DIRICHLET-PROCESS-MIXTURE-BASED BAYESIAN NONPARAMETRIC METHOD FOR MARKOV SWITCHING PROCESS ESTIMATION

Clément Magnant [1],[3] *Audrey Giremus* [3] *Eric Grivel* [3] *Laurent Ratton* [2] *Bernard Joseph* [1]

Thales Systèmes Aéroportés S.A., [1] Pessac, [2] Elancourt, FRANCE
[3]University of Bordeaux - Bordeaux INP - IMS - UMR CNRS 5218, Talence, FRANCE

ABSTRACT

Dirichlet process (DP) mixtures were recently introduced to deal with switching linear dynamical models (SLDM). They assume the system can switch between an *a priori* infinite number of state-space representations (SSR) whose parameters are on-line inferred. The estimation problem can thus be of high dimension when the SSR matrices are unknown. Nevertheless, in many applications, the SSRs can be categorized in different classes. In each class, the SSRs are characterized by a known functional form but differ by a reduced set of unknown hyperparameters. To use this information, we thus propose a new hierarchical model for the SLDM wherein a discrete variable indicates the SSR class. Conditionally to this class, the distributions of the hyperparameters are modeled by DPs. The estimation problem is solved by using a Rao-Blackwellized particle filter. Simulation results show that our model outperforms existing methods in the field of target tracking.

Index Terms— Bayesian non-parametric methods, Dirichlet process mixtures, particle filter, Rao-Blackwellization, interactive multiple models, target tracking.

1. INTRODUCTION

Recursive estimation plays a key role in various fields such as financial time series, weather forecast, hydrology, radar processing, or GPS navigation. In a Bayesian setting, this estimation is based on the so-called state-space representation (SSR) of the system. The latter significantly impacts the estimation algorithm performance. In many applications such as target tracking, a single SSR may not be well-suited if the system exhibits dynamical changes over time. Classically, in this context, switching linear dynamical (SLD) models can be considered and the state vector can be estimated by using a multiple-model (MM) algorithm [1] such as the interacting MM (IMM) [2] and the variable-structure IMM [3]. However, these approaches rely on the strong assumption that the system can only switch between a finite number of *a priori* known models. More recently, Bayesian non-parametric (BNP) methods, which are popular in statistics or machine learning, have been suggested as an alternative [4, 5, 6]. Thanks to these approaches, which are more and more used in various applications such as [7], any assumption regarding the model distributions can be relaxed.

On the one hand, in [4], the state-transition matrix is *a priori* defined but the distributions of the state model and the measurement noises are assumed to be unknown. Therefore, they are modeled by Dirichlet process mixtures (DPM) which can be seen as infinite mixtures of probability density functions (pdf), for instance Gaussian ones. This amounts to considering that the noise covariance matrices can switch among an unknown number of values, some of which may reappear more or less frequently.

On the other hand, in [6], Fox *et al.* address the case of both the unknown state-transition and observation matrices. For this purpose, they make use of an extension of the sticky hierarchical Dirichlet-process hidden-Markov model (HDP-HMM) to learn an unknown number of persistent dynamical modes. This approach is flexible and can apply to a wide range of applications. Its drawback lies in the dimension of the model parameters to be learnt which can be very high depending on the the state-vector size. To decrease the dimensionality of the problem, contrary to [4] and [6], we propose to take advantage that the candidate SSRs can be categorized in a reduced number of classes for a given application. The SSRs within each class all share the same functional form and are characterized by only a reduced number of time-switching hyperparameters to be estimated. For instance, in object tracking, the motion can be described by one of the following well-identified model classes: constant velocity (CV), constant acceleration (CA), constant turn (CT), etc. Each one depends on few unknown time-switching parameters such as the acceleration variance.

Therefore, in this paper, we present a new hierarchical model that can be seen as an extension of the one proposed in [4]. We assume that the system dynamics can switch among a finite number of SSR classes. The proposed model thus includes a discrete variable that indicates the SSR class. Conditionally to this class, the distributions of the hyperparameters are modeled by Dirichlet processes (DPs). The system evolution is hence described by a mixture of DPs. On the basis of our hierarchical model, we perform joint Bayesian inference of the current model class, the state vector and the state-transition pdf at each

This work is part of the common research activities between Bordeaux campus and Thales (GIS Albatros).

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

instant. Since the model is conditionally linear Gaussian, a Rao-Blackwellized particle filter (RBPF) is used [8]. In the simulation results, the proposed model and the resulting estimation algorithm are applied to target tracking.
Our paper is organized as follows: in section 2, the proposed hierarchical model is motivated. In section 3, the RBPF that solves the estimation problem is presented. Then, in section 4, our approach is applied in the framework of target tracking and compared with an IMM and an approach derived from [4]. Finally, some conclusions and perspectives are drawn in section 5.

2. BAYESIAN MODELING OF THE PROBLEM

2.1. Problem statement

Linear dynamical systems are described by a SSR as follows:

$$\begin{aligned} \boldsymbol{x}_t &= \Phi_t \boldsymbol{x}_{t-1} + u_{t-1} \\ \boldsymbol{y}_t &= H_t \boldsymbol{x}_t + b_t \end{aligned} \tag{1}$$

where $\boldsymbol{x}_t$ is the state vector at time t, Φ_t the transition matrix, $\boldsymbol{y}_t$ the observation vector and H_t the observation matrix. In addition, b_t is the observation noise which is uncorrelated with the model noise u_{t-1}.
Equivalently, the SSR (1) can be rewritten in terms of conditional pdf, namely the state-transition pdf $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ and the likelihood $p(\boldsymbol{y}_t|\boldsymbol{x}_t)$. They are used by filtering techniques to sequentially calculate the posterior state pdf $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})$ from the set of observations from time 1 to time t. Then, the state estimate can be obtained as the mean or the mode of this posterior distribution.
In many applications, a unique SSR is not sufficient to plainly represent the system evolution. Since the seminal work of Bar-Shalom [2], SLD systems have been considered to address this issue. The system is then assumed to switch between various SSRs over time. However, these approaches require to *a priori* define a finite set of relevant models.
Very recently, BNP methods have been proposed to alleviate this difficulty. In this case, both the state-transition pdf and the likelihood function are assumed to be unknown and are represented by infinite mixtures of pdfs. In various applications, information about the measurement-noise distribution can be obtained. The main difficulty thus stands in conveniently designing the state-vector dynamical evolution. Therefore, in the remaining of the paper, we focus on the state-transition pdf. The latter can be represented by the following model:

$$p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) = \int_{\boldsymbol{\theta}_t} p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{\theta}_t) dG(\boldsymbol{\theta}_t) \tag{2}$$

where G is the so-called mixing distribution, $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{\theta}_t)$ is Gaussian and $\boldsymbol{\theta}_t$ is a vector of latent variables including for instance the components of both the state transition matrix and the model-noise covariance matrix.
The originality of these BNP methods is that G is unknown and assumed to be random. In a Bayesian setting, it is assigned a prior classically chosen as a DP such as $G \sim \mathcal{DP}(G_0, \alpha)$ with G_0 its base distribution and α the scale parameter. DPs are therefore defined as probability measures on the space of probability measures.
In the remainder of this section, the theoretical background on DPs for density estimation in the general case is first provided. Then, a new DP-based hierarchical model to efficiently deal with SLD systems is presented.

2.2. DP properties

The realizations G of a DP are infinite distributions. By using the stick-breaking representation, they can be expressed as:

$$G(\boldsymbol{\theta}_t) = \sum_{j=1}^{+\infty} \pi_j \delta_{\boldsymbol{U}_j}(\boldsymbol{\theta}_t) \tag{3}$$

where $\delta_{\boldsymbol{U}_j}(\boldsymbol{\theta}_t)$ is the Dirac distribution centered in $\boldsymbol{U}_j$, $\boldsymbol{U}_j \sim G_0$, $\pi_j = \beta_j \prod_{l=1}^{j-1}(1-\beta_l)$ and $\beta_j \sim \mathcal{B}(1,\ \alpha)$, where $\mathcal{B}$ stands for the Beta law. By inserting (3) into (2), it turns out:

$$p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) = \sum_{j=1}^{+\infty} \pi_j p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{U}_j) \tag{4}$$

In the above equation, $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ corresponds to an infinite mixture of the pdf $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{U}_j)$ with mixture weights π_j and latent variables contained in $\boldsymbol{U}_j$.
Note that estimating G is an infinite-dimensional problem. For this reason, these approaches are known as non-parametric methods. However, Blackwell *et al.* showed in [9] that by using DP, the inference procedure boils down to the estimation of the latent vector $\boldsymbol{\theta}_t$. Indeed, the predicted distribution of $\boldsymbol{\theta}_t$ given the latent variables $\boldsymbol{\theta}_{1:t-1}$ can be directly computed by marginalizing G. It leads to the Polya urn representation:

$$p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{1:t-1}, \alpha) = \frac{1}{\alpha+t-1}\sum_{j=1}^{t-1} \delta_{\boldsymbol{\theta}_j}(\boldsymbol{\theta}_t) + \frac{\alpha}{\alpha+t-1} G_0(\boldsymbol{\theta}_t) \tag{5}$$

It can be interpreted as follows: given the previous latent variables $\boldsymbol{\theta}_{1:t-1}$, a new sample can either be drawn from the distribution G_0 with probability $\frac{\alpha}{\alpha+t-1}$ or take the same value as a previous sample with probability $\frac{t-1}{\alpha+t-1}$. Therefore, the scale parameter plays a key role. If α is small, the same value of $\boldsymbol{\theta}_t$ is drawn several times whereas if α tends to infinity, the samples become *iid* from G_0.
As for the estimation problem, it becomes all the more difficult as the dimension of $\boldsymbol{\theta}_t$ is high. For instance, $\boldsymbol{\theta}_t$ can be composed of all the elements of the state matrices. However, in many applications, the state transition and model-noise covariance matrices can only take specific functional forms. A state model can thus be characterized by a reduced number of hyper-parameters.
In the next subsection, we suggest a new hierarchical model to take advantage of this information. It consists of a mixture of DPs.

2.3. Proposed hierarchical model

In the following, we extend the representation (2) by introducing a latent variable that refers to a class of state models. All the state models in a class share the same functional form but differ with one another by a reduced number of hyperparameters.
Let $z_t \in \{1, ..., M\}$ denote the index of the actual state-model class at time t. Here, the sequence $\{z_t\}_{t>0}$ is assumed to be a Markov chain with transition probability matrix (TPM) denoted as $\Pi = \{\pi_{ij}\}_{i=1,...,M}^{j=1,...,M}$. Note that unlike standard MM-based approaches, z_t does not refer to a well-defined model but to a class that comprises an infinity of models.
Conditionally to z_t, the state-transition pdf is therefore described by a mixture of DPs as suggested in (2). The specificity of our work is hence that there are as many DPs as possible model classes. They are denoted as $\{G^m\}_{m=1,...,M}$. Each one is characterized by its own base distribution G_0^m and its scale parameter α^m.
In addition, for each model class, the functional forms of the state matrices are known. Hence, $\boldsymbol{\theta}_t$ is no longer of high dimension but is only composed of few parameters that fully characterize the matrices to be estimated.
At this stage, given the above considerations, the relationships between z_t, G^m, $\boldsymbol{\theta}_t$, $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$ can be described by the following hierarchical model:

$$G^m \sim \mathcal{DP}(G_0^m, \alpha^m) \text{ for } m = 1, ..., M \tag{6}$$

$$z_t | z_{t-1} \sim \pi_{z_t z_{t-1}} \tag{7}$$

$$\boldsymbol{\theta}_t | \{z_t, \{G^m\}_{m=1,...,M}\} \sim G^{z_t}(\boldsymbol{\theta}_t) \tag{8}$$

$$\boldsymbol{x}_t | \{\boldsymbol{x}_{t-1}, \boldsymbol{\theta}_t, z_t\} \sim p(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}, \boldsymbol{\theta}_t, z_t) \tag{9}$$

$$\boldsymbol{y}_t | \boldsymbol{x}_t \sim p(\boldsymbol{y}_t | \boldsymbol{x}_t) \tag{10}$$

In this work, linear Gaussian SSRs of the form (1) are considered. Therefore, the pdfs $p(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}, \boldsymbol{\theta}_t, z_t)$ and $p(\boldsymbol{y}_t | \boldsymbol{x}_t)$ are Gaussian.
Similarly to subsection 2.2., the unknown distributions $\{G^m\}_{m=1,...,M}$ can be integrated out based on the Polya urn representation. However, the latter must be modified to take into account the switching between the different model classes. The predictive distribution of $\boldsymbol{\theta}_t$ becomes:

$$\boldsymbol{\theta}_t | \boldsymbol{\theta}_{1:t-1}, z_t \sim \frac{1}{\alpha^{z_t} + n_{z_t}} \sum_{\substack{j=1 \\ s.t.\ z_j = z_t}}^{t-1} \delta_{\boldsymbol{\theta}_j}(\boldsymbol{\theta}_t) + \frac{\alpha^{z_t}}{\alpha^{z_t} + n_{z_t}} G_0^{z_t}(\boldsymbol{\theta}_t) \tag{11}$$

where n_{z_t} is the number of times the model class z_t has previously appeared.
The hierarchical model defined by (6)-(10) thus reduces to (7), (11), (9), and (10) as depicted by Fig. 1. The objective is then to on-line estimate the joint posterior distribution of all the unknown parameters $p(\boldsymbol{x}_{0:t}, z_{0:t}, \boldsymbol{\theta}_{0:t} | \boldsymbol{y}_{1:t})$. The latter does not admit a closed-form expression due to the non-linearity and non-Gaussianity of the proposed model.
In the next section, we show how to address this issue by using a Rao-Blackwellized particle filter.

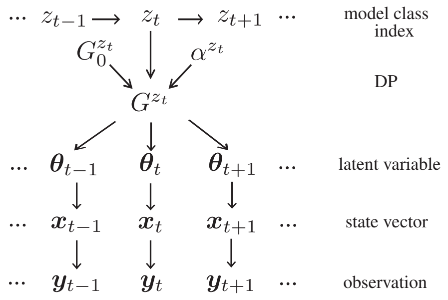

Fig. 1. Graphical representation of the hierarchical model

3. PARTICLE FILTERING ESTIMATION

The posterior distribution can be factorized as follows:

$$\begin{aligned} &p(\boldsymbol{x}_{0:t}, z_{0:t}, \boldsymbol{\theta}_{0:t} | \boldsymbol{y}_{1:t}) \\ &\quad = p(\boldsymbol{x}_{0:t} | z_{0:t}, \boldsymbol{\theta}_{0:t}, \boldsymbol{y}_{1:t}) p(\boldsymbol{\theta}_{0:t}, z_{0:t} | \boldsymbol{y}_{1:t}) \end{aligned} \tag{12}$$

Then, it should be noted that if the sequences of model classes $z_{0:t}$ and model parameters $\boldsymbol{\theta}_{0:t}$ were known, the SSR of the considered system would become linear and Gaussian. Therefore, the distribution $p(\boldsymbol{x}_{0:t} | z_{0:t}, \boldsymbol{\theta}_{0:t}, \boldsymbol{y}_{1:t})$ would be Gaussian and its mean vector and covariance matrix could be calculated by optimal Kalman filtering. To take advantage of this structure, a RBPF can be used to perform the estimation. The principle is to analytically solve a part of the estimation problem so that only the distribution $p(\boldsymbol{\theta}_{0:t}, z_{0:t} | \boldsymbol{y}_{1:t})$ is approximated using sequential Monte Carlo sampling. More precisely, this posterior distribution is approximated by a discrete distribution:

$$\widehat{p}(\boldsymbol{\theta}_{0:t}, z_{0:t} | \boldsymbol{y}_{1:t}) = \sum_{i=1}^{N} w_t^i \delta_{\{\boldsymbol{\theta}_{0:t}^i, z_{0:t}^i\}}(\{\boldsymbol{\theta}_{0:t}, z_{0:t}\}), \tag{13}$$

where the support points $\{\boldsymbol{\theta}_{0:t}^i, z_{0:t}^i\}_{i=1,...,N}$, called the particles, are generated recursively using sequential importance sampling and the weights $\{w_t^i\}_{i=1,...,N}$ sum to one. As for the estimation of the state vector $\boldsymbol{x}_t$, each particle is associated with a Kalman filter (KF) that recursively updates the mean vector and the covariance matrix, denoted $\boldsymbol{\mu}_t^i$ and Σ_t^i respectively, of the following conditional distributions:

$$p(\boldsymbol{x}_t | z_{0:t}^i, \boldsymbol{\theta}_{0:t}^i, \boldsymbol{y}_{1:t}) = \mathcal{N}(\boldsymbol{x}_t; \boldsymbol{\mu}_t^i, \Sigma_t^i). \tag{14}$$

Finally, by marginalizing out $\boldsymbol{\theta}_{0:t}$, the posterior distribution of the state vector is approximated by a mixture of Gaussian distributions:

$$\widehat{p}(\boldsymbol{x}_t | \boldsymbol{y}_{1:t}) = \sum_{i=1}^{N} w_t^i \mathcal{N}(\boldsymbol{x}_t; \boldsymbol{\mu}_t^i, \Sigma_t^i) \tag{15}$$

Note that the particles are propagated according to the prior models (7), (9) and (11), hence the weights are merely proportional to the likelihood $p(\boldsymbol{y}_t | \boldsymbol{y}_{1:t-1}, z_t^i, \boldsymbol{\theta}_t^i)$. The latter is computed at the update step of the KF.

Based on (13) and (15), the state estimates are computed. We consider the maximum and the mean of the posterior distributions for the model class and the state vector respectively.

$$\hat{z}_t^{\text{RBPF}} = \underset{m\in\{1,\ldots,M\}}{\text{argmax}} \sum_{i=1}^{N} \delta_{z_t^i}(m) \text{ and } \hat{\boldsymbol{x}}_t^{\text{RBPF}} = \sum_{i=1}^{N} w_t^i \boldsymbol{\mu}_t^i \quad (16)$$

In the next section, our approach is applied in the context of target tracking.

4. APPLICATION TO TARGET TRACKING

We consider that the target can follow either a CV or a CA motion model at each instant.
In this case, the state vectors satisfy[1]:

$$\boldsymbol{x}_t^{CV} = [x_t, \dot{x}_t]^T \;\; ; \;\; \boldsymbol{x}_t^{CA} = [x_t, \dot{x}_t, \ddot{x}_t]^T \quad (17)$$

where x_t denotes the position, $\dot{x}_t$ the velocity and $\ddot{x}_t$ the acceleration. In addition, the transition and observation matrices of both model classes are defined as follows:

$$\begin{aligned} \Phi^{CV} &= \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \; ; \; \Phi^{CA} = \begin{bmatrix} 1 & T & T^2/2 \\ 0 & 1 & T \\ 0 & 0 & 1 \end{bmatrix} \\ H^{CV} &= [1\;0] \qquad ; \; H^{CA} = [1\;0\;0] \end{aligned} \quad (18)$$

where T is the sampling period. The model-noise covariance matrices for both motion models can be respectively defined as:

$$Q^{CV} = \sigma_{CV}^2 \begin{bmatrix} \frac{T^3}{2} & \frac{T^2}{2} \\ \frac{T^2}{2} & T \end{bmatrix} ; Q^{CA} = \sigma_{CA}^2 \begin{bmatrix} \frac{T^5}{20} & \frac{T^4}{8} & \frac{T^3}{6} \\ \frac{T^4}{8} & \frac{T^3}{3} & \frac{T^2}{2} \\ \frac{T^3}{6} & \frac{T^2}{2} & T \end{bmatrix} \quad (19)$$

where σ_{CV}^2 and σ_{CA}^2 are the acceleration and jerk variances respectively. The associated so-called precision parameters are thus $\gamma_{CV} = \frac{1}{\sigma_{CV}^2}$ and $\gamma_{CA} = \frac{1}{\sigma_{CA}^2}$.

4.1. Resulting setting of the hierarchical model

In this subsection, let us give some information about the model class index, the latent variable and the DP.
As two model classes are considered, $z_t \in \{1,\, 2\}$. The model class index refers to the CV-motion model and the CA-motion model when it is equal to 1 and 2 respectively. In addition, its Markov chain is characterized by the TPM denoted as Π_{DP}.
Given (19), the functional forms of both model-noise covariance matrices are known and only depend on the model-noise variances which can switch between different values over time. At each instant, the latent variable $\boldsymbol{\theta}_t$ hence contains γ_{z_t} which is the precision parameter (i.e. the variance inverse) of the motion model indexed by z_t.
In addition, as suggested in [4], the DPM base distribution G_0 is defined from a $Gamma$ conjugate prior $\Gamma(a_{z_t},\, b_{z_t})$ on the precision parameter of the hierarchical Dirichlet process. In order to consider a weakly informative setting for both motion models, $a_1 = 4$, $b_1 = 40$ and $a_2 = 10$, $b_2 = 0.01$. Both resulting distributions are represented in Fig. 2. Note that the span of possible values taken by $\{\gamma_{z_t}\}_{z_t=1,2}$ are not the same due to the different units of the respective model-noise variances.

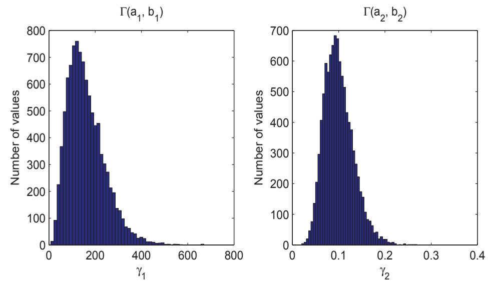

Fig. 2. $Gamma$ prior on the accuracy parameter γ

4.2. Simulation protocol

The relevance of our approach is analyzed by applying it to a set of trajectories generated from $M = 2$ motion models: one CA motion model and one CV motion model. At each instant ($T = 1$), the system can switch between both motion models according to the following TPM: $\begin{bmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{bmatrix}$.
As long as the target is described by the same model class, the model-noise standard deviation (std) remains unchanged and can take one among two possible values. Thus, $\sigma_{CV} \in \{5.10^{-2};\, 10^{-1}\} m.s^{-\frac{3}{2}}$ leading to $\gamma_{CV} \in \{400; 100\}$, while $\sigma_{CA} \in \{2.5,\, 4.5\} m.s^{-\frac{5}{2}}$ leading to $\gamma_{CA} \in \{0.05,\, 0.16\}$. In addition, the measurement-noise variance is set at $25m^2$. The results are obtained for target trajectories of 500 samples and our algorithm is run with $N = 2000$ particles. Three different settings are tested for the TPM. Thus, its diagonal elements denoted p_{ii} are respectively set at 0.95, 0.9 and 0.8 corresponding to a mean sojourn time of $20T$, $10T$ and $5T$.
Our hierarchical DP-based approach, denoted *DP*, is then compared with several filtering approaches:
1/ a KF based on a CA motion model, denoted *Kal-CA*, whose model-noise std is set at $\sigma_{CA} = 4.5m.s^{-\frac{5}{2}}$.
2/ an *IMM* combining two KFs [2]. The first one is based on a CV motion model with $\sigma_{CV} = 10^{-1} m.s^{-\frac{3}{2}}$. The second one is based on a CA motion model with $\sigma_{CA} = 4.5m.s^{-\frac{5}{2}}$. Once again, three different settings are tested for the TPM Π_{IMM}.
3/ A method derived from Caron's work [4], denoted by *DPM*. It was implemented with a CA motion model where all the elements of its model-noise covariance matrix are assumed unknown and are hence estimated.

4.3. Simulation results

First, let us give some comments on the computational costs. According to our tests, DP-based approaches are more computationally expensive than the *IMM* and *Kal-CA*. Nevertheless, compared to *DPM*, the computational cost of the proposed algorithm is lesser. Then, let us analyze the root mean square error (RMSE) between the state vector and its estimate for 500 samples. Note that the RMSE related to the measurement noise,

[1]Note that for the sake of simplicity, we focus on the x dimension, but for 2D or 3D target tracking, it should be generalized to the y and z axis.

i.e. without applying any filtering approach, is equal to 4.01.

Kal-CA	*IMM* [2]		
	$p_{ii} = 0.95$	$p_{ii} = 0.9$	$p_{ii} = 0.8$
3.48	3.22	3.39	3.61
DPM [4]	*DP*		
	$p_{ii} = 0.95$	$p_{ii} = 0.9$	$p_{ii} = 0.8$
3.34	**2.71**	2.89	3.02

Table 1. RMSE averaged over 200 Monte Carlo simulations.

According to Table 1., the RMSE of *DPM* is smaller than the one associated to *Kal-CA* because it allows the time-switching model-noise covariance matrix to be estimated over time. However, its RMSE is higher than the one of the IMM because this latter provides better estimates when the trajectory corresponds to a CV model.

In addition, the proposed approach *DP* outperforms both the implemented IMM algorithm and *DPM*. Indeed, the model-noise variances of the motion models are estimated at each instant with our approach whereas they are set at predefined values when an IMM is used. Moreover, the model class index[2] is better estimated with the proposed approach than with the IMM. This is illustrated in Fig. 3 where the model class index estimation obtained with both the IMM algorithm and our method is given for one realization of the target trajectory. When averaging the model index estimation over 500 Monte Carlo simulations, one can notice that the proposed approach estimates the actual model class index for 96% of the samples whereas it is 91% for the IMM. Indeed, due to the IMM merging strategy, the weights in favor of a given model class are not necessarily clear-cut.

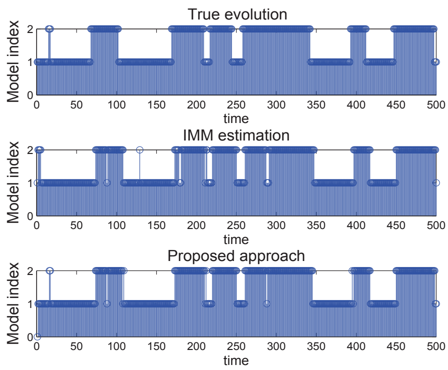

Fig. 3. Comparison of the model class estimation

Concerning the setting of the proposed approach and the IMM, if the diagonal elements of their corresponding TPMs are not set at 0.95 but at different values, their performance is slightly degraded. However, our approach is more reliable than the IMM since it does not require the setting of the model-noise variance. Finally, the relevance of our approach is confirmed by the estimation of the model-noise distribution on the position component x_t, which is compared with the actual one. See Fig. 4.

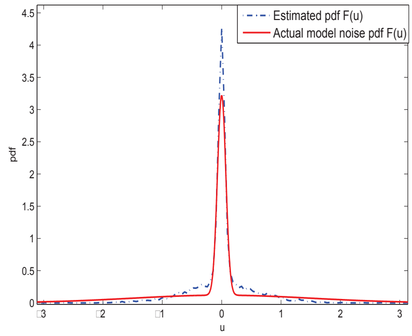

Fig. 4. Model-noise distribution estimate using *DP*

5. CONCLUSIONS AND PERSPECTIVES

In this paper, we propose a new DP-based hierarchical model for SLD systems. Our approach has the advantage of being flexible while decreasing the dimensionality of the estimation problem compared to a complete non-parametric method. We apply our approach in the field of target tracking and show its relevance compared to existing methods. As a perspective, we are investigating the performance of the proposed approach when the classes chosen for the estimator do not necessarily match the ones used for the target trajectory. We also plan to study issues such as stability and tracking speed [10].

[2]When using the IMM, the model class index at each instant corresponds to the model with the so-called highest *a posteriori* weight [2].

6. REFERENCES

[1] X. R. Li and V. P. Jilkov, "Multiple-Model Estimation with Variable Structure: Design and Evaluation of Model-Group Switching Algorithm," *IEEE Trans. on Aerospace and Elec. Syst.*, vol. 35, pp. 225–241, 1999.

[2] Y. Bar-Shalom and X. R. Li, *Multitarget-Multisensor Tracking: Principles and Techniques*, YBS Publishing, Storrs, 1995.

[3] X. R. Li and V. P. Jilkov, "Survey of Maneuvering Target Tracking. Part V: Multiple Model Methods," *IEEE Trans. on Aerospace and Elec. Syst.*, vol. 41, pp. 1255–1321, Oct. 2005.

[4] F. Caron, M. Davy, A. Doucet, E. Duflos, and P. Vanheeghe, "Bayesian Inference for Linear Dynamic Models with Dirichlet Process Mixtures," *IEEE Trans. on Signal Processing*, vol. 56, pp. 71–84, 2008.

[5] E. Fox, E. B. Sudderth, M. I. Jordan, and A. Willsky, "Bayesian Nonparametric Methods for Learning Markov Switching Processes," *IEEE Signal Processing Magazine*, vol. 27, pp. 43–54, 2010.

[6] E. Fox, E. B. Sudderth, M. I. Jordan, and A. Willsky, "Bayesian nonparametric inference of switching dynamic linear models," *IEEE Trans. on Signal Processing*, vol. 59, pp. 1569–1585, 2011.

[7] V. Pereira, G. Ferre, A. Giremus, and E. Grivel, "Relevance of Dirichlet Processes Mixtures for Modeling Interferences in Underlay Cognitive Radio," in *International Conf. on Signal Processing, EUSIPCO 2014*, 2014, pp. 176–180.

[8] A. Doucet, N. de Freitas, K. Murphy, and S. Russell, "Rao-Blackwellized Particle Filtering for Dynamic Bayesian Networks," in *Conf. Uncertainty Artificial Intell.*, 2000, pp. 176–183.

[9] D. Blackwell and B. MacQueen, "Ferguson Distributions Via Polya Urn Schemes," *The annals of statistics*, vol. 1, pp. 353–355, 1973.

[10] Lin Hai and P.J. Antsaklis, "Stability and Stabilizability of Switched Linear Systems: A Survey of Recent Results," *IEEE Trans. on Automatic Control*, vol. 54, pp. 308–322, 2009.