



Bayesian Inference for Dynamic Models with Dirichlet Process Mixtures

Francois Caron, Manuel Davy, Arnaud Doucet, Emmanuel Duflos, Philippe Vanheeghe

► **To cite this version:**

Francois Caron, Manuel Davy, Arnaud Doucet, Emmanuel Duflos, Philippe Vanheeghe. Bayesian Inference for Dynamic Models with Dirichlet Process Mixtures. 9th IEEE International Conference on Information Fusion, 2006, Florence, Italy. inria-00119993

HAL Id: inria-00119993

<https://hal.inria.fr/inria-00119993>

Submitted on 12 Dec 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Bayesian Inference for Dynamic Models with Dirichlet Process Mixtures

Francois Caron¹, Manuel Davy¹, Arnaud Doucet², Emmanuel Duflos¹, Philippe Vanheeghe¹

¹ LAGIS (UMR CNRS 8146), Ecole Centrale de Lille BP48, Villeneuve d'Ascq, France
firstname.lastname@ec-lille.fr

² Department of Statistics and Department of Computer Science,
University of British Columbia, Vancouver, Canada, arnaud@cs.ubc.ca

Abstract - *Using Kalman techniques, it is possible to perform optimal estimation in linear Gaussian state-space models. We address here the case where the noise probability density functions are of unknown functional form. A flexible Bayesian nonparametric noise model based on mixture of Dirichlet processes is introduced. Efficient Markov chain Monte Carlo and Sequential Monte Carlo methods are then developed to perform optimal estimation in such contexts.*

Keywords: Bayesian nonparametrics, Dirichlet Process Mixture, Monte Carlo Markov Chain, Rao-Blackwellisation, Particle filter.

1 Introduction

Let us consider the following dynamic linear model

$$\mathbf{x}_t = F_t \mathbf{x}_{t-1} + C_t \mathbf{u}_t + G_t \mathbf{v}_t \quad (1)$$

$$\mathbf{z}_t = H_t \mathbf{x}_t + \mathbf{w}_t \quad (2)$$

where \mathbf{x}_t is the hidden state vector, \mathbf{z}_t is the observation, \mathbf{v}_t and \mathbf{w}_t are sequence of independent random variables. F_t and H_t are the known state and observation matrices, \mathbf{u}_t is a known input, C_t the input transfert matrix and G_t is the state transfert matrix.

For any sequence $\{\mathbf{a}_t\}$, we write $\mathbf{a}_{i:j}$ for $(\mathbf{a}_i, \mathbf{a}_{i+1}, \dots, \mathbf{a}_j)$. We are here interested in estimating the hidden state \mathbf{x}_t given the observations $\mathbf{z}_{1:t}$ (filtering) or $\mathbf{z}_{1:T}$ for $T \geq t$ (smoothing). If the noise sequences are assumed Gaussian with known parameters, optimal estimation can be performed using the Kalman filter/smoother. However, these Gaussian assumptions can be very inappropriate for many real-world models. In this paper, we address the problem of *optimal state estimation when the probability density functions (pdf) of the noise sequences are unknown and need to be estimated on-line or off-line from the data*. To simplify the presentation, we will limit ourselves here to the estimation of the pdf of the state noise sequence $\{\mathbf{v}_t\}$

and will assume that $\{\mathbf{w}_t\}$ is Gaussian. However, our algorithms can be extended straightforwardly to the case where the observation noise pdf should also be estimated.

Our methodology relies on the introduction of a Dirichlet Process Mixtures (DPM) for the state noise pdf. DPM is a flexible Bayesian nonparametric model which have become very popular in statistics over the last few years to perform nonparametric density estimation. To the best of our knowledge, this powerful class of models has never been used in the context of dynamic models. We demonstrate here that such tools can help to improve drastically the performance of standard algorithms when the noise pdfs are unknown.

The rest of this paper is organized as follows. In Section 2, we recall the basics of Bayesian nonparametric density estimation and in Section 3 we derive the dynamic model with unknown noise distribution. In Section 4, we develop an efficient Markov chain Monte Carlo algorithm to perform optimal estimation in the batch case. In Section 5, we develop an efficient Sequential Monte Carlo/Particle filter to perform optimal estimation in the sequential case. All these algorithms can be interpreted as generalized Rao-Blackwellised methods. Finally, we demonstrate our algorithms on two applications: blind deconvolution of impulse processes and robust regression.

2 Bayesian nonparametric density estimation

We review here briefly modern Bayesian tools to perform nonparametric density estimation.

2.1 Density estimation

Let $\mathbf{y}_1, \dots, \mathbf{y}_n$ be a statistically exchangeable sequence distributed according to the pdf F

$$\mathbf{y}_k \sim F(\cdot).$$

We are interested here in estimating F and we consider the following nonparametric model

$$F(\mathbf{y}) = \int_{\Theta} f(\mathbf{y}|\theta) d\mathbb{G}(\theta) \quad (3)$$

where $\theta \in \Theta$ is called the latent variable or cluster variable, $f(\cdot|\theta)$ is the mixed pdf and \mathbb{G} is the mixing distribution. Under a Bayesian framework, it is assumed that \mathbb{G} is a Random Probability Measure (RPM) distributed according to a prior distribution (i.e., a distribution over the set of probability distributions). We will select here the RPM to follow a Dirichlet Process (DP) prior.

2.2 Dirichlet Processes

Ferguson [1] introduced the Dirichlet Process (DP) as a probability measure on the space of probability measures. Given a probability measure \mathbb{G}_0 on a (measurable) space $(\mathcal{T}, \mathcal{A})$ and a positive real number α , a probability distribution \mathbb{G} distributed according to a DP of base distribution \mathbb{G}_0 and scale factor α , denoted $\mathbb{G} \sim DP(\mathbb{G}_0, \alpha)$, satisfies for any partition A_1, \dots, A_k of \mathcal{T} and any k

$$(\mathbb{G}(A_1), \dots, \mathbb{G}(A_k)) \sim \mathcal{D}(\mathbb{G}_0(A_1), \dots, \mathbb{G}_0(A_k), \alpha)$$

where \mathcal{D} is a standard Dirichlet distribution. Sethuraman [2] established that the realizations of a Dirichlet process are discrete with probability one and admit the so-called stick-breaking representation

$$\mathbb{G} = \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k} \quad (4)$$

with $\theta_k \sim \mathbb{G}_0$, $\pi_k = \beta_k \prod_{j=1}^{k-1} (1 - \beta_j)$ and $\beta_k \sim \mathcal{B}(1, \alpha)$ where \mathcal{B} denotes the Beta distribution and δ_{θ_k} denotes the Dirac delta measure located in θ_k . Using (3), it comes that the following flexible prior model is adopted for the unknown distribution F

$$F(\mathbf{y}) = \sum_{k=1}^{\infty} \pi_k f(\mathbf{y}|\theta_k).$$

Apart from its flexibility, a fundamental motivation to use the DP model is the simplicity of the posterior update. Let $\theta_1, \dots, \theta_n$ be n random samples from \mathbb{G}

$$\theta_k | \mathbb{G} \stackrel{\text{i.i.d.}}{\sim} \mathbb{G}$$

where $\mathbb{G} \sim DP(\mathbb{G}_0, \alpha)$ then the posterior distribution of $\mathbb{G}|\theta_{1:n}$ is also a DP

$$\mathbb{G}|\theta_{1:n} \sim DP\left(\frac{\alpha}{\alpha+n}\mathbb{G}_0 + \frac{1}{\alpha+n} \sum_{k=1}^n \delta_{\theta_k}, \alpha+n\right)$$

Moreover, it can be shown that the predictive distributions, computed by integrating out the RPM \mathbb{G} , admits the following Polya urn representation [3]

$$\theta_{n+1}|\theta_{1:n} \sim \frac{1}{\alpha+n} \sum_{k=1}^n \delta_{\theta_k} + \frac{\alpha}{\alpha+n} \mathbb{G}_0.$$

2.3 Dirichlet Process Mixture

Using these modelling tools, it is now possible to reformulate the density estimation problem using the following hierarchical model known as DPM [4]

$$\begin{aligned} \mathbb{G} &\sim DP(\mathbb{G}_0, \alpha), \\ \theta_k | \mathbb{G} &\sim \mathbb{G} \\ \mathbf{y}_k | \theta_k &\sim f(\cdot|\theta_k) \end{aligned}$$

The objective of density estimation boils down to estimating the posterior distribution $p(\theta_{1:n}|\mathbf{y}_{1:n})$. Although DPM were introduced in the 70's, these models were too complex to handle numerically before the introduction of MCMC [5, 6]. They have now become extremely popular.

3 Dynamic Linear Model with Unknown Noise Distribution

In this article, we are interested in the class of models (1)-(2) where $\{\mathbf{w}_t\}$ are distributed according to $\mathbf{w}_t \sim \mathcal{N}(0, R_t)$ whereas the pdf of $\{\mathbf{v}_t\}$ is unknown. We adopt a flexible Bayesian nonparametric model for this noise pdf. $\{\mathbf{v}_t\}$ is supposed to be distributed according to a DPM of base mixed distribution $\mathcal{N}(\mu_t, \Sigma_t)$, scale parameter α and Normal-inverse Wishart base distribution \mathbb{G}_0 [7] denoted $\mathbb{G}_0 = \mathcal{NIW}(\mu_0, \kappa_0, \nu_0, \Lambda_0)$ with $\mu_0, \kappa_0, \nu_0, \Lambda_0$ are hyperparameters that are assumed to be known and fixed. To summarize, we have the following model

$$\begin{aligned} \mathbb{G} &\sim DP(\mathbb{G}_0, \alpha), \\ (\mu_t, \Sigma_t) &\sim \mathbb{G}, \\ \mathbf{v}_t &\stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_t, \Sigma_t). \end{aligned}$$

This model is much more flexible than a standard mixture of Gaussians. It is equivalent to say that \mathbf{v}_t is sampled from a fixed but unknown distribution F which admits the following representation

$$F(\mathbf{v}_t) = \int \mathcal{N}(\mathbf{v}_t; \mu, \Sigma) d\mathbb{G}(\mu, \Sigma). \quad (5)$$

F is a countable infinite mixture of Gaussians pdf of unknown parameters, and the mixing distribution \mathbb{G} is sampled from a Dirichlet process. We will denote $\theta_t = \{\mu_t, \Sigma_t\}$ the latent cluster variables giving the mean and covariance matrix for that cluster.

The MCMC algorithms available in the literature to estimate these Bayesian nonparametric models – e.g. [5, 6] – do not apply in our case because we do not observe the sequence $\{\mathbf{v}_t\}$ but only the observations $\{\mathbf{z}_t\}$ generated by the dynamic model (1)-(2). In the following sections, we propose two strategies to perform inference in this more complex framework. Despite the complexity of the model, we develop efficient computational methods based on a Rao-Blackwellisation approach.

4 MCMC algorithm

Assume we are interested in estimating the posterior $p(\mathbf{x}_{0:T}, \theta_{1:T} | \mathbf{z}_{1:T})$, where \mathbf{x} is the state vector and $\theta_t = \{\mu_t, \Sigma_t\}$ is the latent variable as defined above. This posterior satisfies

$$p(\mathbf{x}_{0:T}, \theta_{1:T} | \mathbf{z}_{1:T}) = p(\mathbf{x}_{0:T} | \theta_{1:T}, \mathbf{z}_{1:T}) p(\theta_{1:T} | \mathbf{z}_{1:T}). \quad (6)$$

Conditional upon θ_t , (1) may be rewritten as

$$\mathbf{x}_t = F_t \mathbf{x}_{t-1} + \mathbf{u}'_t(\theta_t) + G_t \mathbf{v}'_t(\theta_t)$$

where $\mathbf{u}'_t(\theta_t) = C_t \mathbf{u}_t + G_t \mu_t$ is a known input and $\mathbf{v}'_t(\theta_t)$ is a centered white Gaussian noise of known covariance matrix Σ_t . Thus $p(\mathbf{x}_{0:T} | \theta_{1:T}, \mathbf{z}_{1:T})$ is a Gaussian distribution whose parameters can be computed using a Kalman smoother [8] where the marginal posterior $p(\theta_{1:T} | \mathbf{z}_{1:T})$ can be approximated through MCMC using the following Gibbs sampler [9]:

Algorithm 1 Gibbs sampler to sample from $p(\theta_{1:T} | \mathbf{z}_{1:T})$

• **Initialization:** For $t = 1, \dots, T$, sample $\theta_t^{(1)}$

• **Iteration $i, i \geq 2$**

For $t = 1, \dots, T$,

sample $\theta_t^{(i)} \sim p(\theta_t | \mathbf{z}_{1:T}, \theta_{-t}^{(i)})$ where $\theta_{-t}^{(i)} = \{\theta_1^{(i)}, \dots, \theta_{t-1}^{(i)}, \theta_{t-1}^{(i-1)}, \dots, \theta_T^{(i-1)}\}$

The Gibbs sampler needs to sample from $p(\theta_t | \mathbf{z}_{1:T}, \theta_{-t})$ where

$$p(\theta_t | \mathbf{z}_{1:T}, \theta_{-t}) \propto p(\mathbf{z}_{1:T} | \theta_{1:T}) p(\theta_t | \theta_{-t}).$$

From the Polya urn representation, $p(\theta_t | \theta_{-t})$ equals

$$\frac{1}{\alpha + T - 1} \sum_{k=1, k \neq t}^T \delta_{\theta_k}(\theta_t) + \frac{\alpha}{\alpha + T - 1} \mathbb{G}_0(\theta_t).$$

Thus $p(\theta_t | \mathbf{z}_{1:T}, \theta_{-t})$ is proportional to

$$p(\mathbf{z}_{1:T} | \theta_{1:T}) \times \left(\sum_{k=1, k \neq t}^T \delta_{\theta_k}(\theta_t) + \alpha \mathbb{G}_0(\theta_t) \right).$$

We can sample from this density with a Metropolis-Hastings step, where the candidate pdf is the conditional prior $p(\theta_t | \theta_{-t})$. This is given by algorithm 2.

Algorithm 2 Metropolis-Hastings step to sample from $p(\theta_t | \mathbf{z}_{1:T}, \theta_{-t})$

• Sample a candidate cluster

$$\theta_t^{(i)*} \sim \frac{1}{\alpha + T - 1} \sum_{k=1, k \neq t}^T \delta_{\theta_k} + \frac{\alpha}{\alpha + T - 1} \mathbb{G}_0$$

• With probability $\rho(\theta_t^{(i)}, \theta_t^{(i)*})$ where

$$\rho(\theta_t^{(i)}, \theta_t^{(i)*}) = \min \left(1, \frac{p(\mathbf{z}_{1:T} | \theta_t^{(i)*}, \theta_{-t}^{(i)})}{p(\mathbf{z}_{1:T} | \theta_t^{(i)}, \theta_{-t}^{(i)})} \right)$$

set $\theta_t^{(i)} = \theta_t^{(i)*}$, otherwise $\theta_t^{(i)} = \theta_t^{(i-1)}$.

The computation of the acceptance probability requires to compute the likelihood $p(\mathbf{z}_{1:T} | \theta_t^{(i)}, \theta_{-t}^{(i)})$. This can be done in $O(T)$ operations using a Kalman filter. However, this has to be done for $t = 1, \dots, T$ and one finally obtains an algorithm of computational complexity in $O(T^2)$. Here, we propose to use instead the backward-forward recursion developed in [10], to obtain an algorithm of complexity in $O(T)$. This algorithm uses the following likelihood decomposition:

$$p(\mathbf{z}_{1:T} | \theta_{1:T}) = p(\mathbf{z}_{1:t-1} | \theta_{1:t-1}) p(\mathbf{z}_t | \theta_{1:t}, \mathbf{z}_{1:t-1}) \quad (7)$$

$$\times \int_{\mathcal{X}} p(\mathbf{z}_{t+1:T} | \mathbf{x}_t, \theta_{t+1:T}) p(\mathbf{x}_t | \mathbf{z}_{1:t}, \theta_{1:t}) d\mathbf{x}_t$$

with $p(\mathbf{z}_{t:T} | \mathbf{x}_{t-1}, \theta_{t:T}) =$

$$\int_{\mathcal{X}} p(\mathbf{z}_{t+1:T} | \mathbf{x}_{t-1}, \theta_{t:T}) p(\mathbf{z}_t, \mathbf{x}_t | \theta_t, \mathbf{x}_{t-1}) d\mathbf{x}_t \quad (8)$$

The first two terms of the r.h.s. in Eq. (7) are computed by a forward recursion based on the Kalman filter [10]. The third term can be evaluated by a backward recursion according to Eq. (8), see Appendix for details. Based on Eq. (7), the density $p(\theta_t | \mathbf{z}_{1:T}, \theta_{-t})$ is expressed by

$$p(\theta_t | \mathbf{z}_{1:T}, \theta_{-t}) \propto p(\theta_t | \theta_{-t}) p(\mathbf{z}_t | \theta_{1:t}, \mathbf{z}_{1:t-1})$$

$$\times \int_{\mathcal{X}} p(\mathbf{z}_{t+1:T} | \mathbf{x}_t, \theta_{t+1:T}) p(\mathbf{x}_t | \mathbf{z}_{1:t}, \theta_{1:t}) d\mathbf{x}_t$$

The full steps are given in algorithm 3.

It can be easily established that the simulated Markov chain $\{\theta_{1:T}^{(i)}\}$ is ergodic with limiting distribution $p(\theta_{1:T} | \mathbf{z}_{1:T})$. After N iterations of the algorithm,

Algorithm 3 MCMC algorithm to sample from $p(\theta_{1:T}|\mathbf{z}_{1:T})$

Initialization $i = 1$

- For $t = 1, \dots, T$, sample $\theta_t^{(1)}$.

Iteration $i, i \geq 2$

- Backward recursion: For $t = T, \dots, 1$, compute and store $P'_{t|t+1}(\theta_{t+1:T}^{(i-1)})$ and $P'_{t|t+1}(\theta_{t+1:T}^{(i-1)})m'_{t|t+1}(\theta_{t+1:T}^{(i-1)})$
- Forward recursion: For $t = 1, \dots, T$
 - Perform a Kalman filter step with $\theta_t = \theta_t^{(i-1)}$, store $m_{t|t}(\theta_{1:t-1}^{(i)}, \theta_t^{(i-1)})$ and $P_{t|t}(\theta_{1:t-1}^{(i)}, \theta_t^{(i-1)})$.
 - Metropolis-Hastings step

- * Sample a candidate cluster

$$\theta_t^{(i)*} \sim \frac{1}{\alpha + T - 1} \sum_{k=1, k \neq t}^T \delta_{\theta_k} + \frac{\alpha}{\alpha + T - 1} \mathbb{G}_0$$

- * Perform a Kalman filter step with $\theta_t = \theta_t^{(i)*}$, store $m_{t|t}(\theta_{1:t-1}^{(i)}, \theta_t^{(i)*})$ and $P_{t|t}(\theta_{1:t-1}^{(i)}, \theta_t^{(i)*})$

- * With probability $\rho(\theta_t^{(i)}, \theta_t^{(i)*})$ where

$$\rho(\theta_t^{(i)}, \theta_t^{(i)*}) = \min \left(1, \frac{p(\mathbf{z}_{1:T}|\theta_t^{(i)*}, \theta_{-t}^{(i)})}{p(\mathbf{z}_{1:T}|\theta_t^{(i)}, \theta_{-t}^{(i)})} \right)$$

set $\theta_t^{(i)} = \theta_t^{(i)*}$, otherwise $\theta_t^{(i)} = \theta_t^{(i-1)}$.

State post-sampling

- For $i = 1, \dots, N$, compute $\mathbf{x}_t^{(i)} = \mathbb{E}(\mathbf{x}_t|\theta_{1:T}^{(i)})$ for all t with a Kalman smoother.
-

the MMSE estimates of $\theta_{1:T}$ and $\mathbf{x}_{0:T}$ are computed using

$$\hat{\theta}_{1:T} = \frac{1}{N} \sum_{k=1}^N \theta_{1:T}^{(k)}, \quad \hat{\mathbf{x}}_t = \frac{1}{N} \sum_{k=1}^N \mathbf{x}_t^{(k)}$$

5 Rao-Blackwellized Particle Filter algorithm

In many applications, we want to process the data on-line. In this case, MCMC are inadequate as these are batch iterative algorithms. We propose here some original Sequential Monte Carlo methods, also known as particle filters, to approximate on-line the sequence of probability distributions $\{p(\mathbf{x}_{0:t}, \theta_{1:t}|\mathbf{z}_{1:t}), t = 1, 2, \dots\}$. Similar to the batch case, it is possible to exploit the structure of the dy-

namic model to reduce the dimension of the parameter space. We have the following decomposition of the posterior distribution

$$p(\mathbf{x}_{0:t}, \theta_{1:t}|\mathbf{z}_{1:t}) = p(\mathbf{x}_{0:t}|\theta_{1:t}, \mathbf{z}_{1:t})p(\theta_{1:t}|\mathbf{z}_{1:t}).$$

As $p(\mathbf{x}_{0:t}|\theta_{1:t}, \mathbf{z}_{1:t})$ can be computed using Kalman techniques, we only need to estimate through particle methods the marginal posterior $p(\theta_{1:t}|\mathbf{z}_{1:t})$. This is a generalization of the Rao-Blackwellized particle filter [11] to DPM. At time t , it follows that $p(\mathbf{x}_t, \theta_{1:t}|\mathbf{z}_{1:t})$ is approximated through a set of N particles $\theta_{1:t}^{(1)}, \dots, \theta_{1:t}^{(N)}$ by the following empirical distribution

$$P_N(\mathbf{x}_t, \theta_{1:t}|\mathbf{z}_{1:t}) = \sum_{i=1}^N \tilde{w}_t^{(i)} p(\mathbf{x}_t|\theta_{1:t}^{(i)}, \mathbf{z}_{1:t})$$

with

$$p(\mathbf{x}_t|\theta_{1:t}^{(i)}, \mathbf{z}_{1:t}) = \mathcal{N}(\hat{\mathbf{x}}_{t|t}(\theta_{1:t}^{(i)}), \Sigma_{t|t}(\theta_{1:t}^{(i)}))$$

The parameters $\hat{\mathbf{x}}_{t|t}(\theta_{1:t}^{(i)})$ and $\Sigma_{t|t}(\theta_{1:t}^{(i)})$ are computed recursively for each particle i using the Kalman filter [8]. In order to build the algorithm, we note that

$$p(\theta_{1:t}^{(i)}|\mathbf{z}_{1:t}) \propto p(\theta_{1:t-1}^{(i)}|\mathbf{z}_{1:t-1})p(\mathbf{z}_t|\theta_{1:t}^{(i)}, \mathbf{z}_{1:t-1})p(\theta_t^{(i)}|\theta_{1:t-1}^{(i)})$$

where

$$\begin{aligned} p(\mathbf{z}_t|\theta_{1:t}^{(i)}, \mathbf{z}_{1:t-1}) &= p(\mathbf{z}_t|\theta_t^{(i)}, \theta_{1:t-1}^{(i)}, \mathbf{z}_{1:t-1}) \\ &= \mathcal{N}(\hat{\mathbf{z}}_t(\theta_{1:t}^{(i)}), S_t(\theta_{1:t}^{(i)})) \end{aligned}$$

and

$$\begin{aligned} \hat{\mathbf{z}}_t(\theta_{1:t}^{(i)}) &= H_t \left[F_t \left\{ \hat{\mathbf{x}}_{t-1|t-1}(\theta_{1:t-1}^{(i)}) \right\} + C_t \mathbf{u}_t + G_t \mu_t^{(i)} \right] \\ S_t(\theta_{1:t}^{(i)}) &= H_t \left[F_t \left\{ \Sigma_{t-1|t-1}(\theta_{1:t-1}^{(i)}) \right\} F_t^{\mathbf{T}} + G_t \Sigma_t^{(i)} G_t^{\mathbf{T}} \right] \\ &\quad \times H_t^{\mathbf{T}} + R_t \end{aligned}$$

The Rao-Blackwellized Particle Filter (RBPF) steps are given in algorithm 4.

As in other particle filtering algorithms, the performance depends highly on the importance distribution selected. Here, the optimal importance distribution is $q(\theta_t|\theta_{1:t-1}^{(i)}, \mathbf{z}_{1:t}) = p(\theta_t|\theta_{1:t-1}^{(i)}, \mathbf{z}_{1:t})$, which equals

$$\begin{aligned} &\frac{\sum_{j=1}^{t-1} \mathcal{N}(\mathbf{z}_t; \hat{\mathbf{z}}_t(\theta_{1:t-1}^{(i)}, \theta_t), S_t(\theta_{1:t-1}^{(i)}, \theta_t)) \delta_{\theta_j^{(i)}}(\theta_t)}{\sum_{j=1}^{t-1} \mathcal{N}(\mathbf{z}_t; \hat{\mathbf{z}}_t(\theta_{1:t-1}^{(i)}, \theta_j), S_t(\theta_{1:t-1}^{(i)}, \theta_j)) + \alpha I(\theta_{1:t-1}^{(i)})} \\ &+ \frac{\alpha I(\theta_{1:t-1}^{(i)}) \times H_0(\theta_{1:t-1}^{(i)}, \theta_t)}{\sum_{j=1}^{t-1} \mathcal{N}(\mathbf{z}_t; \hat{\mathbf{z}}_t(\theta_{1:t-1}^{(i)}, \theta_j), S_t(\theta_{1:t-1}^{(i)}, \theta_j)) + \alpha I(\theta_{1:t-1}^{(i)})} \end{aligned}$$

with

$$\begin{aligned} H_0(\theta_{1:t-1}^{(i)}, \theta_t) &= \frac{\mathcal{N}(\mathbf{z}_t; \hat{\mathbf{z}}_t(\theta_{1:t-1}^{(i)}, \theta_t), S_t(\theta_{1:t-1}^{(i)}, \theta_t)) \mathbb{G}_0(\theta_t)}{\int_{\Theta} \mathcal{N}(\mathbf{z}_t; \hat{\mathbf{z}}_t(\theta_{1:t-1}^{(i)}, \theta_t), S_t(\theta_{1:t-1}^{(i)}, \theta_t)) \mathbb{G}_0(\theta_t) d\theta_t}, \\ I(\theta_{1:t-1}^{(i)}) &= \int_{\Theta} \mathcal{N}(\hat{\mathbf{z}}_t(\theta_{1:t-1}^{(i)}, \theta_t), S_t(\theta_{1:t-1}^{(i)}, \theta_t)) \mathbb{G}_0(\theta_t) d\theta_t. \end{aligned}$$

Algorithm 4 Rao-Blackwellised Particle Filter to sample from $p(\theta_{1:t}|\mathbf{z}_{1:t})$

At time 1.

- For $i = 1, \dots, N$, sample $(\mathbf{x}_0^{(i)}, \Sigma_{0|0}^{(i)}) \sim p_0(\mathbf{x}_0)$.

- Set $w_0^{(i)} \leftarrow \frac{1}{N}$

At each time t ($t \geq 2$)

- For $i = 1, \dots, N$
 - Sample $\tilde{\theta}_t^{(i)} \sim q(\theta_t|\theta_{1:t-1}, \mathbf{z}_{1:t})$
 - Compute

$$(\hat{\mathbf{x}}_{t|t-1}(\theta_{1:t}^{(i)}), \Sigma_{t|t-1}(\theta_{1:t}^{(i)}), \hat{\mathbf{x}}_{t|t}(\theta_{1:t}^{(i)}), \Sigma_{t|t}(\theta_{1:t}^{(i)})) \\ = KF_1step(\hat{\mathbf{x}}_{t-1|t-1}(\theta_{1:t-1}^{(i)}), \Sigma_{t-1|t-1}(\theta_{1:t-1}^{(i)}), \tilde{\theta}_t^{(i)}, \mathbf{z}_t)$$

- For $i = 1, \dots, N$, update the weights

$$\tilde{w}_t^{(i)} \propto w_{t-1}^{(i)} \frac{p(\mathbf{z}_t|\theta_{1:t-1}^{(i)}, \tilde{\theta}_t^{(i)}, \mathbf{z}_{1:t-1})p(\tilde{\theta}_t^{(i)}|\theta_{1:t-1}^{(i)})}{q(\tilde{\theta}_t^{(i)}|\theta_{1:t-1}^{(i)}, \mathbf{z}_{1:t})}, \\ \sum_{i=1}^N \tilde{w}_t^{(i)} = 1.$$

- Compute $N_{\text{eff}} = \left[\sum_{i=1}^N (\tilde{w}_t^{(i)})^2 \right]^{-1}$

- If $N_{\text{eff}} \leq \eta$, resample the particles and set $w_t^{(i)} = \frac{1}{N}$
-

The weights are then updated with

$$\tilde{w}_t^{(i)} \propto w_{t-1}^{(i)} I(\theta_{1:t-1}^{(i)}).$$

However, this optimal importance distribution cannot be used, as the associated importance weights do not admit a closed-form expression. However it is possible to derive various approximations of it so as to design efficient importance distributions $q(\theta_t|\theta_{1:t-1}, \mathbf{z}_{1:t})$.

From the particles, the MMSE estimate and posterior covariance matrix of \mathbf{x}_t are given by

$$\hat{\mathbf{x}}_{t|t} = \sum_{i=1}^N \tilde{w}_t^{(i)} \hat{\mathbf{x}}_{t|t}^{(i)}$$

$$\Sigma_{t|t} = \sum_{i=1}^N \tilde{w}_t^{(i)} \left[\Sigma_{t|t}^{(i)} + (\hat{\mathbf{x}}_{t|t}^{(i)} - \hat{\mathbf{x}}_{t|t})(\hat{\mathbf{x}}_{t|t}^{(i)} - \hat{\mathbf{x}}_{t|t})^T \right]$$

6 Applications

In this section, we present two applications of the above model and algorithms. We address, first, blind deconvolution and, second, robust regression. In both cases, we assume that the statistics of the state noise are unknown, and modelled as a DPM.

6.1 Blind deconvolution of impulse processes

Blind deconvolution finds many applications in various fields of engineering and physics, such as image

de-blurring, spectroscopic data analysis, audio source restoration, etc. We follow here the model presented in [12], which is recalled below.

6.1.1 Statistical Model

Let $H = \begin{pmatrix} 1 & h_1 & \dots & h_L \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{h} \end{pmatrix}$ and $\mathbf{x}_t = \begin{pmatrix} v_t & v_{t-1} & \dots & v_{t-L} \end{pmatrix}^T$. The observed signal z_t is the convolution of the sequence \mathbf{x}_t with a finite impulse response filter H , observed in additive white Gaussian noise w_t . The observation model is then

$$z_t = H\mathbf{x}_t + w_t$$

where $w_t \sim \mathcal{N}(0, \sigma_w^2)$ with σ_w^2 is the assumed known variance of w_t . The state space model can be written as follows:

$$\mathbf{x}_t = F\mathbf{x}_{t-1} + Gv_t$$

where $F = \begin{pmatrix} 0 & 0_{1 \times L} \\ 0_{L \times 1} & I_L \end{pmatrix}$, $G = \begin{pmatrix} 1 \\ 0_{L \times 1} \end{pmatrix}$, $0_{m \times n}$ is the null matrix of size $m \times n$ and I_m is the identity matrix of size $m \times m$. The state transition noise v_t is supposed to be independent from w_t , and distributed according to the mixture

$$v_t \sim \lambda F + (1 - \lambda)\delta_0 \quad (9)$$

where δ_0 is the Dirac delta function at 0 and F is a DPM of Gaussians defined in Eq. (5). In other words, the noise is alternatively zero, or distributed according to a DPM of Gaussians.

For simplicity reasons, we introduce latent Bernoulli variables $r_t \in \{0, 1\}$ such that $\Pr(r_t = 1) = \lambda$ and $v_t|(r_t = 1) \sim f(\cdot|\theta_t)$, $v_t|(r_t = 0) \sim \delta_0$. Consider the cluster variable ϕ_t defined by $\phi_t = \theta_t$ if $r_t = 1$ and $\phi_t = (0, 0)$ (i.e. parameters corresponding to the delta-mass) if $r_t = 0$, that is, $\phi_t \sim \lambda F + (1 - \lambda)\delta_{(0,0)}$. By integrating out F , one has

$$\phi_t|\phi_{-t} \sim \lambda p(\phi_t|\phi_{-t}, r_t = 1) + (1 - \lambda)\delta_{(0,0)}$$

where $p(\phi_t|\phi_{-t}, r_t = 1)$ is the Polya urn representation on the set $\tilde{\phi}_{-t} = \{\phi \in \phi_{-t} | \phi \neq \delta_{(0,0)}\}$ of size T' given by

$$\phi_t|(\phi_{-t}, r_t = 1) \sim \frac{\sum_{k=1, k \neq t}^{T'} \delta_{\tilde{\phi}_k} + \alpha \mathbb{G}_0}{\alpha + T'}$$

The hyperparameters are $\Phi = (\alpha \quad \mathbf{h} \quad \lambda)$ (the hyperparameters of the base distribution \mathbb{G}_0 are assumed fixed and known). These hyperparameters are assumed random with prior distribution $p(\Phi) = p(\alpha)p(\mathbf{h})p(\lambda)$, where

$$p(\alpha) = \mathcal{G}\left(\frac{\eta}{2}, \frac{\nu}{2}\right), \quad p(\mathbf{h}) = \mathcal{N}(0, \Sigma_{\mathbf{h}}), \quad p(\lambda) = \mathcal{B}(\zeta, \tau)$$

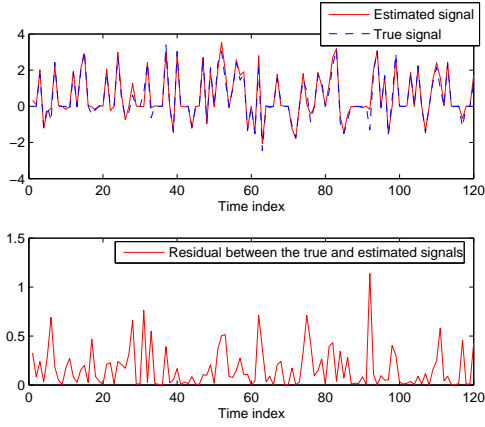


Figure 1: Top picture: True and estimated signal $\mathbf{v}_{1:T}$. Bottom picture: residual between the true and estimated signals.

where η , ν , $\Sigma_{\mathbf{h}}$, ζ and τ are known. We have the following conditional posteriors

$$p(\mathbf{h}|\mathbf{x}_{1:T}, \phi_{1:T}, \alpha, \lambda, \mathbf{z}_{1:T}) \propto p(\mathbf{h}) \times p(\mathbf{z}_{1:T}|\mathbf{x}_{1:T}, \mathbf{h})$$

and $p(\lambda|\mathbf{x}_{1:T}, \phi_{1:T}, \alpha, \mathbf{h}, \mathbf{z}_{1:T}) = p(\lambda|\mathbf{r}_{1:T})$ with

$$p(\lambda|\mathbf{r}_{1:T}) = \mathcal{B}\left(\zeta + \sum_{t=1}^T r_t, \tau + \sum_{t=1}^T (1 - r_t)\right)$$

where $r_t = 0$ if $\phi_t^v = (0, 0)$ and $r_t = 1$ otherwise. The aim is to approximate by MCMC the joint posterior pdf $p(\mathbf{v}_{1:T}, \phi_{1:T}, \Phi|\mathbf{z}_{1:T})$. This is done by applying Algorithm 3 for the cluster variable, whereas the other variables are sampled by Metropolis-Hastings or direct sampling w.r.t their conditional posterior.

6.1.2 Simulation results

This model has been simulated with the following parameters: $T = 120$, $L = 3$, $\mathbf{h} = (-1.5 \ 0.5 \ -0.2)$, $\lambda = 0.6$, $\sigma_w^2 = 0.1$, $F = 0.7\mathcal{N}(2, 1) + 0.3\mathcal{N}(-1, 1)$. For the estimation, 5000 MCMC iterations are performed. Fig. 1 (top) displays the MMSE estimate of $\mathbf{v}_{1:T}$ together with its true value. As can be seen in Fig. 1 (bottom), the residual is very small. Also, as can be seen in Fig. 2, the estimated pdf F is quite close to the true one.

6.2 Robust Regression

We now consider the problem of robust regression as described in [13]. The statistical model considered here is presented below.

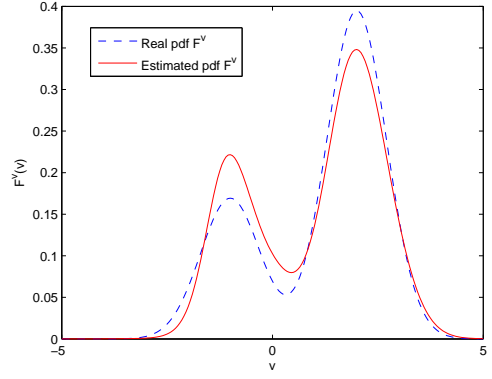


Figure 2: True and estimated pdf F

6.2.1 Statistical model

Let us consider the problem of estimating a continuous regression function $g(\cdot)$ defined on a domain $\mathcal{I} \subset \mathbb{R}$. Suppose we have T observations z_t , $t = 1, \dots, T$ in additive Gaussian noise

$$z_t = g(y_t) + w_t$$

where $y_0 < y_1 < \dots < y_T$ and $w_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma_w^2)$. The function $g(\cdot)$ is supposed to follow the stochastic differential equation $g''(y) = v(y|\theta_t) \forall y \in [y_{t-1}, y_t]$, where $v(y|\theta_t) \sim \mathcal{N}(\mu_t, \sigma_t^2)$, i.e., the second derivative of the function g is piecewise constant and follows a normal pdf of mean and variance given by $\theta_t = \{\mu_t, \sigma_t^2\}$. Conditional on θ_t , this may be formulated under the following state space form

$$\mathbf{x}_t = F\mathbf{x}_{t-1} + \mathbf{u}'_t + v'_t$$

$$z_t = H\mathbf{x}_t + w_t$$

where $\mathbf{x}_t = (g(y_t) \ g'(y_t))^T$, $F = \begin{pmatrix} 1 & \Delta t \\ 0 & 1 \end{pmatrix}$, $\mathbf{u}'_t = \begin{pmatrix} \frac{\Delta t^2}{2} \\ \Delta t \end{pmatrix} \mu_t$, \mathbf{v}'_t are independent $\mathcal{N}(0, \sigma_t^2 \times V(\Delta t))$, $V(\Delta t) = \begin{pmatrix} \frac{\Delta t^3}{3} & \frac{\Delta t^2}{2} \\ \frac{\Delta t^2}{2} & \Delta t \end{pmatrix}$, $H = (1 \ 0)$. Δt is the constant sampling period.

6.2.2 Simulation results

We present simulation results obtained with the following settings: $\sigma_w^2 = 0.1$, $\Delta t = 1$, $T = 50$. The true function g is zero over $[0, 10]$, it switches to $4 + \text{sinc}(0.2 * (t - 10))$ over $[10, 20]$ and switches again to $-5 * \text{sinc}(0.2(t - 20))$ over $[20, 50]$. The regression function has been estimated with both the MCMC algorithm (5000 iterations) and the Rao-Blacwellized particle filter (1000 particles) respectively. The true and estimated regression functions, together with the measurements, are plotted in Figures 3 and 4.

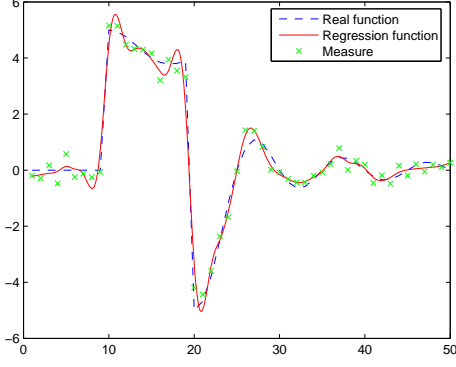


Figure 3: True and regression functions with the MCMC algorithm with 5000 particles.

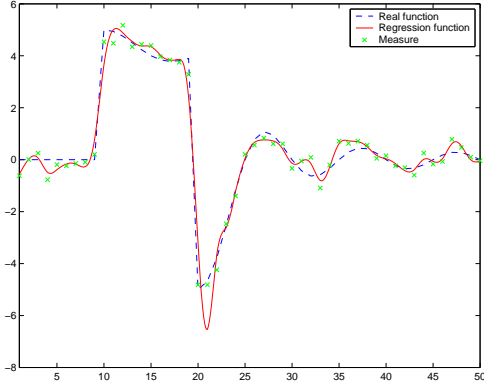


Figure 4: True and regression functions with the Rao-Blackwellized particle filter algorithm with 1000 particles

7 Conclusion

In this article we have presented a Bayesian nonparametric model which allows us to estimate the state noise pdf in a linear dynamic model. The methodology presented here can be straightforwardly extended to estimate the observation noise pdf. The Dirichlet process mixture considered here is flexible and we have presented two simulation-based algorithms based on Rao-Blackwellisation which allows us to perform efficiently inference. We are currently investigating the following extensions of our methodology. First, it would be of interest to consider nonlinear dynamic models. Second, it would be important to develop non-stationary Dirichlet process mixture models in cases where the noise statistics are assumed to evolve over time.

A Backward forward recursion

We recall that, knowing θ_t , the state space model may be written as

$$\mathbf{x}_t = F_t \mathbf{x}_{t-1} + \mathbf{u}'_t(\theta_t) + G_t \mathbf{v}'_t(\theta_t)$$

with $\mathbf{u}'_t(\theta_t) = C_t \mathbf{u}_t + G_t \mu_t$ is a known input and $\mathbf{v}'_t(\theta_t)$ is a centered white Gaussian noise of known covariance matrix Σ_t .

The forward backward algorithm uses the following likelihood decomposition:

$$p(\mathbf{z}_{1:t} | \theta_{1:t}) = p(\mathbf{z}_{1:t-1} | \theta_{1:t-1}) p(\mathbf{z}_t | \theta_{1:t}, \mathbf{z}_{1:t-1}) \quad (10)$$

$$\times \int_{\mathcal{X}} p(\mathbf{z}_{t+1:T} | \mathbf{x}_t, \theta_{t+1:T}) p(\mathbf{x}_t | \mathbf{z}_{1:t}, \theta_{1:t}) d\mathbf{x}_t \quad (11)$$

with

$$\begin{aligned} & p(\mathbf{z}_{t:T} | \mathbf{x}_{t-1}, \theta_{t:T}) \\ &= \int_{\mathcal{X}} p(\mathbf{z}_{t+1:T} | \mathbf{x}_{t-1}, \theta_{t:T}) p(\mathbf{z}_t, \mathbf{x}_t | \theta_t, \mathbf{x}_{t-1}) d\mathbf{x}_t \quad (12) \end{aligned}$$

It is shown in [10] that if $\int_{\mathcal{X}} p(\mathbf{z}_{t:T} | \mathbf{x}_{t-1}, \theta_{t:T}) d\mathbf{x}_{t-1} < \infty$ then $\frac{p(\mathbf{z}_{t:T} | \mathbf{x}_{t-1}, \theta_{t:T})}{\int_{\mathcal{X}} p(\mathbf{z}_{t:T} | \mathbf{x}_{t-1}, \theta_{t:T}) d\mathbf{x}_{t-1}}$ is a Gaussian distribution w.r.t. \mathbf{x}_{t-1} , of mean $m'_{t-1|t}(\theta_{t:T})$ and covariance $P'_{t-1|t}(\theta_{t:T})$. $P'_{t-1|t}(\theta_{t:T})$ and $P'_{t-1|t}(\theta_{t:T}) m'_{t-1|t}(\theta_{t:T})$ always satisfy the following backward information filter recursion.

- Initialization

$$P'_{T|T}(\theta_T) = H_T^T R_T^{-1} H_T$$

$$P'_{T|T}(\theta_T) m'_{T|T}(\theta_T) = H_T^T R_T^{-1} \mathbf{z}_T$$

- Backward recursion. For $t = T - 1, 1$,

$$\Delta_{t+1} = \left[I_{n_v} + B^T(\theta_{t+1}) P'_{t+1|t+1}(\theta_{t+1:T}) B(\theta_{t+1}) \right]^{-1}$$

$$\begin{aligned} P'_{t|t+1}(\theta_{t+1:T}) &= F_{t+1}^T P'_{t+1|t+1}(\theta_{t+1:T}) \\ &\quad \times (I_{n_x} - B(\theta_{t+1}) \Delta_{t+1}(\theta_{t+1:T})) \\ &\quad \times B^T(\theta_{t+1}) P'_{t+1|t+1}(\theta_{t+1:T}) F_{t+1} \end{aligned}$$

$$\begin{aligned} & P'_{t|t+1}(\theta_{t+1:T}) m'_{t|t+1}(\theta_{t+1:T}) \\ &= F^T(\theta_{t+1}) \times (I_{n_x} - P'_{t+1|t+1}(\theta_{t+1:T})) \\ &\quad \times B(\theta_{t+1}) \Delta_{t+1}(\theta_{t+1:T}) B^T(\theta_{t+1}) \\ &\quad \times P'_{t+1|t+1}(\theta_{t+1:T}) \left(m'_{t+1|t+1}(\theta_{t+1:T}) - \mathbf{u}'_{t+1}(\theta_{t+1}) \right) \end{aligned}$$

$$P'_{t|t}(\theta_{t:T}) = P'_{t|t+1}(\theta_{t+1:T}) + H_t^T R_t^{-1} H_t$$

$$\begin{aligned} P'_{t|t}(\theta_{t:T}) m'_{t|t}(\theta_{t:T}) &= P'_{t|t+1}(\theta_{t+1:T}) m'_{t|t+1}(\theta_{t+1:T}) \\ &\quad + H_t^T R_t^{-1} \mathbf{z}_t \end{aligned}$$

where $B(\theta_t) = G_t \times \text{chol}(\Sigma_t)^T$.

For the Metropolis Hasting ratio, we need to compute the acceptance probability only with a probability constant

$$p(\mathbf{z}_{1:T}|\theta_{1:T}) \propto p(\mathbf{z}_t|\theta_{1:t}, \mathbf{z}_{1:t-1}) \times \int_{\mathcal{X}} p(\mathbf{z}_{t+1:T}|\mathbf{x}_t, \theta_{t+1:T})p(\mathbf{x}_t|\mathbf{z}_{1:t}, \theta_{1:t})d\mathbf{x}_t \quad (13)$$

If $P_{t|t}(\theta_{1:t}) \neq 0$ then it exists $\Pi_{t|t}(\theta_{1:t})$ and $Q_{t|t}(\theta_{1:t})$ such that $P_{t|t}(\theta_{1:t}) = Q_{t|t}(\theta_{1:t})\Pi_{t|t}(\theta_{1:t})Q_{t|t}^T(\theta_{1:t})$. The matrices $Q_{t|t}(\theta_{1:t})$ and $\Pi_{t|t}(\theta_{1:t})$ are straightforwardly obtained using the singular value decomposition of $P_{t|t}(\theta_{1:t})$. Matrix $\Pi_{t|t}(\theta_{1:t})$ is a $n_t \times n_t$, $1 \leq n_t \leq n_x$ diagonal matrix with the nonzero eigenvalues of $P_{t|t}(\theta_{1:t})$ as elements. Then one has

$$p(\mathbf{z}_{1:T}|\theta_{1:T}) \propto \mathcal{N}(\tilde{\mathbf{z}}_{t|t-1}(\theta_{1:t}), S_t(\theta_{1:t})) \times \left[\Pi_{t|t}(\theta_{1:t})Q_{t|t}^T(\theta_{1:t})P'_{t|t+1}^{-1}(\theta_{t+1:T})Q_{t|t}(\theta_{1:t}) + I_{n_t} \right]^{-\frac{1}{2}} \times \exp\left(-\frac{1}{2}m_{t|t}^T(\theta_{1:t})P'_{t|t+1}^{-1}(\theta_{t+1:T})m_{t|t}(\theta_{1:t}) - 2m_{t|t}^T(\theta_{1:t})P'_{t|t+1}^{-1}(\theta_{t+1:T})m'_{t|t+1}(\theta_{t+1:T}) - (m'_{t|t+1}(\theta_{t+1:T}) - m_{t|t}(\theta_{1:t}))^T \times P'_{t|t+1}^{-1}(\theta_{t+1:T})A_{t|t}(\theta_{1:t}) \times P'_{t|t+1}^{-1}(\theta_{t+1:T})(m'_{t|t+1}(\theta_{t+1:T}) - m_{t|t}(\theta_{1:t}))\right)$$

where

$$A_{t|t}(\theta_{1:t}) = Q_{t|t}(\theta_{1:t}) \times \left[\Pi_{t|t}^{-1}(\theta_{1:t}) + Q_{t|t}^T(\theta_{1:t})P'_{t|t+1}^{-1}(\theta_{t+1:T})Q_{t|t}(\theta_{1:t}) \right]^{-1} \times Q_{t|t}^T(\theta_{1:t})$$

The quantities $m_{t|t}(\theta_{1:t})$, $P_{t|t}(\theta_{1:t})$, $\tilde{\mathbf{z}}_{t|t-1}(\theta_{1:t})$ and $S_t(\theta_{1:t})$ are, resp., the one-step ahead filtered estimate and covariance matrix of \mathbf{x}_t , the innovation at time t , and the covariance of this innovation. These quantities are given by the Kalman filter, the system being linear Gaussian conditional upon $\theta_{1:t}$.

Acknowledgment

François Caron is supported by the Centre National de la Recherche Scientifique (CNRS) and the Région Nord-Pas de Calais.

References

- [1] T.S. Ferguson. A Bayesian analysis of some nonparametric problems. *The annals of statistics*, 1:209–230, 1973.
- [2] J. Sethuraman. A constructive definition of Dirichlet priors. *Statistica Sinica*, 4:639–650, 1994.

- [3] D. Blackwell and J.B. MacQueen. Ferguson distributions via Polya urn schemes. *The annals of statistics*, 1:353–355, 1973.
- [4] C.E. Antoniak. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *The annals of statistics*, 2:1152–1174, 1974.
- [5] M.D. Escobar and M. West. Bayesian density estimation and inference using mixtures. *Journal of the american statistical association*, 90:577–588, 1995.
- [6] R.M. Neal. Markov chain sampling methods for Dirichlet process mixture models. *Journal of computational and graphical statistics*, 9:249–265, 2000.
- [7] A. Gelman, J.B. Carlin, H.S. Stern, and D.B. Rubin. *Bayesian data analysis*. Chapman and Hall, 1995.
- [8] B.D.O. Anderson and J.B. Moore. *Optimal filtering*. Prentice-Hall, 1979.
- [9] C.P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer-Verlag, 1999.
- [10] A. Doucet and C. Andrieu. Iterative algorithms for state estimation of jump Markov linear systems. *IEEE transactions on signal processing*, 49(6):1216–1227, 2001.
- [11] Arnaud Doucet, Nando de Freitas, and Neil Gordon, editors. *Sequential Monte Carlo Methods in practice*. Springer-Verlag, 2001.
- [12] A. Doucet and P. Duvaut. Bayesian estimation of state-space models applied to deconvolution of Bernoulli-Gaussian processes. *Signal Processing*, 57:147–161, 1997.
- [13] M. West, P. Muller, and M.D. Escobar. Hierarchical priors and mixture models, with application in regression and density estimation. In P.R. Freeman and A.F.M. Smith, editors, *Aspects of uncertainty*, pages 363–386. John Wiley, 1994.