

A Particle filtering Technique for Jump Markov Systems

Christophe Andrieu*, Manuel Davy† and Arnaud Doucet‡

ABSTRACT

This paper presents a particle filtering strategy in order to estimate the state of Jump Markov Systems (JMS). These processes are often met in signal communications, when the Bayesian model changes with time. Our algorithm takes advantage of the structure of the process.

1 Introduction

Many estimation problems arising in communications can be cast in the following form. A unique dynamic model represents the *signal* of interest $\{x_t\}$ ($t \geq 0$ and $x_t \in \mathbb{R}^{n_x}$), which is not observed but statistically related to the *observations* $\{y_t\}$ ($t \geq 1$ and $y_t \in \mathbb{R}^{n_y}$). More formally the process is typically described with a Markov transition density $p(x_t|x_{0:t-1}) = f(x_t|x_{t-1})$, where, for a set of variables l_t , we denote $l_{a:b} \triangleq \{l_a, l_{a+1}, \dots, l_b\}$. The observations $y_{1:t}$ are usually assumed to be independent conditional upon the signal process $\{x_t\}$, and marginally distributed according to $p(y_t|x_{0:t}) = g(y_t|x_t)$. One is then interested in estimating the sequence of posterior densities $p(x_{0:t}|y_{1:t})$ and typically their marginals $p(x_t|y_{1:t})$. This so-called optimal filtering problem generally does not admit a closed-form solution, and one usually resorts to particle filtering techniques [1].

The problem addressed in this paper is that of the development of efficient particle filtering techniques to perform on-line detection and estimation for a very important class of dynamical systems, named Jump Markov Systems (JMS) or Markov Switching State Space Models. JMS are a class of models that extends significantly the models described above and, as we shall see, specific particle filtering methods can be developed to perform optimal estimation.

1.1 Jump Markov Systems

1.1.1 Statistical Model

Let $\{r_t\}$ be a stationary, finite, discrete, first order homogeneous Markov chain taking its values in a set S , with transition probabilities $\pi_{ij} \triangleq \Pr\{r_{t+1} = j|r_t =$

$i\}$, ($i, j \in S$). We define s the finite number of elements of S . Now consider a family of s^2 densities $\{f_{ij}(x'|x)\}$ where $x \in \mathbb{R}^{n_x}$ and $x' \in \mathbb{R}^{n_{x'}}$, and define the state transition conditional densities,

$$p(x_t|x_{1:t-1}, r_{1:t}) = f_{r_{t-1}r_t}(x_t|x_{t-1}). \quad (1)$$

The initial state x_0 is distributed according to a distribution h . Note that *the dimension, or nature, of x_t might be a function of the sequence $\{r_t\}$* , but we do not make this dependence explicit in order to alleviate notation. Neither the process $\{r_t\}$ nor $\{x_t\}$ are observed. Instead, we observe $\{y_t\}$ where

$$p(y_t|x_{0:t}, r_{1:t}, y_{1:t-1}) = g_{r_t}(y_t|y_{1:t-1}, x_t), \quad (2)$$

with $y_t \in \mathbb{R}^{n_{y_t}}$ (the number of observations can vary over time). It is possible to add exogenous variables $\{u_t\}$ in the equations, *i.e.* $\{f_{ij}\}$ and $\{g_j\}$ can also depend on u_t , but we omit them to simplify notation.

Example 1. Variable length channel. Consider a channel h_t of length L_t that may change with time. The observed signal y_t is the baseband output of the channel corrupted by zero-mean Gaussian additive noise n_t of variance σ_t^2 :

$$y_t = h_t^T s_t + n_t$$

where s_t is the sequence of symbols. The objective is to estimate on-line the channel parameters as well as s_t . In this example, the discrete state r_t consists of the symbols s_t and L_t , and the process x_t includes h_t and σ_t^2 .

Example 2. Variable number of users in CDMA. In the context of CDMA where the number K_t of users evolves over time, the Markov chain state r_t consists of K_t as well as the symbols $s_{k,t}$ (note that $k = 1, \dots, K_t$) and the process x_t consists of the coefficients of the channel, and the noise parameter σ_t^2 .

1.1.2 Estimation Objectives

The aim of optimal filtering is to estimate *sequentially in time* the unknown “hidden” states $\{x_t, r_t\}$ and more precisely the series of posterior distributions $p(x_{0:t}, r_{1:t}|y_{1:t})$. Their marginals, and in particular the filtering densities $p(x_t, r_t|y_{1:t})$, are of interest in practice. A simple application of Bayes’ rule allows for an easy formulation of the recursion that updates the posterior distributions:

$$p(x_{0:t-1}, r_{1:t-1}|y_{1:t-1}) \frac{g_{r_t}(y_t|y_{1:t-1}, x_t) f_{r_{t-1}r_t}(x_t|x_{t-1})}{p(y_t|y_{1:t-1})/\pi_{r_{t-1}r_t}}.$$

*Statistics group, Department of Mathematics, University of Bristol, Bristol BS8 1TW, UK, e-mail: C.Andrieu@bristol.ac.uk

†IRCCyN, 1 rue de la Noë, BP 92101, 44321 Nantes cedex 3, France and Department of Engineering, University of Cambridge, UK, e-mail: Manuel.Davy@ircrcyn.ec-nantes.fr

‡Department of Electrical Engineering, University of Melbourne, Victoria 3010 Australia, e-mail: a.doucet@ee.mu.oz.au

There is no closed form solution to this recursion and for state estimates of the form $\int \sum_{r_{1:t}} \phi(x_{0:t}, r_{1:t}) p(dx_{0:t}, r_{1:t} | y_{1:t})$, which include the Minimum Mean Square Estimate (MMSE) of the state, $\mathbb{E}[x_t | y_{1:t}]$ and its covariance for example. For the sake of simplicity in notation, finite sums will be replaced further on by integrals whenever it is convenient.

1.2 Resolution and Organization of the Paper

We propose here to approximate $p(x_{0:t}, r_{1:t} | y_{1:t})$ using particle filtering methods. The key idea of particle filtering is to use an adaptive stochastic grid approximation of the posterior distribution of the state vector with $N \gg 1$ weighted particles (values of the grid) evolving randomly in time according to a simulation-based rule; that is the density is approximated by a weighted mixture of points,

$$p(dx_{0:t}, r_{1:t} | y_{1:t}) \approx \sum_{i=1}^N w_t^{(i)} \delta_{x_{0:t}}^{(i)}(dx_{0:t}) \mathbb{I}_{\{r_{1:t}^{(i)}\}}(r_{1:t}), \quad (3)$$

$$\sum_{i=1}^N w_t^{(i)} = 1, w_t^{(i)} \geq 0, \text{ so that for example}$$

$$\mathbb{E}[\phi(x_{0:t}, r_{1:t}) | y_{1:t}] \approx \sum_{i=1}^N w_t^{(i)} \phi(x_{0:t}^{(i)}, r_{1:t}^{(i)}).$$

We will further denote $\delta_{x,r}(dx, r) = \mathbb{I}_{\{r\}}(r) \delta_x(dx)$. The adaptive algorithm is designed such that the concentration of particles in a given region of the state space, say A , represents the probability of A under the posterior distribution, *i.e.* $\int_A p(x_{0:t}, r_{1:t} | y_{1:t}) dx_{0:t} dr_{1:t}$. Therefore computational efforts focus on different zones of the state space according to their importance, resulting in efficient algorithms. The particles evolve with time in a series of growing spaces, and can either give birth to offspring particles or die, depending on their ability to represent the different characteristics of interest of the posterior distributions, which are dictated by the observation process and the dynamics of the underlying system. The art of particle filtering consists mainly of the way the particles are updated and propagated through time. In particular it is extremely important to guide the exploration of these particles through the series of state spaces using any available information (*e.g.* the observations $y_{1:t}$) or salient feature of the underlying process (*e.g.* conditional linearity). It is thus no surprise if most efforts in the field have been devoted to these latter aspects. We propose here to develop a generic approach in order to design efficient particle filtering techniques adapted to the class of JMS described earlier. Our approach is an original combination of several methods that have been recently proposed in the literature, mainly the Auxiliary Particle Filter (APF) [4] and a suboptimal deterministic filtering methods, the Unscented Kalman Filter (UKF), which is a particular instance of the Unscented Transform (UT) [2]. This methodology has been successfully applied to, *e.g.*,

nonstationary signal detection and estimation [5], and can be applied as well to JMS in communications. In the next section, we briefly review the basic principles of particle filtering techniques and detail our generic algorithm adapted to JMS. Conclusions are drawn in Section 3.

2 Particle Filtering for JMS

2.1 Sequential Importance Sampling and Resampling

We briefly describe here how to apply the Sequential Importance Sampling Resampling (SISR) method in order to approximately sample from $p(x_{0:t}, r_{1:t} | y_{1:t})$; see [1] for further details. At time $t-1$, assume we have, say, N weighted particles $\{x_{0:t-1}^{(i)}, r_{1:t-1}^{(i)}\}$ associated to the weights $\{w_{t-1}^{(i)}\}$ that represent $p(dx_{0:t-1}, r_{1:t-1} | y_{1:t-1})$. We want to obtain N particles $\{x_{0:t}^{(i)}, r_{1:t}^{(i)}\}$ distributed according to $p(x_{0:t}, r_{1:t} | y_{1:t})$. At time t , we extend each particle $x_{0:t-1}^{(i)}, r_{1:t-1}^{(i)}$ by sampling $\tilde{x}_t^{(i)}, \tilde{r}_t^{(i)}$ according to a conditional distribution q_t to obtain N new particles $\{\tilde{x}_{0:t}^{(i)}, \tilde{r}_{1:t}^{(i)}\}$ where $\tilde{x}_{0:t}^{(i)} \triangleq (x_{0:t-1}^{(i)}, \tilde{x}_t^{(i)})$ and $\tilde{r}_{1:t}^{(i)} \triangleq (r_{1:t-1}^{(i)}, \tilde{r}_t^{(i)})$. To correct for the discrepancy between the distribution of each particle $\tilde{x}_{0:t}^{(i)}, \tilde{r}_{1:t}^{(i)}$ and $p(x_{0:t}, r_{1:t} | y_{1:t})$, we use importance sampling so that $p(x_{0:t}, r_{1:t} | y_{1:t})$ is approximated by the empirical distribution

$$\hat{p}_N(dx_{0:t}, r_{1:t} | y_{1:t}) = \sum_{i=1}^N w_t^{(i)} \delta_{\tilde{x}_{0:t}^{(i)}, \tilde{r}_{1:t}^{(i)}}(dx_{0:t}, r_{1:t}), \quad (4)$$

where the importance weights $w_t^{(i)}$ are equal to

$$w_{t-1}^{(i)} \frac{g_{\tilde{r}_t^{(i)}}(y_t | y_{1:t-1}, \tilde{x}_t^{(i)}) f_{\tilde{r}_{t-1}^{(i)}, \tilde{r}_t^{(i)}}(\tilde{x}_t^{(i)} | \tilde{x}_{t-1}^{(i)}) \pi_{\tilde{r}_{t-1}^{(i)}, \tilde{r}_t^{(i)}}}{q_t(\tilde{x}_{0:t}^{(i)}, \tilde{r}_{1:t}^{(i)} | \tilde{x}_{0:t-1}^{(i)}, \tilde{r}_{1:t-1}^{(i)}, y_{1:t})}. \quad (5)$$

The performance of the algorithm depends on the importance density q_t . The “optimal” importance density, that is the density minimizing the conditional variance of the weights conditional upon $y_{1:t-1}$, is proportional to [1]

$$g_{r_t}(y_t | y_{1:t-1}, x_t) f_{r_t}(x_t | x_{t-1}) \pi_{r_{t-1}, r_t},$$

and the associated importance weight are proportional to $w_{t-1} p(y_t | y_{1:t-1}, x_{t-1}, r_{t-1})$. This scenario is referred to as “full adaption” in [4]. Finally, one obtains N particles $\{x_{0:t}^{(i)}, r_{1:t}^{(i)}\}$ approximately distributed according to $p(x_{0:t}, r_{1:t} | y_{1:t})$ by resampling/selection from the weighted empirical distribution given in Eq. (4). There are several resampling procedures available in the literature. We adopt here the stratified sampling scheme [1]. This “optimal” importance sampling case deserves special attention. Indeed, the importance weights w_t is proportional to $p(y_t | x_{t-1}, r_{t-1}, y_{1:t-1})$, the predictive likelihood, do not actually depend on $\{x_t, r_t\}$. This means that resampling/selection can be performed before extending trajectories, thus selecting the most promising

trajectories before extension. However, in most practical cases, it is impossible to use the “optimal” importance sampling density as the predictive likelihoods of particles do not admit a closed form expression. However this scenario motivates an alternative particle filtering method known as APF [4], see Section 3, where one analytically approximates the predictive likelihoods, or its behaviour, whenever necessary.

2.2 Strategies for Efficient Particle Filtering

2.2.1 Auxiliary Particle Filter

The idea behind APF is, at time t , to extend existing trajectories $\{x_{0:t-1}^{(i)}, r_{1:t-1}^{(i)}\}$ that are the most promising, in the sense that their predictive likelihoods $p(y_t|y_{1:t-1}, x_{t-1}^{(i)}, r_{t-1}^{(i)})$ are large. However the analytical computation of these predictive likelihoods might prove to be intractable and approximation is needed. Recall that $p(y_t|y_{1:t-1}, x_{t-1}^{(i)}, r_{t-1}^{(i)})$ is equal to

$$\sum_{r_t \in S} \pi_{r_{t-1}^{(i)}, r_t} \int g_{r_t}(y_t|y_{1:t-1}, x_t) f_{r_{t-1}^{(i)}, r_t}(x_t|x_{t-1}^{(i)}) dx_t.$$

In [4], the authors propose simple approximations of the integrals with $g_{r_t}(y_t|y_{1:t-1}, \zeta(x_{t-1}^{(i)}, r_t))$, where $\zeta(x_{t-1}^{(i)}, r_t)$ is the mode or mean of $f_{r_{t-1}^{(i)}, r_t}(x_t|x_{t-1}^{(i)})$. In many applications, especially if $g_{r_t}(y_t|y_{1:t-1}, x_t)$ varies significantly over the significant regions of $f_{r_{t-1}^{(i)}, r_t}(x_t|x_{t-1}^{(i)})$, then the approximation of the predictive likelihood can be very poor and lead to performance far below that of the SIS algorithm. Indeed, one ends up biasing the exploration of the space towards uninteresting regions. It is thus fundamental to be able to approximate properly the predictive likelihood. An obvious solution would consist of using a second-stage Monte Carlo method for each particle. It is however too computationally intensive and introduces further Monte Carlo variation. We propose here to approximate this integral by $\psi_{r_t}(x_{t-1}^{(i)}, r_{t-1}^{(i)})$, a deterministic mapping/integration technique described in the next subsection (we omit the observations $y_{1:t}$ in $\psi_{r_t}(\cdot)$ to simplify notation). Then, an estimate of the desired quantity is

$$\widehat{p}(y_t|y_{1:t-1}, x_{t-1}^{(i)}, r_{t-1}^{(i)}) = \sum_{r_t \in S} \pi_{r_{t-1}^{(i)}, r_t} \psi_{r_t}(x_{t-1}^{(i)}, r_{t-1}^{(i)}).$$

The SIS extends each particle $x_{0:t-1}^{(i)}, r_{1:t-1}^{(i)}$ by sampling $\tilde{x}_t^{(i)}, \tilde{r}_t^{(i)}$ according to a conditional distribution q_t to obtain $\{x_{0:t-1}^{(i)}, \tilde{x}_t^{(i)}\}$ and $\{r_{1:t-1}^{(i)}, \tilde{r}_t^{(i)}\}$, then the associated weight $w_t^{(i)}$ can be rewritten as follows

$$w_{t-1}^{(i)} \widehat{p}(y_t|y_{1:t-1}, x_{t-1}^{(i)}, r_{t-1}^{(i)}) \times \frac{g_{\tilde{r}_t^{(i)}}(y_t|y_{1:t-1}, \tilde{x}_t^{(i)}) f_{\tilde{r}_{t-1}^{(i)}, \tilde{r}_t^{(i)}}(\tilde{x}_t^{(i)}|x_{t-1}^{(i)}) \pi_{\tilde{r}_{t-1}^{(i)}, \tilde{r}_t^{(i)}}}{\widehat{p}(y_t|y_{1:t-1}, x_{t-1}^{(i)}, r_{t-1}^{(i)}) q_t(\tilde{x}_t^{(i)}, \tilde{r}_t^{(i)}|x_{0:t-1}^{(i)}, r_{1:t-1}^{(i)}, y_{1:t})}.$$

The interest of this decomposition is that the first term, which is independent of $\tilde{r}_t^{(i)}, \tilde{x}_t^{(i)}$, mimicks the

weight of the “full adaption” scenario described earlier. It therefore suggests the possibility of resampling $\{x_{0:t-1}^{(i)}, r_{1:t-1}^{(i)}\}$ according to the weights $\lambda_t^{(i)} \propto w_{t-1}^{(i)} \widehat{p}(y_t|y_{1:t-1}, x_{t-1}^{(i)}, r_{t-1}^{(i)})$ in order to obtain N particles $\{\tilde{x}_{0:t-1}^{(i)}, \tilde{r}_{1:t-1}^{(i)}\}$ which are then approximately sampled from a distribution close to $p(x_{0:t-1}, r_{1:t-1}|y_{1:t})$. We then extend each particle by sampling $\tilde{x}_t^{(i)}, \tilde{r}_t^{(i)}$ according to a conditional distribution q_t . Contrary to the full adaption case, it is however necessary to reweight the particles by

$$w_t^{(i)} \propto \frac{g_{\tilde{r}_t^{(i)}}(y_t|y_{1:t-1}, \tilde{x}_t^{(i)}) f_{\tilde{r}_{t-1}^{(i)}, \tilde{r}_t^{(i)}}(\tilde{x}_t^{(i)}|\tilde{x}_{t-1}^{(i)}) \pi_{\tilde{r}_{t-1}^{(i)}, \tilde{r}_t^{(i)}}}{\widehat{p}(y_t|y_{1:t-1}, \tilde{x}_{t-1}^{(i)}, \tilde{r}_{t-1}^{(i)}) q_t(\tilde{x}_t^{(i)}, \tilde{r}_t^{(i)}|\tilde{x}_{0:t-1}^{(i)}, \tilde{r}_{1:t-1}^{(i)}, y_{1:t})},$$

as the samples $\{\tilde{x}_{0:t-1}^{(i)}, \tilde{r}_{1:t-1}^{(i)}\}$ are no longer distributed according to $p(x_{0:t-1}, r_{1:t-1}|y_{1:t})$ (even as $N \rightarrow \infty$). The problem of constructing an efficient deterministic mapping ψ_r for our problem is the subject of the following subsection.

2.2.2 The Unscented Transform

The Unscented Kalman Filter (UKF) is an alternative to the Extended Kalman Filter (EKF) which possesses many advantages. Both approaches are motivated by the fact that in most cases a single dynamic model, such as that of Subsection 1.1, can alternatively be represented in the following manner

$$x_t = \varphi(x_{t-1}, v_t), \quad y_t = \gamma(x_t, \varepsilon_t), \quad (6)$$

where $\{v_t\}$ and $\{\varepsilon_t\}$ are typically mutually independent zero-mean i.i.d. sequences, φ and γ are nonlinearities (similarly we will introduce φ_{ij} and γ_{ij} for JMS). Both the EKF and UKF rely on approximations of the system (6), but are of different nature. Nevertheless, for both scenarios the result of such approximations is that the series of predictive and filtering densities $\{p(x_t|y_{1:t-1})\}$ and $\{p(x_t|y_{1:t})\}$ are replaced with series of Gaussian distributions $\{\mathcal{N}(x_t; m_{t|t-1}, P_{t|t-1})\}$ and $\{\mathcal{N}(x_t; m_{t|t}, P_{t|t})\}$. Such approximations allow, in principle, for the application of the Kalman filter recursions in order to compute $m_{t|t-1}, P_{t|t-1}$ from $m_{t-1|t-1}, P_{t-1|t-1}$ (this makes use of the evolution equation) and $m_{t|t}, P_{t|t}$ from $m_{t|t-1}, P_{t|t-1}$ and the observation y_t (this makes use of the observation equation). The EKF relies on linearizations of the evolution and observation equations (6), followed with a direct application of the Kalman recursions on the first and second order moments. The solution adopted by the UKF is a second order truncation of the statistics of the posterior distributions at hand, followed by the Kalman recursions. More precisely, assume that a set of n points $\{\bar{x}_{t-1}^{(i)}\}$, the “sigma points” [2], possess the correct mean equal to $x_{t-1|t-1}$ and covariance $P_{t-1|t-1}$. Then the mean and sample autocovariance of the set $\{\varphi(\bar{x}_{t-1}^{(i)}, 0)\}$ should be a good approximation of $x_{t|t-1}$ and $P_{t|t-1}$ respectively. Similarly the mean and autocovariance

of $\{\gamma(\varphi(\bar{x}_{t-1}^{(i)}, 0), 0)\}$ should lead to reasonable approximations of $x_{t|t-1} \triangleq \mathbb{E}(\gamma(\varphi(x_{t-1}, v_t), \varepsilon_t) | y_{1:t-1})$ and $cov(y_t | y_{1:t-1})$, which are required for the Kalman filter recursion. Following the same principle the cross-covariance $cov(x_t, y_t | y_{1:t-1})$ can also be computed. Given these quantities it is then possible to take into account the new observation y_t , and calculate $x_{t|t}$ and $P_{t|t}$ with the Kalman filter. Given values for $x_{t|t}$ and $P_{t|t}$, various methods have been proposed in order to generate a new set $\{\bar{x}_t^{(i)}\}$ [2].

2.2.3 Algorithm

Based on the elements presented above, it is possible to propose the following generic particle filtering algorithm for JMS.

At time $t = 0$, **Step 0: Initialization**

- For $i = 1, \dots, N$ sample $x_0^{(i)} \sim p(x_0)$ and set the weights $w_0^{(i)} = 1/N$.

At time $t \geq 1$,

Step 1: Auxiliary variable resampling step

- For $i = 1, \dots, N$, compute $\lambda_t^{(i)}$ as¹

$$\lambda_t^{(i)} \propto w_{t-1}^{(i)} \sum_{r_t \in S} \pi_{r_{t-1} r_t} \psi_{r_t}(x_{t-1}^{(i)}, r_{t-1}^{(i)}), \quad \sum_{i=1}^N \lambda_t^{(i)} = 1, \quad (7)$$

where $\psi_{r_t}(x_{t-1}^{(i)}, r_{t-1}^{(i)})$ are computed using an unscented approximation:

$$\psi_{r_t}(x_{t-1}^{(i)}, r_{t-1}^{(i)}) = \frac{1}{n} \sum_{i=1}^n g_{r_t}(y_t | y_{1:t-1}, \varphi_{r_{t-1} r_t}(x_{t-1}^{(i)}))$$

The same sigma points $\{\bar{x}_{t-1}^{(i)}\}$ are used to compute $m_{t|t}(r_t), P_{t|t}(r_t)$ for all r_t in S .

- Multiply/Discard particles $\{x_{0:t-1}^{(i)}, r_{1:t-1}^{(i)}\}$ and the associated statistics $\{\psi_{r_t}(x_{t-1}^{(i)}, r_{t-1}^{(i)}), m_{t|t}^{(i)}(r_t), P_{t|t}^{(i)}(r_t); r_t \in S\}$ with respect to high/low importance weights $\lambda_t^{(i)}$ to obtain N particles $\{\bar{x}_{0:t-1}^{(i)}, \bar{r}_{1:t-1}^{(i)}\}$ and the associated statistics $\{\psi_{r_t}(\bar{x}_{t-1}^{(i)}, \bar{r}_{t-1}^{(i)}), \bar{m}_{t|t}^{(i)}(r_t), \bar{P}_{t|t}^{(i)}(r_t); r_t \in S\}$.

Step 2: Importance sampling step

- For $i = 1, \dots, N$, sample

$$\begin{aligned} \tilde{r}_t^{(i)} &\sim q(r_t | y_{1:t}, \bar{x}_{0:t-1}^{(i)}, \bar{r}_{0:t-1}^{(i)}) \\ &\propto \pi_{r_{t-1} r_t} \psi_{r_t}(\bar{x}_{t-1}^{(i)}, \bar{r}_{t-1}^{(i)}). \end{aligned}$$

- For $i = 1, \dots, N$, extend the trajectories with²

$$\bar{x}_t^{(i)} \sim \mathcal{N}(x; \bar{m}_{t|t}^{(i)}(\tilde{r}_t^{(i)}), \bar{P}_{t|t}^{(i)}(\tilde{r}_t^{(i)})).$$

¹For $t=1$, π_{r_{t-1}, r_t} in Eq. (7) should be replaced with the stationary distribution of the discrete Markov chain.

²Any other distribution, such as a heavy tailed distribution (e.g. a t -distribution) could also be used.

- Compute the importance weights as $w_t^{(i)} \propto$

$$\frac{g_{\tilde{r}_t^{(i)}}(y_t | \bar{x}_t^{(i)}) f_{\tilde{r}_{t-1} \tilde{r}_t^{(i)}}(\bar{x}_t^{(i)} | \bar{x}_{t-1}^{(i)})}{\psi_{\tilde{r}_t^{(i)}}(\bar{x}_{t-1}^{(i)}, \bar{r}_{t-1}^{(i)}) \mathcal{N}(\bar{x}_t^{(i)}; \bar{m}_{t|t}^{(i)}(\tilde{r}_t^{(i)}), \bar{P}_{t|t}^{(i)}(\tilde{r}_t^{(i)})}$$

- Rename the particles $\{\bar{x}_t^{(i)}, \tilde{r}_{t-1}^{(i)}\}$ into $\{x_t^{(i)}, r_{t-1}^{(i)}\}$
-

2.2.4 Discussion

The combination of the APF together with the UKF ensures good statistical properties of the proposed algorithm, as it tends to approximate the “full adaptation” scenario of [4]. However we would like here to stress on its computational efficiency, as some quantities are computed once, but used twice in **Step 1** and **Step 2**.

The efficiency of this algorithm for JMS is mostly a consequence of efficient sampling of the discrete state $\tilde{r}_t^{(i)}$ according to an estimate of the predicted likelihood of each state $r_t \in S$ (this is a consequence of the joint use of APF and UKF). This yields better efficiency than sampling from e.g. the prior distribution $\pi_{\tilde{r}_{t-1} r_t}$. In the stationary case (no jumps), the algorithm is still very efficient for two reasons: First, the APF step avoids sampling unlikely particles that would not contribute significantly to the approximation (and thus are wasted because the number of particles is fixed). Second, the UKF yields an importance distribution which consists of a locally Gaussian approximation of the actual target distribution.

3 Conclusion

In this paper we develop efficient particle filtering techniques especially tailored for Jump Markov systems. The efficiency of our approach has already been demonstrated in [5] for the estimation and detection of time-varying autoregressive models with unknown and variable number of poles.

References

- [1] Doucet A., de Freitas N. and Gordon N.J., Sequential Monte Carlo Methods in Practice, Springer-Verlag, May 2001.
- [2] Julier S.J. and Uhlmann J.K., “A general method for approximating nonlinear transformations of probability distributions”, Technical Report, Engineering Dept., University of Oxford.
- [3] Kim C.-J. and Nelson C. R., State-Space Models with Regime Switching, MIT Press, 1999.
- [4] Pitt M.K. and Shephard N., “Filtering via simulation: auxiliary particle filter,” J. Am. Statist. Ass., 94, 590-599, 1999.
- [5] Andrieu C., Davy M. and Doucet A., “Efficient Particle Filtering for Jump Markov Systems”, Technical report CUED/F-INFENG/TR.427, University of Cambridge, February 2002