ELSEVIER

A flat Dirichlet process switching model for Bayesian estimation of hybrid systems

H. Wu, F. Noé[*]

Free University of Berlin, Arnimallee 6, 14195 Berlin, Germany

Abstract

Hybrid systems are often used to describe many complex dynamic phenomena by combining multiple modes of dynamics into whole systems. In this paper, we present a flat Dirichlet process switching (FDPS) model that defines a prior on mode switching dynamics of hybrid systems. Compared with the classical Markovian jump system (MJS) models, the FDPS model is nonparametric and can be applied to the hybrid systems with an unbounded number of potential modes. On the other hand, the probability structure of the new model is simpler and more flexible than the recently proposed hierarchical Dirichlet process (HDP) based MJS. Furthermore, we develop a Markov chain Monte Carlo (MCMC) method for estimating the states of hybrid systems with FDPS prior. And the numerical simulations of a hybrid system in different conditions are employed to show the effectiveness of the proposed approach.

Keywords: Dirichlet process, Markov chain Monte Carlo, hybrid system, Bayesian nonparametric estimation

Notation

Let $\mathbb{R}$, $\mathbb{R}^n$ and $\mathbb{R}^{m \times n}$ denote the sets of real numbers, real n-vectors and real $m \times n$ matrices, respectively. For $a, b \in \mathbb{R}$, $a \wedge b = \min\{a, b\}$. The cardinality of a set S is denoted by $|S|$. Given a sequence $\{x_t\}$, we denote the $\{x_k, x_{k+1}, \ldots, x_l\}$ by $x_{k:l}$, and $x_{k:l} = y$ means $x_k = x_{k+1} = \ldots = x_l = y$. For a finite sequence $x_{1:T}$, the notation x_{-t} stands for $(x_{1:t-1}, x_{t+1:T})$, $x_{-(t:s)}$ stands for $(x_{1:t-1}, x_{s+1:T})$, and the set of distinct values of $x_{1:T}$ is denoted by $\text{SD}(x_{1:T})$. $\mathcal{D}(\alpha_1, \ldots, \alpha_k)$ denotes the Dirichlet distribution of order k with parameters $\alpha_1, \ldots, \alpha_k$, $\mathcal{B}(\alpha, \beta)$ denotes the Beta distribution with parameters α, β, $\mathcal{G}(\alpha, \beta)$ denotes the Gamma distribution with parameters α, β, $\mathcal{U}(\alpha, \beta)$ denotes the uniform distribution over the interval $[\alpha, \beta]$, $\mathcal{N}(\mu, \Sigma)$ denotes the multivariate normal (MVN) distribution with mean μ and covariance matrix Σ, and $p_{\mathcal{N}}(\cdot | \mu, \Sigma)$ denotes the probability density function (pdf) of $\mathcal{N}(\mu, \Sigma)$. Given a probability measure G on a measurable space $(\mathcal{T}, \mathcal{A})$, the product measure G^2 on $(\mathcal{T} \times \mathcal{T}, \mathcal{A} \times \mathcal{A})$ is defined by

$$G^2(A \times B) = G(A)G(B), \quad \forall A, B \in \mathcal{A}.$$

---
[*]Corresponding author
Email addresses: hwu@zedat.fu-berlin.de (H. Wu), frank.noe@fu-berlin.de (F. Noé)

1. Introduction

Hybrid systems (also known as multiple model systems) [1] are a very important class of models, which are widely used in several fields of signal processing, including maneuvering target tracking [2], fault detection [3], model reduction of molecular dynamics [4] and so on. The major difference between the hybrid systems and standard dynamic systems with continuous state space is that a hybrid system combines hierarchically discrete/continuous state spaces, and each discrete state (or called mode) is associated with a continuous-state dynamical process. The state space model of the class of discrete-time hybrid systems under study here is given as below, and many hybrid systems (*e.g.*, the systems considered in [5, 6]) can be represented in this form:

$$\begin{cases} x_{t+1} &= f(\theta_{t+1}, x_t, w_{t+1}) \\ y_t &= g(\theta_t, x_t, v_t) \end{cases}, \quad t = 1, 2, \ldots, T \tag{1}$$

where f is the transition function, g is the measurement function, $x_t \in \mathbb{R}^{n_x}$ is the base (continuous) state variable, $y_t \in \mathbb{R}^{n_y}$ is the available measurement, $w_t \in \mathbb{R}^{n_w}$ and $v_t \in \mathbb{R}^{n_v}$ are process and measurement noise, and $\theta_t \in \mathbb{R}^{n_\theta}$ is the modal (discrete) state variable. Neither the base state process $x_{1:T}$ nor the modal state process $\theta_{1:T}$ is observed, and only the noisy measurement process $y_{1:T}$ is available. The major challenge of estimation for hybrid systems arises from the mode uncertainty, for it is impossible to investigate all the possible combinations of the modal states at different times.

In the research of hybrid systems, the most popular assumption concerning the mode is that the modal state process is a Markov chain (MC), and the systems satisfying this assumption are called Markovian jump systems (MJSs) [1, 7]. In the past, a variety of algorithms have been proposed for solving the state estimation problem of MJSs with completely known transition probabilities of modal states. such as generalized pseudo-Bayes (GPB) [8], interacting multiple model (IMM) [9] and expectation propagation (EP) [10] algorithms. And some researchers have proposed the Bayesian estimation algorithms for the unknown transition probability matrices (TPMs) within the framework of IMM [11, 7]. These algorithms approximate the posterior distributions of states by finite mixture models, and are able to get the estimates with low computation costs. But they may fail if the mixture models cannot approximate the true distributions accurately, especially when the system contains nonlinear/nonGaussian modes. To solve the problem of estimation of more general hybrid systems, some Monte Carlo estimation techniques including Markov chain Monte Carlo (MCMC) [12] and sequential Monte Carlo (SMC) [6, 13] have attracted much attention in this field. They provide a flexible framework to incorporate the heuristic approaches developed previously, and allow the Bayesian estimation without any model approximation or linearization.

In recent years, the Dirichlet Process (DP) [14] approach has been one of the most important approaches to nonparametric statistics. As a prior model of discrete distributions with infinite components, DP provides a powerful tool for Bayesian clustering and multiple model analysis. And the original DP has been used to estimate the Gaussian mixture noise density of dynamic systems [15]. But only a few studies have investigated the application of DP in general hybrid systems, and most of them are based on the hierarchical DPs (HDPs) proposed by Teh [16]. An HDP model consists of multiple DPs which are organized in a hierarchical structure, and can be used to develop an infinite discrete-state hidden Markov model (HMM). Fox [17, 18] modified the HDP-HMM and presented the HDP based MJS. The major advantage of the HDP-MJS model is that both the number and the values of modes appearing in the processes can be estimated in a purely Bayesian manner.

In this paper, we present an alternative and more flexible DP prior model for hybrid systems with unknown mode sets, and develop an MCMC algorithm for the Bayesian estimation of states. In comparison with the HDP-MJS, the new model contains only one DP, which greatly simplifies the probability model structure and makes the sampling easier. The remaining part of the paper is organized as follows. Section 2 reviews the basics of DPs and HDPs. In Section 3, we propose the new DP model for mode switching of hybrid systems. After describing the model, in Section 4 we present a Metropolis-within-Gibbs (MG) approach for state estimation, and in Section 5 provide results from an example problem. In Section 6, we briefly discuss the online estimation problem. Section 7 is a summary of the work in this paper.

2. Background

In order to make the paper self-contained, we provide a brief overview of DPs and HDPs in this section.

2.1. Dirichlet processes

A DP is a probability measure on the space of probability measures. Let G_0 be a probability measure on $(\mathcal{T}, \mathcal{A})$ and α be a positive real number. We say a random measure G is distributed according to a DP with scaling parameter α and base distribution G_0, denoted $G \sim \mathrm{DP}(\alpha, G_0)$, if for any finite measurable partition $\{A_1, \ldots, A_k\}$ of $\mathcal{T}$,

$$(G(A_1), \ldots, G(A_k)) \sim \mathcal{D}(\alpha G_0(A_1), \ldots, \alpha G_0(A_k)).$$

From the definition of DP, one can obtain two important equivalent descriptions of the DP model:

1. Stick-breaking representation. Sethuraman [19] proved that a $G \sim \mathrm{DP}(\alpha, G_0)$ can be written in the explicit form

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\eta_k}, \quad \pi_k = \pi'_k \prod_{l=1}^{k-1} \left(1 - \pi'_l\right) \tag{2}$$

with

$$\pi'_k | \alpha, G_0 \stackrel{\mathrm{iid}}{\sim} \mathcal{B}(1, \alpha), \quad \eta_k | \alpha, G_0 \stackrel{\mathrm{iid}}{\sim} G_0.$$

According to (2), a realization of DP is discrete with probability one.

2. Blackwell-MacQueen urn scheme. Suppose we draw a G from $\mathrm{DP}(\alpha, G_0)$, and independently draw random variables $\{\theta_t\}$ from G:

$$\begin{aligned} G | \alpha, G_0 &\sim \mathrm{DP}(\alpha, G_0), \\ \theta_t | G &\stackrel{\mathrm{iid}}{\sim} G. \end{aligned}$$

Integrating out G, the distributions of $\theta_1, \theta_2, \ldots$ can be provided by the Blackwell-MacQueen urn scheme [20]:

$$\theta_t | \theta_{1:t-1}, \alpha, G_0 \sim \frac{1}{t-1+\alpha} \sum_{l=1}^{t-1} \delta_{\theta_l} + \frac{\alpha}{t-1+\alpha} G_0. \tag{3}$$

(In this paper we denote the joint distribution of $\theta_{1:t}$ defined in (3) by $p_{\mathrm{MDP}}(\theta_{1:t} | \alpha, G_0)$.) Thus, $\{\theta_t\}$ forms a stochastic process. At each time t, θ_t takes a value from $\theta_{1:t-1}$ with a positive probability, and the probability is proportional to the number of times the value has occurred.

2.2. Hierarchical Dirichlet processes

The HDP is an extension DP model for solving problems involving groups of data. The generative model for an HDP is represented as

$$\begin{aligned} G_0 | \gamma, H &\sim \mathrm{DP}(\gamma, H), \\ G_i | \alpha_i, G_0 &\stackrel{\mathrm{ind}}{\sim} \mathrm{DP}(\alpha_i, G_0), \quad i = 1, 2, \ldots \end{aligned}$$

Under this hierarchical structure, the base distribution for each G_i is also a realization of DP. Therefore all the G_i share the common set of mixture components. The HDP can be applied to the Bayesian inference of HMMs, and the resulting model is called HDP-HMM [16, 21]. Suppose that $\{\theta_t\}$ is the state sequence of an HMM with countable state set $\Phi = \{\phi_1, \phi_2, \ldots\}$ and Markov kernel $K(\phi_i, \cdot) = \mathrm{Pr}(\theta_{t+1} \in \cdot | \theta_t = \phi_i)$. Then the HDP can be used to construct the prior models of $\{K(\phi_i, \cdot)\}$ under the case that both Φ and K are unknown, since all the $K(\phi_i, \cdot)$ have the same support within the HDP prior model constraints.

3. Flat Dirichlet process switching model for hybrid systems

It is clear that the dynamics of the hybrid system defined in (1) relies on the probability distribution of $\theta_{1:T}$, but which is generally unknown in practical applications. As mentioned above, the HDP-HMM can be used to construct a prior model for $p(\theta_{1:T})$, if the total number of modes appearing in the process is uncertain. However, the HDP-HMM involves multiple DPs which are associated with each other and form a hierarchical structure. It causes that the estimation procedure is very complex with a large number of auxiliary variables. In this section, we develop a more simple DP model for handling the mode switching dynamics.

3.1. Flat Dirichlet process switching model

Unlike the HDP-HMM, here we consider applying the DP model with base distribution G_0^2 to the sequence of transition pairs $\bar{\theta}_{1:T}$ instead of $\theta_{1:T}$, where G_0 is a continuous distribution on $\mathbb{R}^{n_\theta}$ and $\bar{\theta}_t = (\theta'_t, \theta_t)$ denotes the transition pair from time $t-1$ to time t. Certainly θ'_t should satisfy

$$\theta'_{t+1} = \theta_t. \tag{4}$$

However, the realization of $p_{\mathrm{MDP}}\left(\bar{\theta}_{1:T}|\alpha, G_0^2\right)$ does not satisfy (4) in the general case. Therefore we adopt the following modified prior model of $\bar{\theta}_{1:T}$:

$$\bar{\theta}_{1:T}|\alpha, G_0, \sigma_d^2 \sim p_{\mathrm{FDPS}}\left(\bar{\theta}_{1:T}|\alpha, G_0, \sigma_d^2\right) \propto \exp\left(-\frac{V\left(\bar{\theta}_{1:T}\right)}{\sigma_d^2}\right) p_{\mathrm{MDP}}\left(\bar{\theta}_{1:T}|\alpha, G_0^2\right) \tag{5}$$

where

$$V\left(\bar{\theta}_{1:T}\right) = \frac{1}{2}\sum_{t=1}^{T-1}\left(\theta_t - \theta'_{t+1}\right)^{\mathrm{T}} D^{-1}\left(\theta_t - \theta'_{t+1}\right).$$

is an energy function which we use to impose the soft constraints $\theta'_{t+1} \approx \theta_t$ on the values of $\bar{\theta}_{1:T}$, $D \in \mathbb{R}^{n_\theta \times n_\theta}$ is a positive definite matrix, and σ_d^2 is a small positive number. The model based on this prior with σ_d^2 small enough tends to make θ'_{t+1} and θ_t be approximately equal and could be therefore appropriate for describing the switching dynamics of $\theta_{1:T}$.

Remark 1. Compared to the HDP-HMM, the structure of the prior defined in (5) is "flat" and only a single DP is required. In this sense, we call the proposed model the flat DP switching (FDPS) model.

This prior model can also explained by the following virtual model $\mathcal{M}$:

$$
\begin{aligned}
G|\alpha, G_0 &\sim \mathrm{DP}\left(\alpha, G_0^2\right), \\
\bar{\theta}_t|G &\overset{\mathrm{iid}}{\sim} G, \\
u_t &\overset{\mathrm{iid}}{\sim} \mathcal{N}\left(0, \sigma_d^2 D\right), \\
y'_t &= \theta_t - \theta'_{t+1} + u_t,
\end{aligned}
$$

where u_t is virtual noise and y'_t is a virtual noisy measurement of the difference between θ_t and θ'_{t+1}. It is obvious that P_{FDPS} is equal to the marginal conditional distribution of $\bar{\theta}_{1:T}$ given all $y'_t = 0$ under the virtual model, *i.e.*,

$$p_{\mathrm{FDPS}}\left(\bar{\theta}_{1:T}|\alpha, G_0, \sigma_d^2\right) = p_{\mathcal{M}}\left(\bar{\theta}_{1:T}|\alpha, G_0, \sigma_d^2, y'_{1:T-1} = 0\right)$$

where $p_{\mathcal{M}}$ denotes the distribution under model $\mathcal{M}$.

Remark 2. If the discrete distribution G is given in the model $\mathcal{M}$, the conditional distribution of $\bar{\theta}_{1:T}$ can be written as

$$p_{\mathcal{M}}\left(\bar{\theta}_{1:T}|G, \sigma_d^2, y'_{1:T-1} = 0\right) \propto \exp\left(-\frac{V\left(\bar{\theta}_{1:T}\right)}{\sigma_d^2}\right)\prod_{t=1}^{T} G\left(\bar{\theta}_t\right).$$

The right-hand-side (rhs) of the equation consists of two terms. The first term stands for the soft constraints. And the second term is the product of weights of all the transition pairs defined by the G, which is equivalent to a Boltzmann chain model [22], which is an extension of MC. Therefore the proposed FDPS model can be treated as a prior for the approximate Boltzmann chain model, and the resulted hybrid system model has more flexibility than common MJS models.

Remark 3. For simplicity, we assume that $G_0 = \mathcal{N}(\mu_0, \Sigma_0)$ and $D = \Sigma_0$ in this paper. Then $\theta'_{1:T}$ can be integrated out in analysis and estimation.

3.2. Marginal distribution of FDPS model

We now consider the marginal distribution of the FDPS prior model:

$$p_{\text{MFDPS}}\left(\theta_{1:T}|\alpha, G_0, \sigma_d^2\right) = \int p_{\text{FDPS}}\left(\bar{\theta}_{1:T}|\alpha, G_0, \sigma_d^2\right) d\theta'_{1:T} = p_{\mathcal{M}}\left(\theta_{1:T}|\alpha, G_0, \sigma_d^2, y'_{1:T-1} = 0\right). \tag{6}$$

Suppose that a given $\theta_{1:T}$ has m distinct values $\{\phi_1, \dots, \phi_m\}$. Let c_t denote the corresponding indicators with $c_t = i \iff \theta_t = \phi_i$. It is obvious that the corresponding $\bar{\theta}_{1:T}$ also has m distinct values $\left\{\left(\phi'_1, \phi_1\right), \dots, \left(\phi'_m, \phi_m\right)\right\}$ and satisfies $\bar{\theta}_t = \left(\phi'_{c_t}, \phi_{c_t}\right)$ with probability one under model $\mathcal{M}$. From the above, we have

$$
\begin{aligned}
p_{\mathcal{M}}\left(\theta_{1:T}|\alpha, G_0, \sigma_d^2, y'_{1:T-1} = 0\right) \quad &\propto \quad p_{\mathcal{M}}\left(y'_{1:T-1} = 0, \theta_{1:T}|\alpha, G_0, \sigma_d^2\right) \\
&= \quad p_{\text{MDP}}\left(\theta_{1:T}|\alpha, G_0\right) p_{\mathcal{M}}\left(y'_{1:T-1} = 0|\phi_{1:m}, c_{1:T}, G_0, \sigma_d^2\right) \\
&= \quad p_{\text{MDP}}\left(\theta_{1:T}|\alpha, G_0\right) \\
&\qquad \cdot \int \left(\prod_{i=1}^{m} \left(p_{\mathcal{N}}\left(\phi'_i|\mu_0, \Sigma_0\right)\right)\right) \left(\prod_{t=1}^{T-1} p_{\mathcal{N}}\left(\phi_{c_t} - \phi'_{c_{t+1}}|0, \sigma_d^2 \Sigma_0\right)\right) d\phi'_{1:m}.
\end{aligned}
$$

Note that the function integrated in the second term on the rhs represents an MVN distribution of $\left(\phi_{1:m}, \phi'_{1:m}\right)$. Therefore we can get

$$p_{\text{MFDPS}}\left(\theta_{1:T}|\alpha, G_0, \sigma_d^2\right) = p_{\mathcal{M}}\left(\theta_{1:T}|\alpha, G_0, \sigma_d^2, y'_{1:T-1} = 0\right) = \frac{1}{Z\left(\alpha, \sigma_d^2, G_0\right)} p_{\text{MDP}}\left(\theta_{1:T}|\alpha, G_0\right) p_{\mathcal{N}}\left(\phi_{1:m}|\mu_r, \Sigma_r\right) \tag{7}$$

where $\mu_r \in \mathbb{R}^{mn_\theta}$ and $\Sigma_r \in \mathbb{R}^{mn_\theta \times mn_\theta}$ are both functions of $\left(c_{1:T}, \sigma_d^2, \mu_0, \Sigma_0\right)$ which can be easily computed by the Kalman filter (KF) in practice, and $Z\left(\alpha, \sigma_d^2, G_0\right)$ denotes the normalized constant.

4. MCMC estimation

4.1. State estimation

In this subsection, we consider the Bayesian inference of both base and modal states of the hybrid system conditioned on the measurements under the FDPS prior with hyperparameters of the FDPS model assumed to be known. (In the rest paper we will drop the hyperparameters of the FDPS model from the notation where there is no ambiguity.)

From the above results, the Bayesian inference relies on the posterior distribution

$$p\left(\theta_{1:T}, x_{1:T}|y_{1:T}\right) \propto \gamma\left(\theta_{1:T}, x_{1:T}\right) = p_{\text{MFDPS}}\left(\theta_{1:T}\right) \prod_{t=1}^{T} \psi_t\left(\theta_t, x_t, x_{t-1}\right) \tag{8}$$

where

$$\psi_t\left(\theta_t, x_t, x_{t-1}\right) = \begin{cases} p_{x_1}\left(x_1|\theta_1\right) p_g\left(y_1|x_1, \theta_1\right), & t = 1 \\ p_f\left(x_t|x_{t-1}, \theta_t\right) p_g\left(y_t|x_t, \theta_t\right), & t > 1 \end{cases},$$

p_f denotes the transition probability density with $x_t|x_{t-1}, \theta_t \sim p_f\left(x_t|x_{t-1}, \theta_t\right)$, p_g denotes the measurement probability density with $y_t|x_t, \theta_t \sim p_g\left(y_t|x_t, \theta_t\right)$, and p_{x1} denotes the distribution of the initial state x_1 with $x_1|\theta_1 \sim p_{x_1}\left(x_1|\theta_1\right)$. And the posterior can be approximated by MCMC methods based on MG sampling [23] method, which uses the Metropolis sampling technique to draw each variable (or group of variables) from its conditional distribution while holding all the other variables fixed. For the estimation problem in this section, each iteration of the MG sampler draws the following samples:

$\boldsymbol{\theta_{t:t+1}|\theta_{-(t:t+1)}, x_{1:T}, y_{1:T}}$ **for** $\boldsymbol{t = 1, \dots, T - 1}$**.** From (8), we have

$$p\left(\theta_{t:t+1}|\theta_{-(t:t+1)}, x_{1:T}, y_{1:T}\right) \propto p_{\text{MFDPS}}\left(\theta_{t:t+1}|\theta_{-(t:t+1)}\right) \psi_t\left(\theta_t, x_t, x_{t-1}\right) \psi_{t+1}\left(\theta_{t+1}, x_{t+1}, x_t\right)$$

where $p_{\text{MFDPS}}(\theta_{t:t+1}|\theta_{-(t:t+1)})$ is a finite mixture distribution of a discrete distribution and multiple MVN distributions, and can be easily derived from the expression for $p_{\text{MFDPS}}(\theta_{1:T})$ in (7). Then we generate a new sample for $\theta_{t:t+1}$ as

$$\theta_{t:t+1}^{\text{new}} \sim p_{\text{MFDPS}}(\theta_{t:t+1}|\theta_{-(t:t+1)})$$

and let $\theta_{t:t+1} \leftarrow \theta_{t:t+1}^{\text{new}}$ with probability

$$\frac{\psi_t\left(\theta_t^{\text{new}}, x_t, x_{t-1}\right) \psi_{t+1}\left(\theta_{t+1}^{\text{new}}, x_{t+1}, x_t\right)}{\psi_t\left(\theta_t, x_t, x_{t-1}\right) \psi_{t+1}\left(\theta_{t+1}, x_{t+1}, x_t\right)} \wedge 1.$$

Remark 4. Here we sample θ_t and θ_{t+1} instead of a single θ_t at each time. The reason is that the variance of $p_{\text{MFDPS}}(\theta_t|\theta_{-t})$ is often very small during the sampling procedure due to the soft constraint $\theta_t \approx \theta'_{t+1}$, and the sampler thereafter may be inefficient if we only sample a single θ_t at each time.

$x_{t:t+L-1}|\theta_{1:T}, x_{-(t:t+L-1)}, y_{1:T}$ **for** $t = 1, 1 + L, 1 + 2L, \ldots$ Here we adopt the similar strategy as in sampling θ_t, *i.e.*, divide $x_{1:T}$ into blocks with length L and sample a block at each step. The conditional distribution of $x_{t:t+L-1}$ is

$$p\left(x_{t:t+L-1}|\theta_{1:T}, x_{-(t:t+L-1)}, y_{1:T}\right) \propto \prod_{k=t}^{t+L} \psi_k\left(\theta_k, x_k, x_{k-1}\right).$$

And we can get an MVN distribution q_{x_t} which is an approximation of the conditional distribution by KF or unscented Kalman filter (UKF) [24, 25]. Then we draw

$$x_{t:t+L-1}^{\text{new}} \sim q_{x_{t:t+L-1}}(x_{t:t+L-1})$$

and let $x_{t:t+L-1} \leftarrow x_{t:t+L-1}^{\text{new}}$ with probability

$$\frac{\psi_t\left(\theta_t, x_t^{\text{new}}, x_{t-1}\right) \psi_{t+L}\left(\theta_{t+L}, x_{t+L}, x_{t+L-1}^{\text{new}}\right) \prod_{k=t+1}^{t+L-1} \psi_k\left(\theta_k, x_k^{\text{new}}, x_{k-1}^{\text{new}}\right) q_{x_{t:t+L-1}}(x_{t:t+L-1})}{\psi_t\left(\theta_t, x_t, x_{t-1}\right) \psi_{t+L}\left(\theta_{t+L}, x_{t+L}, x_{t+L-1}\right) \prod_{k=t+1}^{t+L-1} \psi_k\left(\theta_k, x_k, x_{k-1}\right) q_{x_{t:t+L-1}}\left(x_{t:t+L-1}^{\text{new}}\right)} \wedge 1.$$

$\phi_{1:m}|c_{1:T}, x_{1:T}, y_{1:T}$. Here the definitions of $\phi_{1:m}$ and $c_{1:T}$ are the same as in Subsection 3.2, and we will sample the values of $\theta_{1:T}$ with $c_{1:T}$ fixed in this step. According to (7) and (8), the conditional distribution of $\phi_{1:m}$ can be written as

$$p(\phi_{1:m}|c_{1:T}, x_{1:T}, y_{1:T}) \propto \left(\prod_{i=1}^{m} p_{\mathcal{N}}(\phi_i|\mu_0, \Sigma_0)\right) \cdot p_{\mathcal{N}}(\phi_{1:m}|\mu_r, \Sigma_r) \cdot \left(\prod_{t=1}^{T} \psi_t(\phi_{c_t}, x_t, x_{t-1})\right). \tag{9}$$

Then we draw

$$\phi_{1:m}^{\text{new}} \sim q_\phi(\phi_{1:m})$$

where q_ϕ is a proposal distribution, and let $\phi_{1:m} \leftarrow \phi_{1:m}^{\text{new}}$ with probability

$$\frac{\left(\prod_{i=1}^{m} p_{\mathcal{N}}(\phi_i^{\text{new}}|\mu_0, \Sigma_0)\right) \cdot p_{\mathcal{N}}(\phi_{1:m}^{\text{new}}|\mu_r, \Sigma_r) \cdot \left(\prod_{t=1}^{T} \psi_t(\phi_{c_t}^{\text{new}}, x_t, x_{t-1})\right) q_\phi(\phi_{1:m})}{\left(\prod_{i=1}^{m} p_{\mathcal{N}}(\phi_i|\mu_0, \Sigma_0)\right) \cdot p_{\mathcal{N}}(\phi_{1:m}|\mu_r, \Sigma_r) \cdot \left(\prod_{t=1}^{T} \psi_t(\phi_{c_t}, x_t, x_{t-1})\right) q_\phi(\phi_{1:m}^{\text{new}})} \wedge 1.$$

According to (9), we can also use KF or UKF to develop an MVN proposal q_ϕ such that $q_\phi(\phi_{1:m}) \approx p(\phi_{1:m}|c_{1:T}, x_{1:T}, y_{1:T})$.

After $M' + M$ iterations of the MG sampling (including M' burn-in iterations), the minimum mean-square error (MMSE) estimates of states can be computed as

$$\hat{x}_t = \mathbb{E}\left[x_t|y_{1:T}\right] \approx \frac{1}{M} \sum_{i=M'+1}^{M'+M} x_t^{(i)}, \quad \hat{\theta}_t = \mathbb{E}\left[\theta_t|y_{1:T}\right] \approx \frac{1}{M} \sum_{i=M'+1}^{M'+M} \theta_t^{(i)}$$

where $\left(\theta_{1:T}^{(i)}, x_{1:T}^{(i)}\right)$ denotes the sample of i-th iteration.

4.2. Estimation of α and σ_d^2

We now consider the case that the hyperparameters α and σ_d^2 are unknown. It seems reasonable to jointly estimate hyperparameters and states by MG sampling based on

$$p\left(\alpha, \sigma_d^2 | \theta_{1:T}, x_{1:T}, y_{1:T}\right) = p\left(\alpha, \sigma_d^2 | \theta_{1:T}\right) \propto \frac{p\left(\alpha, \sigma_d^2\right)}{Z\left(\alpha, \sigma_d^2, G_0\right)} p_{\text{MDP}}\left(\theta_{1:T} | \alpha, G_0\right) p_{\mathcal{N}}\left(\phi_{1:m} | \mu_r, \Sigma_r\right)$$

where $p\left(\alpha, \sigma_d^2\right)$ denotes the prior distribution of α and σ_d^2. But it is impossible to sample directly for $Z\left(\alpha, \sigma_d^2, G_0\right)$ is intractable. So here we approximate the $p\left(\alpha, \sigma_d^2\right)$ by a discrete distribution

$$\hat{p}\left(\alpha, \sigma_d^2\right) = \sum_{i=1}^{N_h} \frac{1}{N_h} \delta_{\alpha_i, \sigma_{d,i}^2}\left(\alpha, \sigma_d^2\right)$$

and draw

$$\left(\alpha, \sigma_d^2\right) \sim \hat{p}\left(\alpha, \sigma_d^2 | \theta_{1:T}\right) \propto \sum_{i=1}^{N_h} \frac{1}{Z_i} p_{\text{MDP}}\left(\theta_{1:T} | \alpha_i, G_0\right) p_{\mathcal{N}}\left(\phi_{1:m} | \mu_r, \Sigma_r\right) \delta_{\alpha_i, \sigma_{d,i}^2}\left(\alpha, \sigma_d^2\right)$$

where $\left\{\left(\alpha_i, \sigma_{d,i}^2\right)\right\}$ are N_h samples from $p\left(\alpha, \sigma_d^2\right)$ and $Z_i = Z\left(\alpha_i, \sigma_{d,i}^2, G_0\right)$. Although the values of Z_i are still unknown, the modified Wang-Landau algorithm [26] can be utilized to get the approximates $\hat{Z}_i \approx Z_i / \left(\sum_j Z_j\right)$ through sampling. (Due to space limitations, we omit details of the algorithm.)

Remark 5. Generally speaking, N_h should be large enough such that the discrete distribution can well approximate the true distribution. If we apply the proposed method to estimate (μ_0, Σ_0), which contains $O\left(n_\theta^2\right)$ random numbers, the required value of N_h may be so large that the estimation approach implemented here is impractical. So how to achieve the Bayesian inference of all the hyperparameters of FDPS model remains an open problem. And μ_0, Σ_0 are set by trials and errors in this paper.

5. Simulations

In this section, the approach proposed in this paper will be applied to a hybrid system described by

$$\begin{cases} x_{t+1} &= a_{t+1} x_t + b_{t+1} + 25 \frac{x_t}{1+x_t^2} + w_{t+1} \\ y_t &= \frac{x_t^2}{20} + v_t \end{cases}, \quad t = 1, 2, \ldots, T$$

with $T = 300$, $x \sim \mathcal{N}(0, 1)$, $w_t \sim \mathcal{N}(8 \cos(1.2(t-1)), 0.1)$, $v_t \sim \mathcal{N}(0, 1)$. And $\theta_t = (a_t, b_t)$ switches between three modes $\phi_1 = (1, 0)$, $\phi_2 = (1, 3)$ and $\phi_3 = (0.5, 0)$. In simulations, $\theta_{1:T}$ are generated according to the following two models, denoted M1 and M2.

- M1: $\theta_{1:T}$ is an MC with

$$\Pr\left(\theta_{t+1} = \phi_j | \theta_t = \phi_i\right) = \begin{cases} 0.8, & i = j \\ 0.1, & i \neq j \end{cases} \quad \text{and} \quad \Pr\left(\theta_1 = \phi_i\right) = \frac{1}{3}.$$

- M2: The duration of the i-th mode follows the Poisson distribution with parameter ϵ_i, $\Pr(\theta_1 = \phi_i) = \frac{1}{3}$ and $\Pr\left(\theta_{t+1} = \phi_j | \theta_t = \phi_i, \theta_{t+1} \neq \theta_t\right) = w_{ij}$, where

$$(\epsilon_1, \epsilon_2, \epsilon_3) = (10, 20, 35) \quad \text{and} \quad w_{12} = w_{23} = w_{31} = 0.2, \quad w_{13} = w_{21} = w_{32} = 0.8.$$

This is a nonMarkovian model (see [27] for details).

Table 1: State estimation results based on the FDPS and classical MJS models. This plot shows the mean and variance of the MSEs calculated over 50 independent runs.

	M1				M2			
	MSE_x		MSE_θ		MSE_x		MSE_θ	
	mean	variance	mean	variance	mean	variance	mean	variance
FDPS	0.3913	0.0198	0.2483	0.0054	0.4588	0.0182	0.0498	0.0016
Classical MJS	0.2614	0.0079	0.2016	0.0077	0.5687	0.0268	0.2492	0.0084

The parameters of our algorithm/model are chosen as $M' = 1000$, $M = 3000$, $L = 10$, $\mu_0 = (0, 0)$, $\Sigma_0 = \mathrm{diag}\,(1, 5)$, and $\left\{ \left(\alpha_i, \sigma_{d,i}^2 \right) \right\}$ with $N_h = 200$ are sampled according to $\alpha \sim \mathcal{G}(2, 2)$ and $\log \sigma \sim \mathcal{U} \left(\log 10^{-3}, \log 0.5 \right)$. For comparison we also perform the Bayesian estimation based on the classical MJS model with 3 modes by the MG sampling method [28], where the iteration number is the same. (For this case, the classical MJS model is "more exact" than the HDP-HMM based model since the number of modes is set correctly.)

We repeated the simulation and estimation procedures 50 times. Table 1 summarizes the performance of the FDPS and classical MJS models. The table shows the means and variances of the mean square errors (MSEs) of the state estimates, where

$$\mathrm{MSE}_x = \frac{1}{T} \sum_{t=1}^{T} \|\hat{x}_t - x_t\|^2, \quad \mathrm{MSE}_\theta = \frac{1}{T} \sum_{t=1}^{T} \|\hat{\theta}_t - \theta_t\|^2.$$

Fig. 1 shows the estimation results of modal states obtained from a single run of M1 and M2. Note that the MJS model with 3 modes is an "exact" prior model of M1. By contrast, the proposed FDPS model achieves the similar estimation performance for M1 with the number of modes unknown. For the nonMarkovian jump system model M2, the FDPS model outperforms the MJS model, especially in the mode estimation. This demonstrates the robustness and flexibility of the FPF, (see Remark 2). Moreover, in the simulations in Fig. 1, there are 9 different transition pairs in $\{\theta_{1:T}\}$ ($\phi_1 \to \phi_1, \phi_1 \to \phi_2, \ldots, \phi_3 \to \phi_3$), and the FDPS model can get the correct number in most samples.

6. SMC for online estimation

In some applications, the states of the systems are required to be estimated online. For these cases, we can apply the SMC method to sampling sequentially from distributions $\{p(\theta_{1:t}, x_{1:t}|y_{1:t}) \propto \gamma(\theta_{1:t}, x_{1:t}) | t = 1, 2, \ldots\}$.

For the general JMS model, the SMC method performs online sampling based on the following recursive equation:

$$p(\theta_{1:t}, x_{1:t}|y_{1:t}) \propto \psi(\theta_t, x_t, x_{t-1})\, p(\theta_t|\theta_{1:t-1})\, p(\theta_{1:t-1}, x_{1:t-1}|y_{1:t-1}).$$

However, in the proposed FDPS based hybrid system model, $p_{\mathrm{MFDPS}}(\theta_{1:t})$ cannot be expressed as $p_{\mathrm{MFDPS}}(\theta_{1:t}) = p_{\mathrm{MFDPS}}(\theta_{1:t-1})\, p_{\mathrm{MFDPS}}(\theta_t|\theta_{1:t-1})$ for $\int p_{\mathrm{MFDPS}}(\theta_{1:t})\, d\theta_t \neq p_{\mathrm{MFDPS}}(\theta_{1:t-1})$. Therefore (8) is not a generative but a discriminative probabilistic model, and the estimates cannot be calculated by the ordinary SMC method for JMS models directly. Fortunately, the posterior distributions of states can be converted into a generative form by applying the virtual model $\mathcal{M}$. From (6), we have

$$
\begin{aligned}
\gamma(\theta_{1:t}, x_{1:t}) &\propto p_{\mathcal{M}}\left(\theta_{1:t}, x_{1:t}|y_{1:t}, y'_{1:t-1} = 0\right) \\
&\propto \psi(\theta_t, x_t, x_{t-1})\, p_{\mathcal{M}}\left(\theta_t, y'_{t-1} = 0|\theta_{1:t-1}\right) p_{\mathcal{M}}\left(\theta_{1:t-1}, x_{1:t-1}|y_{1:t-1}, y'_{1:t-2} = 0\right)
\end{aligned}
$$

and the $p_{\mathcal{M}}\left(\theta_t, y'_{t-1} = 0|\theta_{1:t-1}\right)$ can be recursively computed by KF. Then we can develop the SMC method for FDPS based hybrid system models (details will be given elsewhere).
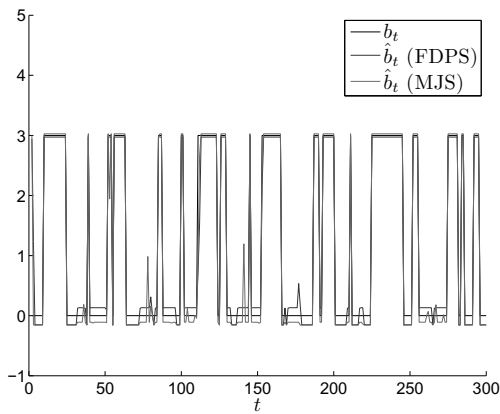
7. Conclusions

In this paper, we have developed a Bayesian nonparametric model and estimator for hybrid systems with unknown mode sets. The main difference between our approach and other nonparametric approaches is that we utilize the DP
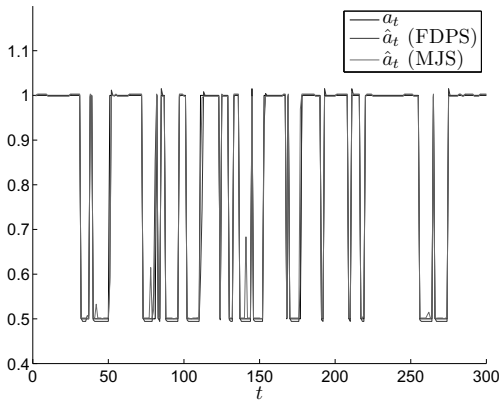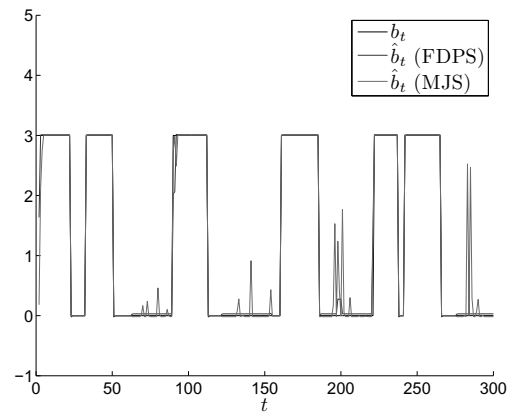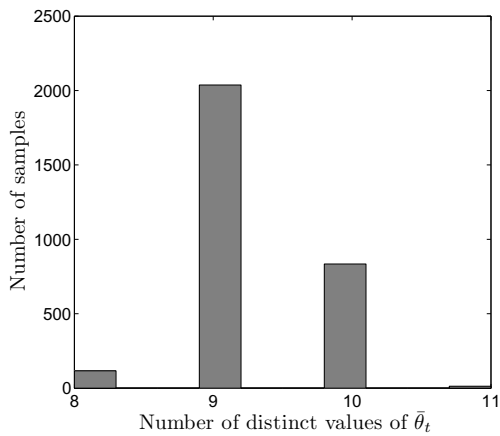
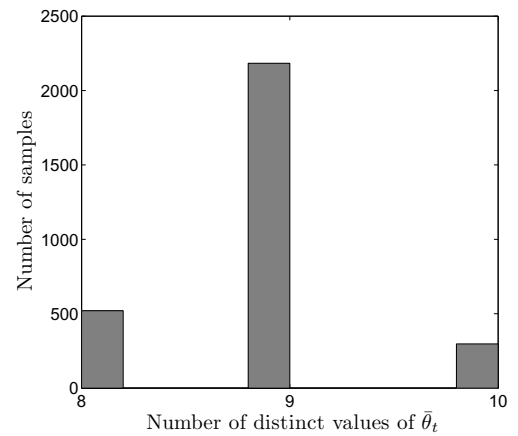(a) Estimated a_t (M1)
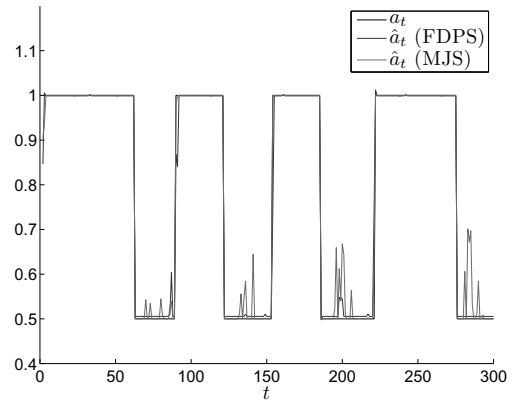
(b) Estimated a_t (M2)

(c) Estimated b_t (M1)

(d) Estimated b_t (M2)

(e) Histogram of $\left|\mathrm{SD}\left(\bar{\theta}_{1:T}^{(i)}\right)\right|$ for $i = M' + 1, \ldots, M' + M$ (M1)

(f) Histogram of $\left|\mathrm{SD}\left(\bar{\theta}_{1:T}^{(i)}\right)\right|$ for $i = M' + 1, \ldots, M' + M$ (M2)

Figure 1: Mode estimation results for data generated by M1 and M2. (a) and (b) Estimates of $a_{1:T}$ using the FDPS model and classical MJS model. (c) and (d) Estimates of $b_{1:T}$. (e) and (f) Histograms of the numbers of distinct transition pairs $\bar{\theta}_t$ appearing in the samples of FDPS model. (Although the complete transition pairs $\bar{\theta}_t$ are not sampled in our method, we have $\left|\mathrm{SD}\left(\bar{\theta}_{1:T}^{(i)}\right)\right| = \left|\mathrm{SD}\left(\theta_{1:T}^{(i)}\right)\right|$.)

to model the distribution of mode transition pairs instead of individual modes. Consequently the proposed FDPS prior does not need multiple DPs as the HDP prior, and the corresponding probability model of hybrid systems is greatly simplified and more flexible. Using the MCMC method, we may efficiently compute state estimates from noisy measurements. Future work will concentrate on schemes for Bayesian inference of the whole hyperparameters of FDPS model. It would also be important to improve the estimation performance by some new sampling techniques.

Acknowledgments

References

[1] D. Sworder, J. Boyd, Estimation problems in hybrid systems, Cambridge Univ. Press.

[2] R. Li, V. Jilkov, Survey of maneuvering target tracking. Part V. Multiple-model methods, IEEE Transactions on Aerospace and Electronic Systems 41 (4) (2005) 1255–1321.

[3] M. Zhong, H. Ye, P. Shi, G. Wang, Fault detection for Markovian jump systems, IEE Proceedings-Control Theory and Applications 152 (4) (2005) 397–402.

[4] I. Horenko, E. Dittmer, C. Schütte, Reduced stochastic models for complex molecular systems, Computing and Visualization in Science 9 (2) (2006) 89–102.

[5] I. Hwang, H. Balakrishnan, C. Tomlin, Observability criteria and estimator design for stochastic linear hybrid systems, in: Proceedings of the IEE European Control Conference, 2003, pp. 3–2.

[6] H. Driessen, Y. Boers, Efficient particle filter for jump Markov nonlinear systems, IEE Proceedings-Radar, Sonar and Navigation 152 (5) (2005) 323–326.

[7] V. Jilkov, X. Li, Online Bayesian estimation of transition probabilities for Markovian jump systems, IEEE Transactions on Signal Processing 52 (6) (2004) 1620–1630.

[8] G. Ackerson, K. Fu, On state estimation in switching environments, IEEE Transactions on Automatic Control 15 (1) (2002) 10–17.

[9] H. Blom, Y. Bar-Shalom, The interacting multiple model algorithm for systems with Markovian switching coefficients, IEEE Transactions on Automatic Control 33 (8) (2002) 780–783.

[10] T. Minka, Expectation propagation for approximate Bayesian inference, in: Proceedings of Uncertainty in Artificial Intelligence, Vol. 17, 2001, pp. 362–369.

[11] A. Doucet, B. Ristic, Recursive state estimation for multiple switching models with unknown transition probabilities, IEEE Transactions on Aerospace and Electronic Systems 38 (3) (2002) 1098–1104.

[12] V. Jilkov, X. Li, D. Angelova, Estimation of markovian jump systems with unknown transition probabilities through bayesian sampling, Numerical Methods and Applications (2003) 307–315.

[13] S. Tafazoli, X. Sun, Hybrid system state tracking and fault detection using particle filters, IEEE Transactions on Control Systems Technology 14 (6) (2006) 1078–1087.

[14] T. Ferguson, A Bayesian analysis of some nonparametric problems, The annals of statistics 1 (2) (1973) 209–230.

[15] F. Caron, M. Davy, A. Doucet, E. Duflos, P. Vanheeghe, Bayesian inference for linear dynamic models with Dirichlet process mixtures, IEEE Transactions on Signal Processing 56 (1) (2007) 71–84.

[16] Y. Teh, M. Jordan, M. Beal, D. Blei, Hierarchical dirichlet processes, Journal of the American Statistical Association 101 (476) (2006) 1566–1581.

[17] E. Fox, E. Sudderth, A. Willsky, Hierarchical Dirichlet processes for tracking maneuvering targets, in: Proceedings of the 10th International Conference on Information Fusion, IEEE, 2007, pp. 1–8.

[18] E. Fox, E. Sudderth, M. Jordan, A. Willsky, Bayesian Nonparametric Methods for Learning Markov Switching Processes, IEEE Signal Processing Magazine 27 (6) (2010) 43–54.

[19] J. Sethuraman, A constructive definition of Dirichlet priors, Statistica Sinica 4 (2) (1994) 639–650.

[20] D. Blackwell, J. MacQueen, Ferguson distributions via Pólya urn schemes, The annals of statistics 1 (2) (1973) 353–355.

[21] M. Beal, Z. Ghahramani, C. Rasmussen, The infinite hidden Markov model, Advances in Neural Information Processing Systems 1 (2002) 577–584.

[22] L. Saul, M. Jordan, Boltzmann chains and hidden Markov models, Advances in neural information processing systems (1995) 435–442.

[23] A. Gelfand, Gibbs Sampling, Journal of the American Statistical Association 95 (452) (2000) 1300–1304.

[24] A. Doucet, S. Godsill, C. Andrieu, On sequential Monte Carlo sampling methods for Bayesian filtering, Statistics and computing 10 (3) (2000) 197–208.

[25] E. Wan, R. Van Der Merwe, The unscented Kalman filter for nonlinear estimation, in: Proceedings of Adaptive Systems for Signal Processing, Communications, and Control Symposium, IEEE, 2002, pp. 153–158.

[26] Y. Atchadé, N. Lartillot, C. Robert, Bayesian computation for statistical models with intractable normalizing constants, Research report, arXiv: 0804.3152 v1.

[27] P. Djuric, J. Chun, An MCMC sampling approach to estimation of nonstationary hidden Markov models, IEEE Transactions on Signal Processing 50 (5) (2002) 1113–1123.

[28] S. Scott, Bayesian Methods for Hidden Markov Models: Recursive Computing in the 21st Century., Journal of the American Statistical Association 97 (457) (2002) 337–352.