a large number of samples $\{(\mathbf{A}^{(k)}, \mathbf{B}^{(k)})\}$ of $(\mathbf{A}, \mathbf{B})$ are drawn from the posterior distribution

$$\begin{array}{rcl} & & p(\mathbf{A}, \mathbf{B}|\{y_t\}) \\ & \propto & p(\mathbf{A}, \mathbf{B}) \sum_{\{s_t\}} p(\{s_t\}|\mathbf{A})\, p(\{y_t\}|\{s_t\}, \mathbf{B}) \end{array} \tag{13}$$

where s_t denotes the discrete hidden state of the infinite HMM at time t. Finally, the i-th eigenvalue $\lambda_i^{(k)}$ and left eigenvector $\mathbf{l}_i^{(k)\intercal}$ of $\mathbf{A}^{(k)}$ are calculated for each k such that the posterior distribution of $(\lambda_i, \mathbf{q}_i)$ can be approximated by the ensemble $\{(\lambda_i^{(k)}, \mathbf{q}_i^{(k)})\}$ with $\mathbf{q}_i^{(k)} = \mathbf{B}^{(k)\intercal}\mathbf{l}_i^{(k)}$.

It is natural for us to utilize one of infinite HMMs such as the HDP-HMM and sticky HDP-HMM mentioned in Subsection 2.2 to design the prior distribution of $(\mathbf{A}, \mathbf{B})$ within the above framework. However, the following simple example shows that the existing SBP based infinite HMMs are *not* applicable to the spectral estimation problem.

Example 2. Let $\{s_t\} = \{s_{n\tau}\}_{n=0}^{1000}$ be a realization of a reversible 3-state Markov chain with transition matrix

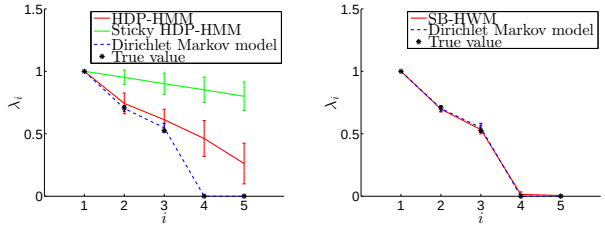
$$\mathbf{A}_0 = \begin{bmatrix} 0.8462 & 0.0769 & 0.0769 \\ 0.1250 & 0.7500 & 0.1250 \\ 0.1818 & 0.1818 & 0.6364 \end{bmatrix} \tag{14}$$

It is clear that $\{s_t\}$ is a Markov chain with large self-transition probabilities and has only three nonzero eigenvalues. We use the prior models of infinite-dimensional transition matrices defined in the HDP-HMM and sticky HDP-HMM to approximate the first 5 eigenvalues of $\{s_t\}$ based on the posterior distribution

$$\begin{array}{rcl} p(\mathbf{A}|\{s_t\}) & \propto & p(\mathbf{A})\, p(\{s_t\}|\mathbf{A}) \\ & \propto & p(\mathbf{A}) \prod_{n=1}^{1000} a_{s_{(n-1)\tau}, s_{n\tau}} \end{array} \tag{15}$$

Fig. 1a illustrates the estimation results obtained by the Markov chain Monte Carlo (MCMC) sampling. It can be observed that both the HDP-HMM and the sticky HDP-HMM give poor estimates of eigenvalues and fail to detect the spectral gap between λ_3 and λ_4.

In a strict sense, the estimation problem in Example 2 is not an "HMM problem" since the hidden state sequence $\{s_t\}$ is exactly known, rather, it is an effective toy example for illustrating the difficulty of nonparametric spectral estimation for the existing infinite HMMs. Roughly speaking, each infinite HMM provides a "sparse" prior distribution of the stationary distribution $\boldsymbol{\pi} = [\pi_i]$ of the transition matrix $\mathbf{A}$, which means for most samples of $\boldsymbol{\pi}$ we can find a small set $\mathcal{S}$ of hidden states such that $\sum_{i \notin \mathcal{S}} \pi_i \approx 0$. Thus, there are only a small number of distinct hidden states that can be detected by the Bayesian inference in general although the prior model contains infinite states. However, the sparsity of $\boldsymbol{\pi}$ cannot guarantee the sparsity of the eigenvalue set, because the transition dynamics between hidden



(a) HDP-HMM and sticky HDP-HMM

(b) SB-HWM

Figure 1: Estimation results of the first 5 eigenvalues of $\{s_t\}$ based on different prior models of $\mathbf{A}$, where error bars represent one standard deviation confidence intervals, and dashed lines represent the estimation results obtained by assuming that $\mathbf{A}$ is a transition matrix with size 3×3 and the prior of each row of $\mathbf{A}$ is $\mathrm{Dir}\,(1/3, 1/3, 1/3)$. Truncated-model-based samplers (see Subsection 4.3 and (Fox et al., 2008)) are applied to sampling $\mathbf{A}$ of infinite HMMs, which approximate $\mathbf{A}$ by a finite-dimensional matrix with size 20×20, and the posterior means and standard deviations are calculated from 5000 MCMC samples with 5000 burn-in samples.

states with small stationary probabilities may also contain large eigenvalues. This is also the reason why the infinite HMMs overestimate the eigenvalues in Example 2. (In fact, for any stationary distribution $\boldsymbol{\pi}$ and $i > 0$ we can construct a sequence of matrices $\{\mathbf{A}^{(k)}\}$ which are reversible w.r.t. $\boldsymbol{\pi}$ and satisfy $\lim_{k\to\infty} \lambda_i^{(k)} \to 1$. See the supplementary material for details.) Note that in contrast with the HDP-HMM, the sticky HDP-HMM encourages longer residence time for each state and tends to generate more "pseudo-dominant eigenvalues", so it performs worse than the HDP-HMM in this example. Furthermore, it is difficult for both HDP-HMM and sticky HDP-HMM to incorporate the reversibility constraint.

In order to overcome disadvantages of existing infinite HMMs in the application of spectral estimation, we present in next section a novel infinite HMM, which approximates the "half-weighted matrix" instead of the transition matrix in a nonparametric way and can provide a sparse prior for eigenvalues.

4 STICK-BREAKING HALF-WEIGHTED MODELS

4.1 HALF-WEIGHTED MATRICES

Before developing our infinite HMM for spectral estimation, we first introduce the definition and some important properties of half-weighted matrices for the purpose of self-containedness. For a Markov chain with transition matrix $\mathbf{A} = [a_{ij}]$ and stationary distribution $\boldsymbol{\pi} = [\pi_i]$, the half-weighted matrix $\mathbf{H} = [h_{ij}]$ is defined by[4]

$$\mathbf{H} = \mathrm{diag}\,(\boldsymbol{\pi})^{\frac{1}{2}}\, \mathbf{A} \mathrm{diag}\,(\boldsymbol{\pi})^{-\frac{1}{2}} \tag{16}$$

(Note dimensions of $\mathbf{A}$, $\boldsymbol{\pi}$ and $\mathbf{H}$ may be infinite here.)

The following two theorems summarize important properties of the half-weighted matrix and provide a criterion for checking if a matrix is a valid half-weighted matrix. (The proofs are in the supplementary material.)

Theorem 3. *If $\mathbf{A}$ is a reversible and positive transition matrix and all eigenvalues $\{\lambda_i\}$ of $\mathbf{A}$ are square summable, then (1) $\mathbf{H}$ is a positive and symmetric matrix. (2) $\|\mathbf{H}\|_F < \infty$. (3) $\mathbf{H}$ and $\mathbf{A}$ have the same eigenvalues, and the i-th eigenvector $\boldsymbol{\psi}_i$ of $\mathbf{H}$ and the i-th left eigenvector $\mathbf{l}_i$ of $\mathbf{A}$ satisfy $\boldsymbol{\psi}_i = \mathrm{diag}\,(\boldsymbol{\pi})^{-\frac{1}{2}} \mathbf{l}_i$. (4) $\sum_{i=m+1}^{\infty} \lambda_i^2 \le \sum_{i>m \vee j>m} h_{ij}^2$ for all $m > 1$.*

Theorem 4. *If $\mathbf{H}$ is a positive and symmetric matrix with $\|\mathbf{H}\|_F < \infty$, and the spectral radius of $\mathbf{H}$ is 1, then $\mathbf{H}$ is a half-weighted matrix of a Markov chain.*

According to (16), the likelihood of a half-weighted matrix $\mathbf{H}$ of a given state sequence $\{s_t\} = \{s_{n\tau}\}_{n=0}^{N}$ is

$$p\,(\{s_t\}|\mathbf{H}) = p\,(s_0) \sqrt{\frac{\pi_{s_{N\tau}}}{\pi_{s_0}}} \prod_{n=1}^{N} h_{s_{(n-1)\tau}, s_{n\tau}} \tag{17}$$

From the above, it can be seen that the half-weighted matrix $\mathbf{H}$ can be used to describe the dynamics of state transitions instead of $\mathbf{A}$, and $\mathbf{H}$ is more numerically stable for eigenvalue decomposition than $\mathbf{A}$ due to the symmetry of $\mathbf{H}$. Moreover, it is interesting to observe that (17) is in fact a Boltzmann chain model (Saul and Jordan, 1995) with the transition energy from state i to state j being $-\ln h_{ij}$.

4.2 MODEL DEFINITION

From the fourth property of half-weighted matrices stated in Theorem 3, it can be seen that if a Markov chain has a half-weighted matrix with all elements except the ones in a small number of rows and columns are close to zero, then there are only a few eigenvalues of the Markov chain that can be significantly larger than zero. This suggests a natural way of constructing a prior distribution over infinite HMMs which encourages the sparsity of eigenvalue sets and satisfies the reversibility.

Based on the above discussion, we now propose the following infinite HMM called stick-breaking half-weighted model (SB-HWM) for spectral estimation:

$$\begin{aligned} \mathbf{H} &= \frac{1}{r}\bar{\mathbf{H}} \\ \mathbf{B} &= (\mathbf{b}_1^\intercal, \mathbf{b}_2^\intercal, \ldots)^\intercal \end{aligned} \tag{18}$$

[4]The definition of half-weighted matrix is in fact a discrete version of the "half-weighted correlation density" proposed in (Noé and Nüske, 2013) for analysis of Markov processes.

with

$$\begin{aligned} \mathbf{w} = [w_i] &\sim \mathrm{GEM}\left(\alpha_w\right) \\ \gamma_{ij} &\overset{\text{iid}}{\sim} \mathrm{Gamma}\left(\alpha_\gamma, \beta_\gamma\right), \quad \text{for } i \geq j \\ \gamma_{ji} &= \gamma_{ij}, \quad \text{for } i < j \\ w_i^d &\overset{\text{iid}}{\sim} \mathrm{Gamma}\left(\alpha_d, \beta_d\right) \\ \bar{\mathbf{H}} = [\bar{h}_{ij}] &= [\gamma_{ij}\left(w_i w_j + w_i w_i^d \cdot 1_{i=j}\right)] \end{aligned} \tag{19}$$

and

$$\mathbf{b}_i \overset{\text{iid}}{\sim} \mathrm{Dir}\left(\alpha_b, \ldots, \alpha_b\right) \tag{20}$$

where r denotes the spectral radius of the "unnormalized half-weighted matrix" $\bar{\mathbf{H}}$, $(\alpha_w, \alpha_\gamma, \beta_\gamma, \alpha_d, \beta_d, \alpha_b)$ are hyperparameters, and it is easy to verify by Theorem 4 that the realization of $\mathbf{H}$ is a valid half-weighted matrix with probability 1. Note that $\bar{\mathbf{H}}$ can be expressed in a more compact form as

$$\bar{\mathbf{H}} = \mathbf{\Gamma} \circ \left(\mathbf{w}\mathbf{w}^\intercal + \mathrm{diag}\left(\mathbf{w} \circ \mathbf{w}^d\right)\right) \tag{21}$$

where $\mathbf{\Gamma} = [\gamma_{ij}]$, $\mathbf{w}^d = [w_i^d]$ and $\circ$ denotes the element-wise product. It can be seen that $\mathbf{w}$ employs a "template vector" to encourage rows and columns of the half-weighted matrix to have the similar sparse structures[5]. Furthermore, it is known that for a metastable Markov process, the hidden states of the equivalent HMM mentioned in Subsection 2.1 often have long residence times since they arise from metastable states of the original process (Noé et al., 2013). So we use $\mathbf{w}^d$ to enhance probabilities of self-transitions of hidden states, which plays the similar role as the sticky factor κ in (11). The following theorem gives a theoretical description of the sparsity of $\boldsymbol{\pi}$ and eigenvalue set $\{\lambda_i\}$ in the SB-HWM (see the supplementary material for the proof):

Theorem 5. *For an SB-HWM $(\mathbf{H}, \mathbf{B})$ generated by the prior defined by (18)-(20), the i-th largest magnitude eigenvalue λ_i and the stationary probability π_i of the i-th hidden satisfy $\mathbb{E}\left[|\lambda_i|\right] = O\left(\left(\frac{\alpha_w}{1+\alpha_w}\right)^{\frac{i}{3}}\right)$ and $\mathbb{E}\left[\pi_i\right] = O\left(\left(\frac{\alpha_w}{1+\alpha_w}\right)^{\frac{i}{3}}\right)$ as $i \to \infty$.*

As a comparison, we also apply the SB-HWM to the data in Example 2 and the estimation results are shown in Fig. 1b. (See Subsection 4.3 for the sampling algorithm.) It can be seen that the SB-HWM achieves the similar estimation performance as the HMM with state number given, and the correct number of dominant spectral components can be easily obtained from samples of the SB-HWM.

[5]For example, if w_i is about zero, the elements in the i-th row and column of $\mathbf{H}$ will also be close to zero with high probabilities.

4.3 SAMPLING INFERENCE

For convenience of computation, we first construct a truncated model to approximate the SB-HWM by replacing the prior distribution of $\mathbf{w}$ in (19) with the following truncated SBP prior (Ishwaran and James, 2002):

$$w_i = V_i \prod_{j=1}^{i-1} \left(1 - V_j\right) \tag{22}$$

with

$$\begin{cases} V_i \overset{\text{iid}}{\sim} \mathrm{Beta}\left(1, \alpha\right), & i < L \\ V_i = 1, & i = L \end{cases} \tag{23}$$

The truncated model is obviously a finite HMM with L states since $w_i = 0$ for $i > L$, and according to Theorem 5, the influence of the truncation on the dynamics of SB-HWM is slight if L is sufficiently large[6]. Then the Markov chain Monte Carlo approach can be utilized to draw samples of $(\mathbf{H}, \mathbf{B})$ from the posterior distribution $p\left(\mathbf{H}, \mathbf{B} | \{y_t\}\right)$ based on the truncated prior. Considering that the presented prior distribution of $\mathbf{H}$ is not a conjugate distribution for the state sequence, here we combine the Metropolis-within-Gibbs algorithm with the block sampling algorithm of classical HMMs to generate samples (see the supplementary material for details) based on the assumption that $\{s_t\}$ is a stationary process, i.e., $p\left(s_0\right) = \pi_{s_0}$.

5 APPLICATIONS

In this section, we demonstrate the performance of the SB-HWM based Bayesian spectral estimation method on three examples of stochastic systems including an HMM, a diffusion process governed by a Brownian dynamics and the molecular dynamics of alanine dipeptide. The detailed settings of simulations and estimation algorithms are provided in the supplementary material.

5.1 HMM DATA

Here we apply the SB-HWM to the simulation data generated by a 3-state HMM with lagtime $\tau = 1$ and 8, and compare its performance with HDP-HMM and sticky HDP-HMM. Estimation results are summarized in Fig. 2. Obviously, both HDP-HMM and sticky HDP-HMM severely overestimate the eigenvalues and result in large errors in estimation of projected eigenfunctions, because their samples contain a lot of "pseudo-dominant spectral components" as mentioned in Section 3. (All the three models achieve small

[6]According to our experience, the empirical performance of truncated SB-HWMs is not sensitive to the choice of L if it is larger than twice or three times of the number of the dominant eigenvalues. We simply set L to be 20 in experiments of this paper, and the theoretical analysis of the truncation error will be published elsewhere.

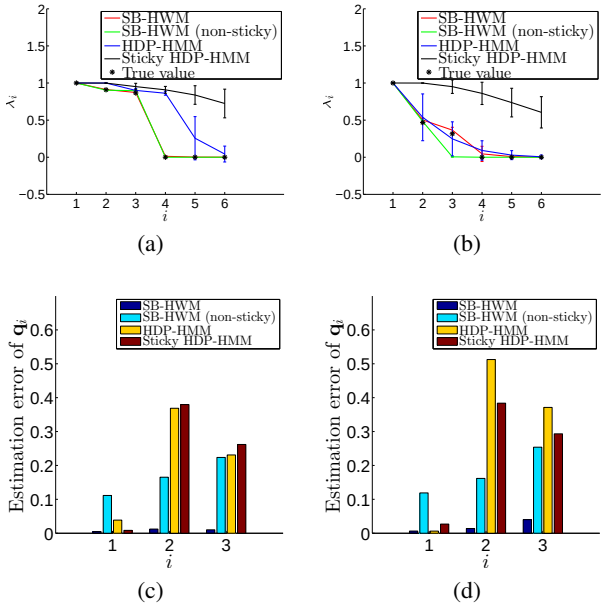


Figure 2: Infinite HMMs applied to data from an HMM. (a,b) Estimates of the first 6 eigenvalues with $\tau = 1$ and 8, where error bars represent one standard deviation confidence intervals. (c,d) Estimation errors of the first 3 projected eigenfunctions with $\tau = 1$ and 8, where the error between the estimate $\hat{\mathbf{q}}_i$ and the true value $\mathbf{q}_i$ is defined by $\|\hat{\mathbf{q}}_i - \mathbf{q}_i\|$.

estimation errors on $\mathbf{q}_1$ as it can easily be estimated as the stationary distribution of $\{y_t\}$.) Of the above three models, only the SB-HWM provides accurate estimates of eigenvalues and dominant eigenfunctions, which allows us to correctly detect the spectral gap and total number of dominant spectral components. Furthermore, we apply a specific SB-HWM with $w_i^d \equiv 0$ for all i to the HMM data in order to verify the usefulness of the sticky term in (19), and the estimates obtained by the non-sticky SB-HWM are also shown in Fig. 2 (see green lines and bars). It can be observed that the estimates of projected eigenfunctions obtained by the non-sticky SB-HWM are much worse than that obtained by the proposed SB-HWM and the non-sticky SB-HWM fails to identify the third dominant spectral component when applied to the HMM data with $\tau = 8$. The main reason for the poor performance of the non-sticky SB-HWM is that it tends to underestimate residence times of hidden states which are key parameters affecting the spectral properties especially for HMMs of metastable systems. (Note that $a_{ii} = O(w_i^2)$ as $i \to \infty$ in the non-sticky SB-HWM, whereas $a_{ii} = O(w_i)$ in the SB-HWM.)

5.2 BROWNIAN DYNAMICS DATA

In this subsection, we consider a two-dimensional system of Brownian dynamics on the domain $\Omega = [-2, 2] \times [-1.5, 2.5]$ with a three-well potential and a Galerkin discretization observation model which are depicted in Fig. 3. The three potential wells implies that the system contains the same number of metastable states and dominant spectral components. Fig. 4 plots spectral estimation results obtained by the SB-HWM, HDP-HMM and MSM, where the MSM estimates spectral components by simply assuming that each bin in Fig. 3 is a discrete state in a Markov chain. It is obvious that the discrete bins cannot accurately capture boundaries between the metastable states in this example, and the poor coarse-graining causes large estimation errors of eigenvalues. (The detailed theoretical analysis on the relationship between the spectral estimation error and the choice of the discretization is reported in (Sarich et al., 2010).) From Fig. 4, we can also see that the HDP-HMM performs even much worse than the simple MSM, which again demonstrates the difficulty of spectral estimation for the existing infinite HMMs. The SB-HWM significantly outperforms the other two models on the spectral estimation in this example.

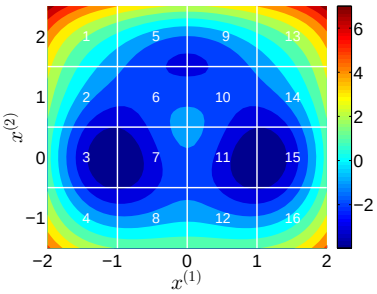


Figure 3: Illustration of the potential function and observation model of a Brownian dynamics system, where each grid represents a bin of the observation model.

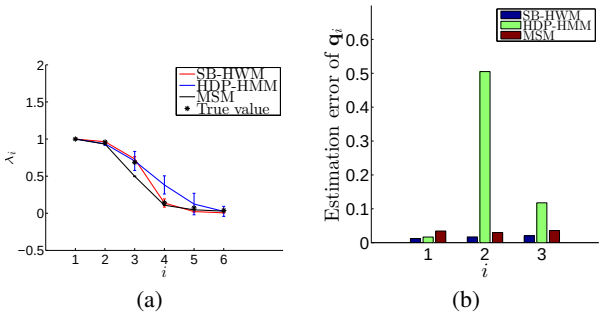


Figure 4: Infinite HMMs and an MSM applied to data from a Brownian dynamics simulation. (a) Estimates of the first 6 eigenvalues, where error bars represent one standard deviation confidence intervals. (b) Estimation errors of the first 3 projected eigenfunctions.

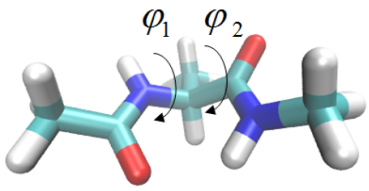


Figure 5: Illustration of the structure of alanine dipeptide

5.3 MOLECULAR DYNAMICS DATA

Alanine dipeptide (sequence acetyl-alanine-methylamide) is a small molecule which consists of two alanine amino acid units. The structural and dynamical properties of this molecule have been thoroughly studied, and it is well known that the configuration space of the alanine dipeptide can be conveniently described by two backbone dihedral angles (see Fig. 5) and contains three metastable states (see Fig. 6). We utilize the SB-HWM, HDP-HMM, 5-state MSM and 23-state MSM to perform the spectral estimation based on a molecular dynamics simulation with length 0.05 millisecond, where the discretization of all models are designed by using the kmeans algorithm and the first three models share the same discretization shown in Fig. 6. Moreover, for convenience of comparison, we construct a very finely discretized MSM with 129 states to estimate spectral components from a molecular dynamics simulation with length 1 millisecond, and use the corresponding estimates as "true values" in this example. It can be observed from Fig. 7 that the HDP-HMM cannot provide any valuable information on spectral components in this example except the first component, and the SB-HWM with observation space $\{1, \ldots, 5\}$ obviously outperforms the MSMs with 5 states and 23 states.

Fig. 8 shows the estimated eigenfunctions calculated according to the estimated projected eigenfunctions provided by the SB-HWM and 129-state MSM respectively. (The calculation details are give in the supplementary material.) By comparing them, it is interesting to note that the SB-HWM is able to well reconstruct dominant eigenfunctions in a low-dimensional function space, which also demonstrates the effectiveness of the proposed spectral estimation method.

6 CONCLUSION

We introduce in this paper a novel infinite HMM, "stick-breaking half-weighted model" (SB-HWM) for identification of dominant spectral components of metastable systems. The main idea is to construct a SBP based infinite-dimensional half-weighted matrix to describe transition dynamics of hidden states. In contrast with the other infinite HMMs, the SB-HWM provides a sparse prior on eigen-

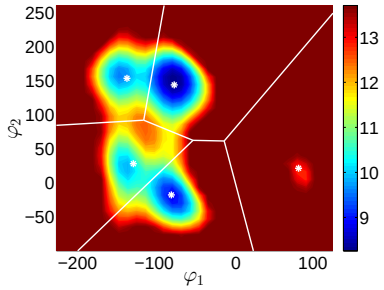
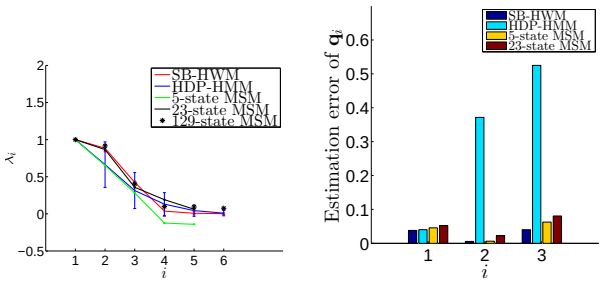


Figure 6: Free energy landscape in the state space of alanine dipeptide, where each grid represents a bin of the observation model.



(a) (b)

Figure 7: Infinite HMMs and MSMs applied to molecular dynamics data. (a) Estimates of the first 6 eigenvalues, where error bars represent one standard deviation confidence intervals. (Note the 5-state MSM has at most 5 nonzero eigenvalues.) (b) Estimation errors of the first 3 projected eigenfunctions.

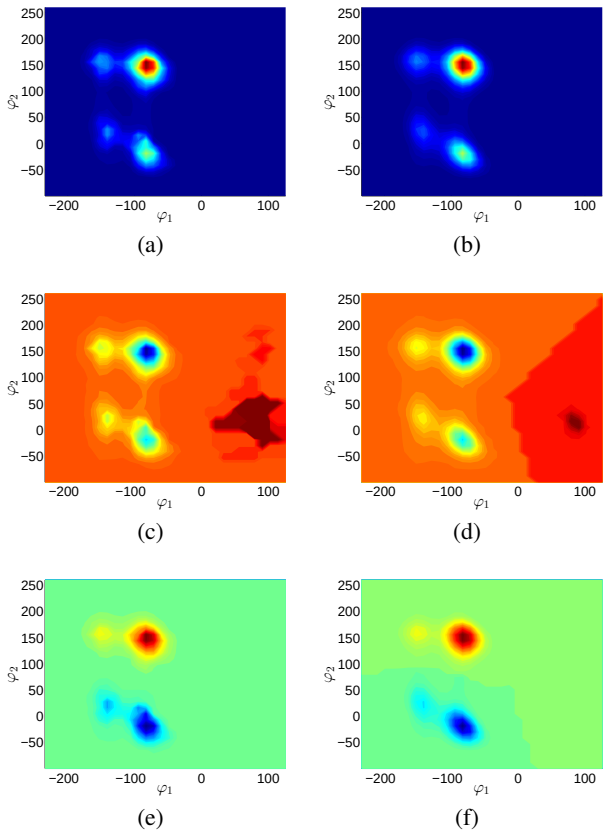


Figure 8: Estimates of the first three eigenfunctions l_1, l_2, l_3. (a,c,e) Estimates obtained by the SB-HWM. (b,d,f) Estimates obtained by the 129-state MSM.

values so that both the values and the numbers of dominant spectral components can be estimated by the Bayesian nonparametric inference. Furthermore, a truncated approximation based sampling inference algorithm for SB-HWMs is developed. Interesting directions of future research include developing a more efficient sampling algorithm and extending the algorithm to non-reversible systems.

ACKNOWLEDGMENT

This work is supported by the Deutsche Forschungsgemeinschaft (DFG) under grant No. WU 744/1-1.

References

Berglund, N. and B. Gentz, 2002: Metastability in simple climate models: Pathwise analysis of slowly driven langevin equations. *Stochastics and Dynamics*, **2 (03)**, 327–356.

Biancalani, T., T. Rogers, and A. McKane, 2012: Noise-induced metastability in biochemical networks. *Physical Review E*, **86 (1)**, 010 106.

Deuflhard, P. and M. Weber, 2005: Robust perron cluster analysis in conformation dynamics. *Linear algebra and its applications*, **398**, 161–184.

Djurdjevac, N., M. Sarich, and C. Schütte, 2010: On markov state models for metastable processes. *Proceedings of the International Congress of Mathematicians*, Hyderabad, 3105–3131.

Ferguson, A. L., A. Z. Panagiotopoulos, P. G. Debenedetti, and I. G. Kevrekidis, 2011: Integrating diffusion maps with umbrella sampling: Application to alanine dipeptide. *Journal of Chemical Physics*, **134 (13)**, 135 103.

Ferguson, T. S., 1973: A Bayesian analysis of some nonparametric problems. *The annals of statistics*, 209–230.

Fox, E. B., E. B. Sudderth, M. I. Jordan, and A. S. Willsky, 2008: An HDP-HMM for systems with state persistence. *Proceedings of the 25th international conference on Machine learning*, ACM, 312–319.

Fox, E. B., E. B. Sudderth, M. I. Jordan, and A. S. Willsky, 2011: A sticky HDP-HMM with application to speaker diarization. *Annals of Applied Statistics*, **5 (2A)**, 1020–1056.

Gelman, A., J. Carlin, H. Stern, and D. Rubin, 2003: *Bayesian data analysis*. 2d ed., CRC, Boca Raton.

Ishwaran, H. and L. James, 2002: Approximate dirichlet process computing in finite normal mixtures: Smoothing and prior information. *Journal of Computational and Graphical Statistics*, **11 (3)**, 1–26.

Ishwaran, H. and L. F. James, 2001: Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, **96 (453)**, 161–173.

Kube, S. and M. Weber, 2007: A coarse graining method for the identification of transition rates between molecular conformations. *Journal of Chemical Physics*, **126 (2)**, 024 103.

Noé, F., S. Doose, I. Daidone, M. Löllmann, M. Sauer, J. D. Chodera, and J. C. Smith, 2011: Dynamical fingerprints for probing individual relaxation processes in biomolecular dynamics with simulations and kinetic experiments. *Proceedings of the National Academy of Sciences*, **108 (12)**, 4822–4827.

Noé, F. and S. Fischer, 2008: Transition networks for modeling the kinetics of conformational change in macromolecules. *Current opinion in structural biology*, **18 (2)**, 154–162.

Noé, F. and F. Nüske, 2013: A variational approach to modeling slow processes in stochastic dynamical systems. *SIAM Journal on Multiscale Modeling and Simulation*, **11 (2)**, 635–655.

Noé, F., H. Wu, J.-H. Prinz, and N. Plattner, 2013: Projected and hidden Markov models for calculating kinetics and metastable states of complex molecules. *Journal of Chemical Physics*, **139 (18)**, 184 114.

Nüske, F., B. G. Keller, A. S. M. Guillermo Pérez-Hernández, and F. Noé, 2013: Variational approach to molecular kinetics. Journal of Chemical Theory and Computation.

Paisley, J. and L. Carin, 2009: Hidden Markov models with stick-breaking priors. *IEEE Transactions on Signal Processing*, **57 (10)**, 3905–3917.

Perez-Hernandez, G., F. Paul, T. Giorgino, G. De Fabritiis, and F. Noé, 2013: Identification of slow molecular order parameters for markov model construction. *Journal of Chemical Physics*, **139 (1)**, 015 102.

Prinz, J.-H., J. D. Chodera, and F. Noé, 2014: Spectral rate theory for two-state kinetics. *Physical Review X*, **4**, 011 020.

Prinz, J.-H., et al., 2011: Markov models of molecular kinetics: Generation and validation. *Journal of Chemical Physics*, **134**, 174 105.

Rohrdanz, M. A., W. Zheng, M. Maggioni, and C. Clementi, 2011: Determination of reaction coordinates via locally scaled diffusion map. *Journal of Chemical Physics*, **134**, 124 116.

Sarich, M., F. Noé, and C. Schütte, 2010: On the approximation quality of markov state models. *Multiscale Modeling & Simulation*, **8 (4)**, 1154–1177.

Saul, L. K. and M. I. Jordan, 1995: Boltzmann chains and hidden Markov models. *Advances in Neural Information Processing Systems*, 435–442.

Teh, Y., M. Jordan, M. Beal, and D. Blei, 2006: Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, **101 (476)**, 1566–1581.

Teh, Y. W. and M. I. Jordan, 2010: *Bayesian Nonparametrics: Principles and Practice*, chap. Hierarchical Bayesian nonparametric models with applications. Cambridge University Press, Cambridge.