

# Linköping University Electronic Press

## Book Chapter

### Mixture Kalman Filters and Beyond

Saikat Saha, Gustaf Hendeby and Fredrik Gustafsson

In *Current Trends in Bayesian Methodology with Applications*, Satyanshu K. Upadhyay, Umesh Singh, Dipak K. Dey, Appaia Loganathan (eds), 2015, pp. 537-562

ISBN: 978-1-4822-3511-1 (print), 978-1-4822-3512-8 (ebook).

DOI: <http://dx.doi.org/10.1201/b18502-27>.

Chapman and Hall/CRC Press, Taylor & Frances

<https://www.crcpress.com/product/isbn/9781482235111>

Available at: Linköping University Electronic Press  
<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-118279>

---

# Mixture Kalman Filters and Beyond

---

## Saikat Saha

*Dept. Electrical Engineering, Div. Automatic Control, Linköping University, SE-581 83 Linköping, Sweden, e-mail: saha@isy.liu.se*

## Gustaf Hendeby

*Dept. Electrical Engineering, Div. Automatic Control, Linköping University, SE-581 83 Linköping, Sweden, e-mail: hendeby@isy.liu.se; and  
Div. Sensor Informatics, Dept. Sensor & EW Systems, Swedish Defence Research Agency (FOI), SE-581 11 Linköping, Sweden*

## Fredrik Gustafsson

*Dept. Electrical Engineering, Div. Automatic Control, Linköping University, SE-581 83 Linköping, Sweden, e-mail: fredrik@isy.liu.se*

## CONTENTS

1	Introduction .....	2
	1.1 Problem background .....	2
	1.2 Organization .....	3
2	Particle filtering .....	4
3	Rao-Blackwellization .....	5
4	Models for mixture Kalman filter .....	7
	4.1 Hierarchical conditionally linear Gaussian model .....	7
	4.2 Mixed linear/nonlinear Gaussian state-space model ....	8
5	Beyond mixture Kalman filter in practice .....	11
	5.1 Hierarchical conditionally finite HMM .....	11
	5.2 Multiple switching nonlinear dynamic models .....	11
	5.2.1 A RBPF implementation .....	13
	5.3 Partially observed Gaussian state-space models .....	15
	5.4 Partially observed finite HMM .....	16
	5.5 Conditionally conjugate latent process model .....	16
	5.5.1 Rao-Blackwellized noise adaptive particle filtering .....	17
6	Terrain navigational problems .....	20
7	Approximate Rao-Blackwellized nonlinear filtering .....	21
8	Concluding remarks .....	22
	Acknowledgment .....	22

---

## 1 Introduction

The discrete time general state-space model is a flexible framework to deal with the nonlinear and/or non-Gaussian time series problems. However, the associated (Bayesian) inference problems are often intractable. Additionally, for many applications of interest, the inference solutions are required to be recursive over time. The *particle filter* (PF) is a popular class of Monte Carlo based numerical methods to deal with such problems in real time. However, PF is known to be computationally expensive and does not scale well with the problem dimensions. If a part of the state space is analytically tractable conditioned on the remaining part, the Monte Carlo based estimation is then confined to a space of lower dimension, resulting in an estimation method known as the *Rao-Blackwellized particle filter* (RBPF).

In this chapter, we present a brief review of Rao-Blackwellized particle filtering. Especially, we outline a set of popular conditional tractable structures admitting such Rao-Blackwellization in practice. For some special and/or relatively new cases, we also provide reasonably detailed descriptions. We confine our presentation mostly to the practitioners' point of view.

### 1.1 Problem background

Let us start with the following discrete time general state-space model relating the latent state  $x_k \in \mathbb{R}^{n_x}$  at time step  $k$ , to the observation  $y_k \in \mathbb{R}^{n_y}$  as

$$x_k = f(x_{k-1}, w_k), \quad (1a)$$

$$y_k = h(x_k, e_k), \quad (1b)$$

where  $f(x_{k-1}, w_{k-1})$  describes how the state propagates driven by the process noise  $w_{k-1}$ , and  $h(x_k, e_t)$  describes how the measurements relates to the state and how the measurement is affected by noise,  $e_k$ . Note that the latent state  $x_k$  in (1) follows a Markovian model and the observation  $y_k$  is conditionally independent given  $x_k$ . The model (1) can be equivalently characterized in terms of known state transition density  $p(x_k|x_{k-1})$  and observation likelihood density  $p(y_k|x_k)$ . For this model, given the density for initial state (*i.e.*,  $p(x_0)$ ) and a stream of observations  $y_{0:k} \triangleq \{y_0, y_1, \dots, y_k\}$  up to time  $k$ , carrying the information about the state, one typical objective is to optimally estimate the sequence of posterior densities  $p(x_{0:k}|y_{1:k})$  and their marginals  $p(x_k|y_{1:k})$ . The latter problem is generally referred to as the *filtering problem*. The filtering problem appears naturally in many applications of interest in applied science, econometrics, and engineering. Moreover, for certain applications (*e.g.*, target tracking), the above densities need to be propagated recursively in real time. The exact posterior densities  $p(x_k|y_{1:k})$  (also known as filter densities) can be obtained analytically and recursively over time, only for a few special cases,

including the *linear Gaussian state-space model* (LGSSM), where the exact posterior is given by the celebrated *Kalman filter* (KF) [4, 48].

However, contrary to Kalman's linear-Gaussian assumptions (*i.e.*,  $f(\cdot)$  and  $h(\cdot)$  are linear,  $w_k$ ,  $v_k$ , and  $x_0$  are Gaussian), many real world systems are nonlinear-non-Gaussian in their nature. Optimal estimates for such models can generally not be obtained in closed form and as a result, many approximative methods have been proposed. Most of these approaches are centered around approximating the state-space models with a linear-Gaussian form (*e.g.*, the extended Kalman filter, Gaussian sum filters, and nonlinear Kalman filters based on different numerical quadrature rules). Since the early 1990's, a class of sequential Monte Carlo simulation based filters, popularly known as *particle filters* have appeared to dominate this field due to their flexibility in adapting nonlinearity and/or non-Gaussianity without any *ad hoc* model assumptions. Particle filtering methods were introduced in their modern forms by the seminal contributions of Gordon et al. [20]. In this framework, the posterior distribution is approximated by an empirical distribution formed by the (weighted) random samples, called particles, and this empirical distribution can be propagated over time recursively (see *e.g.*, [8, 15, 18, 21] for details).

Although the PF has been around for a while, it is computationally demanding and notably, it has severe limitations when scaling to higher dimensions [14, 33, 44]. As many important practical problems are high-dimensional in their nature, this poses a limitation on the general applicability of the PF. However, for many practical applications, severe nonlinearity and/or non-Gaussianity often affects only a small subset of the states. The rest of the states are often mildly nonlinear and/or non-Gaussian. Then one practical solution to the high dimensional problems can be given by first employing a particle filter to the highly nonlinear/non-Gaussian part of the state space, and conditioned on this, apply an (extended) Kalman filter (or variant thereof) to the remaining state space. This technical enabler is often known as the *mixture Kalman filter* (MKF) [10].

The mixture Kalman filter requires a special model structure, *i.e.*, a part of the state space is required to be conditionally linear-Gaussian. Then one can conditionally apply KF, which is analytically tractable. The immediate and foreseeable advantage is that solving part of the problem analytically leaves only a part of the state vector to be solved approximately using particle techniques, hence requiring fewer particles. However, this conditional tractability is not limited to only LGSSM (as in MKF) and it has greater connotation in terms of efficiency improvement. When it is possible to exploit such conditional tractability, the Monte Carlo based estimation is confined to lower dimensional space. Consequently, the estimate obtained is often better and never worse than the estimate provided by the PF targeting the full state-space model. This efficiency arises as an implication of the so called *Rao-Blackwellization*. For this reason, the resulting filtering problem is also known as *Rao-Blackwellized particle filtering* [19].

## 1.2 Organization

The rest of the chapter is organized as follows. In order to make the chapter self containing, we start with brief but necessary backgrounds on PF and Rao-Blackwellization, in Section 2 and Section 3 respectively. This is followed by a detailed description on the models for mixture Kalman filter in Section 4. As pointed out earlier, RBPF is not only confined to the MKF and therefore we outline other popular RBPF approaches in Section 5. Next, we present a real life application in Section 6, where such model frameworks are put in place. For many practical applications, where conditional tractability is missing, the marginalization step can be carried out approximately using Monte Carlo integrations. This idea has led to the emergence of ‘approximated RBPF’, which is briefly discussed in Section 7. Finally, this is followed by the concluding remarks in Section 8.

---

## 2 Particle filtering

In PF, the posterior distribution associated with the density  $p(x_{0:k}|y_{1:k})$  is approximated by an empirical distribution represented by a set of  $N(\gg 1)$  weighted particles (samples) as

$$\hat{P}_N(dx_{0:k}|y_{1:k}) = \sum_{i=1}^N \tilde{w}_k^{(i)} \delta_{x_{0:k}^{(i)}}(dx_{0:k}), \quad (2)$$

where  $\delta_{x_{0:k}^{(i)}}(A)$  is a Dirac measure for a given  $x_{0:k}^{(i)}$  and a measurable set  $A$ . The  $\tilde{w}_k^{(i)}$  is the weight associated with each particle  $x_{0:k}^{(i)}$ , such that  $\tilde{w}_k^{(i)} \geq 0$  and  $\sum_{i=1}^N \tilde{w}_k^{(i)} = 1$ . Given this representation, one can approximate the marginal distribution associated with  $p(x_k|y_{1:k})$  as

$$\hat{P}_N(dx_k|y_{1:k}) = \sum_{i=1}^N \tilde{w}_k^{(i)} \delta_{x_k^{(i)}}(dx_k), \quad (3)$$

and expectations of the form

$$I(g_k) = \int g_k(x_{0:k}) p(x_{0:k}|y_{1:k}) dx_{0:k} \quad (4)$$

as

$$\hat{I}_N(g_k) = \int g_k(x_{0:k}) \hat{P}_N(dx_{0:k}|y_{1:k}) \approx \sum_{i=1}^N \tilde{w}_k^{(i)} g_k(x_{0:k}^{(i)}). \quad (5)$$

Although the distribution  $\hat{P}_N(dx_{0:k}|y_{1:k})$  does not admit a well defined density with respect to the Lebesgue measure, the density  $p(x_{0:k}|y_{1:k})$  is empirically

represented as  $\widehat{p}_N(x_{0:k}|y_{1:k})$  given by

$$\widehat{p}_N(x_{0:k}|y_{1:k}) = \sum_{i=1}^N \widetilde{w}_k^{(i)} \delta(x_{0:k} - x_{0:k}^{(i)}), \quad (6)$$

where  $\delta(\cdot)$  is the Dirac-delta function. The notation used in (6) is not mathematically rigorous; however, it is intuitively easier to follow than the more rigorous measure theoretic notations. This is especially useful when we are not concerned with theoretical convergence studies.

Now suppose at time  $k - 1$ , we have a weighted particle approximation of the posterior  $p(x_{0:k-1}|y_{1:k-1})$  as  $\widehat{P}_N(dx_{0:k-1}|y_{1:k-1}) = \sum_{i=1}^N \widetilde{w}_{k-1}^{(i)} \delta_{x_{0:k-1}^{(i)}}(dx_{0:k-1})$ . With the arrival of a new measurement  $y_k$ , we want to approximate  $p(x_{0:k}|y_{1:k})$ . This is achieved by propagating the particles in time using importance sampling and resampling steps. First, the particles are propagated to time  $k$  by sampling a new state  $x_k^{(i)}$  from a proposal kernel  $\pi(x_k|x_{0:k-1}^{(i)}, y_{1:k})$  and then setting  $x_{0:k}^{(i)} \triangleq (x_{0:k-1}^{(i)}, x_k^{(i)})$ . Since we have the posterior recursion as

$$p(x_{0:k}|y_{1:k}) \propto p(y_k|x_{0:k}, y_{1:k-1}) p(x_k|x_{0:k-1}, y_{1:k-1}) p(x_{0:k-1}|y_{1:k-1}) \quad (7)$$

and using (1), the corresponding (normalized) weights of the particles are updated as

$$w_k^{(i)} \propto \widetilde{w}_{k-1}^{(i)} \frac{p(y_k|x_k^{(i)}) p(x_k^{(i)}|x_{k-1}^{(i)})}{\pi(x_k^{(i)}|x_{0:k-1}^{(i)}, y_{1:k})}; \quad \widetilde{w}_k^{(i)} = \frac{w_k^{(i)}}{\sum_{j=1}^N w_k^{(j)}}. \quad (8)$$

Now, to avoid carrying trajectories with small weights, and to focus on the ones with large weights, the particles are resampled (with replacement), whenever necessary (see *e.g.*, [17] for details). The resampled particles representing the same posterior  $p(x_{0:k}|y_{1:k})$  are equally weighted. The whole procedure is repeated sequentially over time. For a more general introduction to PF, refer to *e.g.*, [18].

---

### 3 Rao-Blackwellization

The foundation of the Rao-Blackwellization technique used in many estimation and filtering applications dates back to the 1940's. In the 1940's Rao [35] and Blackwell [7] showed that an estimator can be improved by making use of conditional expectations, if such structure exist in the problem formulation. They further showed how an estimator based on this knowledge can be constructed as an expected value of a conditioned estimator. This practice has over time become known as Rao-Blackwellization.

The Rao-Blackwell theorem is often used to show how the knowledge of a conditional expected value can improve the variance of an estimate, but this is only a special case of a more general theorem which shows that any convex cost-function is improved using this technique. For completeness, the Rao-Blackwell theorem is given in Theorem 1, [28, p. 51].

**Theorem 1** (The Rao-Blackwell theorem [28]). *Let  $x$  be a random observable with the distribution  $P_\theta \in \mathcal{P} = \{P_{\theta'} \in \Omega\}$ , and let  $t$  be sufficient statistics for  $\mathcal{P}$ <sup>1</sup>. Let  $\hat{\theta}$  be an estimator of an estimand  $g(\theta)$ , and let the loss function  $L(\theta, d)$  be a strictly convex function in  $d$ . Then if  $\hat{\theta}$  has finite expectation and risk,*

$$R(\theta, \hat{\theta}) = \mathbb{E}(L(\theta, \hat{\theta}(x))) \quad \text{and if} \quad \hat{\theta}^{\text{RB}}(t) = \mathbb{E}(\hat{\theta}(x)|t)$$

*the risk of the estimator  $\hat{\theta}^{\text{RB}}(t)$  satisfies*

$$R(\theta, \hat{\theta}^{\text{RB}}) \prec R(\theta, \hat{\theta}),$$

*unless  $\hat{\theta}(x) = \hat{\theta}^{\text{RB}}(t)$  with probability 1, where  $A \prec B$  denotes that  $B - A$  is a positive definite matrix.*

*Proof.* See [28]. □

It is worth pointing out once more that the Rao-Blackwell theorem does not only predict a reduced risk, it also shows how to construct the estimator to obtain the gain as an expected value. This makes the theorem very useful, as will be exemplified a number of times in the remainder of this chapter.

An important and often used special case of Theorem 1 describes the variance properties of a Rao-Blackwell estimator.

**Corollary 1.** *Let  $\hat{\theta}$  be an unbiased estimator that fulfills the requirements in Theorem 1 and  $\hat{\theta}^{\text{RB}}$  the Rao-Blackwell estimator, then*

$$\text{cov}(\hat{\theta}^{\text{RB}}) \prec \text{cov}(\hat{\theta})$$

*unless  $\hat{\theta}(x) = \hat{\theta}^{\text{RB}}(t)$  with probability 1.*

*Proof.* To prove the corollary, assume the cost function

$$L(\theta, \hat{\theta}) = (\theta - \hat{\theta})(\theta - \hat{\theta})^T,$$

which is convex in both  $\theta$  and  $\hat{\theta}$ , and apply Theorem 1. Finally note that  $\text{cov}(\theta) = R(\theta, \hat{\theta}) = \mathbb{E}(L(\theta, \hat{\theta}))$ , which concludes the proof. □

---

<sup>1</sup>The statistic  $t$  is sufficient for  $\theta$  if the conditional distribution of  $x$  given  $t$  does not depend on  $\theta$ .

The covariance gain using Rao-Blackwellization is given by the separation of covariance, [24], which states

$$\text{cov}(x) = \text{cov}(\mathbb{E}(x|y)) + \mathbb{E}(\text{cov}(x|y)),$$

where  $\text{cov}(x|y)$  denotes the covariance matrix of  $x$  conditioned on  $y$ . Assuming that  $x$  is the state estimate, then  $\mathbb{E}(x|y)$  is the improved estimate in the Rao-Blackwell theorem. Hence, compared to the Rao-Blackwellized estimate, the regular estimate has an additional covariance term,  $\mathbb{E}(\text{cov}(x|y))$ , that by definition is positive semi-definite. The covariance gain obtained with a Rao-Blackwellized estimator comes from eliminating this term in the covariance expression.

---

## 4 Models for mixture Kalman filter

The mixture Kalman filter was mentioned earlier as an enabling technique for PF to reasonably high dimensional problems. Since MKF uses conditional KF, it essentially requires a *conditional linear Gaussian state-space model* (CLGSSM). This is an important class of models with many practical applications which has received extensive research attentions in the past [3, 21, 22, 29]. In the sequel, we outline two different model structures conforming to this class: the hierarchical conditional linear Gaussian state-space model and the mixed linear/nonlinear Gaussian state-space model.

### 4.1 Hierarchical conditionally linear Gaussian model

The *hierarchical conditionally linear Gaussian model* (HCLGM) is, as the name suggests, a hierarchical model structure in two layers where a latent process  $c_k$  propagates according to a Markov kernel, independent of the second part of the latent variable  $x_k$ . Conditioned on  $c_k$ ,  $(x_k, y_k)$  follows a linear Gaussian state-space model, parametrized in  $c_k$ . Mathematically the HCLGM can then be described by

$$c_{k+1} \sim p(c_{k+1}|c_k) \tag{9a}$$

$$x_{k+1} = f(c_k) + F(c_k)x_k + G(c_k)w_k \tag{9b}$$

$$y_k = h(c_k) + H(c_k)x_k + H^e(c_k)e_k, \tag{9c}$$

where  $p(c_{k+1}|c_k)$  indicates that  $c_k$  is propagated using a Markov kernel;  $w_k$  and  $e_k$  are Gaussian process and measurement noises, respectively. A typical example of an HCLGM is a linear system with a scaled mixture of Gaussian noises (*e.g.*, Student's t noise distribution [30, 40]); here conditioned on  $c_{0:k}$ , the noise can be suitably represented as Gaussian and consequently KF can be applied on the conditionally linear-Gaussian part. A graphical illustration



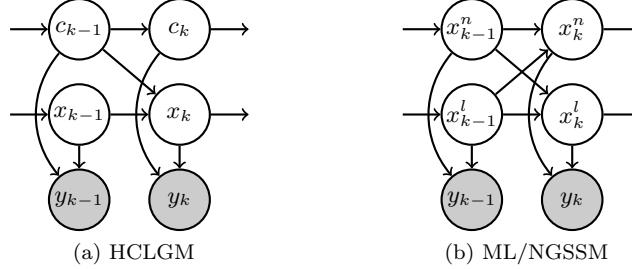


FIGURE 1: Graphical representation of dependencies in *hierarchically linear Gaussian model* (HCLGM) (9) and the *mixed linear/nonlinear Gaussian state-space model* (ML/NGSSM) (11).

of the hierarchical model structure is given in Figure 1(a). This class of models is also known as the *conditional dynamic linear model* (CDLM)<sup>2</sup> [10]. For the inference task, we can decompose the joint density  $p(c_{0:k}, x_k | y_{1:k})$ ,

$$p(c_{0:k}, x_k | y_{1:k}) = p(c_{0:k} | y_{1:k}) p(x_k | c_{0:k}, y_{1:k}), \quad (10)$$

where we target the propagation of  $p(c_{0:k} | y_{1:k})$  using a PF and for each sequence  $c_{0:k}$ , the propagation of  $p(x_k | c_{0:k}, y_{1:k})$  using a KF.

## 4.2 Mixed linear/nonlinear Gaussian state-space model

Another common model following a CGLSSM structure is the *mixed linear/nonlinear Gaussian state-space model* (ML/NGSSM) [43]. The ML/NGSSM is a state-space model where it is possible to divide the state vector into two parts,  $x_k = \begin{pmatrix} x_k^n \\ x_k^l \end{pmatrix}$ , such that conditioned on  $x_k^n$ ,  $x_k^l$  assumes a linear-Gaussian state-space model,

$$x_{k+1}^n = f^n(x_k^n) + F^n(x_k^n)x_k^l + w_k^n \quad (11a)$$

$$x_{k+1}^l = f^l(x_k^n) + F^l(x_k^n)x_k^l + w_k^l \quad (11b)$$

$$y_k = h(x_k^n) + H(x_k^n)x_k^l + e_k, \quad (11c)$$

where  $w_k^n$ ,  $w_k^l$ , and  $e_k$  are mutually independent Gaussian noise processes with variance  $Q_k^n$ ,  $Q_k^l$ , and  $R_k$ , respectively. Furthermore,  $x_0^l$  is assumed to follow a Gaussian distribution. The distribution of  $x_0^n$  can be arbitrary, but is assumed known. In [10], such a model is denoted a *partial conditional dynamic*

<sup>2</sup>Although we restrict  $c_k$  to be a continuous variable for CDLM, in [10] CDLM is defined in a broader sense such that  $c_k$  can be either a continuous or discrete variable. The above model corresponding to a discrete  $c_k$  is popularly known as switching linear Gaussian system (or jump Markov linear system), which, according to our notations, is a special case of model (5.2), discussed later.

*linear model* (partial CDLM). However, there the authors did not consider the influence of  $x_k^l$  in the evolution of  $x_k^n$ . This influence was first considered in [31] and was further popularized in [43]. This framework is extended to the dependent noise processes in [37]. The interconnection between the two parts of the state space,  $x_k^n$  and  $x_k^l$  is best illustrated in a graphical model as in Figure 1(b).

The nonlinear filtering problem can be decomposed into two sub-problems,

$$p(x_k^l, x_{0:k}^n | y_{1:k}) = p(x_{0:k}^n | y_{1:k}) p(x_k^l | x_{0:k}^n, y_{1:k}), \quad (12)$$

where the propagation of  $p(x_{0:k}^n | y_{1:k})$  is targeted using a PF and conditioned on each sequence  $x_{0:k}^n$ ,  $p(x_k^l | x_{0:k}^n, y_{1:k})$  can be propagated analytically using KF. The algorithm descriptions for this model are rather technical. The mixture Kalman filter here, when a bit simplified, works the following way: The nonlinear part of the problem is solved using a particle filter. Attached to each particle is a Kalman filter, which estimates the (conditionally) linear part of the state. To compensate for the interactions between the linear and nonlinear part of the state vector, a virtual measurement update must be applied to the Kalman filter each time the particle filter is propagated forward in time.

Later [25, 26], reformulated this RBPF scheme in terms of a stochastic bank of Kalman filters, in that way highlighting the structural components of the solution. In essence, this RBPF scheme becomes a Kalman filter bank with stochastic pruning and weight computations provided by the particle filter. The algorithmic summary of this approach is given in Algorithm 1.

The mixed linear/nonlinear Gaussian state-space model as in (11) is quite common, especially in tracking and navigation applications, where the unknown state vector typically consists of position, velocity, and sometimes acceleration. When fitted into such a framework, the position typically ends up in the nonlinear part of the state space, whereas the velocity and acceleration in the linear part. These applications were considered in [23, 31] under the name *marginalized particle filter* (MPF).<sup>3</sup>

This model structure is also extensively used in robotics to address the *simultaneous localization and mapping* (SLAM) problem. Here one attempts to localize oneself in an unknown environment and at the same time, build a map of the surrounding environment. In essence, this yields a state-space model with a state comprised of the kinematic state of one self, and the map, which grows over time. Fortunately, conditioned on the state, the map fits in

---

<sup>3</sup>The reason behind that name is that the linear (or, more generally, analytical) part of the state vector is ‘marginalized’ (or, integrated out) in closed form to obtain the posterior distribution of the nonlinear (or, non-analytical) part.

**Algorithm 1** Rao-Blackwellized PF [26]

For the system (11).

**Initialization:** For  $i = 1, \dots, N$ , let

$$x_{0|-1}^{p(i)} \sim \pi_0^n, \quad x_{0|-1}^{k(i)} = x_0^k, \quad P_{0|-1}^{k(i)} = \Pi_0^l, \quad \omega_{0|-1} = \frac{1}{N}.$$

**Iterations:** For  $k = 0, \dots$  do

- Measurement update, for each particle  $i = 1, \dots, N$  do

$$\begin{aligned} \omega_{k|k}^{(i)} &\propto \mathcal{N}(y_k; \hat{y}_k^{(i)}, S_k^{(i)}) \cdot \omega_{k|k-1}^{(i)} \\ x_{k|k}^{(i)} &= x_{k|k-1}^{(i)} + K_k^{(i)}(y_k - \hat{y}_k^{(i)}) \\ P_{k|k}^{(i)} &= P_{k|k-1}^{(i)} - K_k^{(i)} S_k^{(i)} K_k^{(i)T}, \end{aligned}$$

where

$$\begin{aligned} \hat{y}_k^{(i)} &= h(x_{k|k-1}^{n(i)}) + H(x_{k|k-1}^{n(i)})x_{k|k-1}^{l(i)} \\ S_k^{(i)} &= H(x_{k|k-1}^{n(i)})P_{k|k-1}^{(i)}H^T(x_{k|k-1}^{n(i)}) + R_k \\ K_k^{(i)} &= P_{k|k-1}^{l(i)}H^T(x_{k|k-1}^{l(i)})(S_k^{(i)})^{-1}. \end{aligned}$$

- If necessary, resample the particles using standard techniques. When selecting a particle  $x_{k|k}^{n(i)}$  select the matching linear states  $x_{k|k}^{l(i)}$  and  $P_{k|k}^{l(i)}$ , and set  $\omega_{k|k}^{(i)}$  appropriately.
- Time update, for each particle  $i = 1, \dots, N$  do

$$\begin{aligned} \bar{x}_{k+1|k}^{n(i)} &= f^n(x_{k|k}^{n(i)}) + F^n(x_{k|k}^{n(i)})x_{k|k}^{l(i)} \\ \bar{x}_{k+1|k}^{l(i)} &= f^l(x_{k|k}^{n(i)}) + F^l(x_{k|k}^{n(i)})x_{k|k}^{l(i)} \\ \bar{P}_{k+1|k}^{n(i)} &= F^n(x_{k|k}^{n(i)})P_{k+1|k}^{l(i)}(F^n(x_{k|k}^{n(i)}))^T + Q_k^n \\ \bar{P}_{k+1|k}^{nl(i)} &= F^n(x_{k|k}^{n(i)})P_{k+1|k}^{l(i)}(F^l(x_{k|k}^{n(i)}))^T \\ \bar{P}_{k+1|k}^{l(i)} &= F^l(x_{k|k}^{n(i)})P_{k+1|k}^{l(i)}(F^l(x_{k|k}^{n(i)}))^T + Q_k^l. \end{aligned}$$

With  $\xi_{k+1}^{(i)} \sim \mathcal{N}(0, \bar{P}_{k+1|k}^{n(i)})$  do

$$\begin{aligned} x_{k+1|k}^{n(i)} &= \bar{x}_{k+1|k}^{n(i)} + \xi_{k+1}^{(i)} \\ x_{k+1|k}^{l(i)} &= \bar{x}_{k+1|k}^{l(i)} + (\bar{P}_{k+1|k}^{nl(i)})^T (\bar{P}_{k+1|k}^{n(i)})^{-1} \xi_{k+1}^{(i)} \\ P_{k+1|k}^{l(i)} &= \bar{P}_{k+1|k}^{l(i)} - (\bar{P}_{k+1|k}^{nl(i)})^T (\bar{P}_{k+1|k}^{n(i)})^{-1} P_{k+1|k}^{nl(i)}. \end{aligned}$$

a linear-Gaussian model which allows for Rao-Blackwellization. This fact is utilized in the FastSLAM algorithm which is an MKF in disguise [29].

---

## 5 Beyond mixture Kalman filter in practice

Note that the key to efficient Rao-Blackwellization is a (conditionally) tractable substructure, which is not necessarily limited to the conditionally *linear-Gaussian state-space model* (CLGSS). In this section, we present some other existing Rao-Blackwellization approaches that are popular in practice.

### 5.1 Hierarchical conditionally finite HMM

Similar to the LGSS model, when the latent state evolves according to a finite state Markov chain, the posterior of the state given the observations, can be obtained in closed form using the *hidden Markov model* (HMM) forward algorithm [34]. This model structure is known as the finite state-space HMM. Now if part of the state space follows such finite state-space HMM, it is then possible to exploit an RBPF formulation.

One such RBPF framework appears when the state space follows a finite state HMM, whose parameters are modulated by another Markovian process [19]. This model structure is denoted the *hierarchical conditionally finite HMM* (HCfHMM) and can be described as

$$p(c_k|c_{k-1}), \quad p(r_k|c_k, r_{k-1}), \quad p(y_k|c_k, r_k), \quad (13)$$

where  $r_k$  is a finite state-space Markov chain, whose parameter at time  $k$  depends on a Markov process  $c_k$ . Both  $r_k$  and  $c_k$  are latent and are related to the observation  $y_k$ . Note that  $c_k$  is at the top of the hierarchy and the evolution of  $c_k$  is independent of  $r_k$ . The graphical representation is shown in Figure 2(a). Now conditioned on  $c_k$ ,  $r_k$  follows a finite state-space Markov chain of known parameters. Therefore one can implement a Rao-Blackwellized particle filter, where the PF targets only  $p(c_{0:k}|y_{0:k})$  and where  $p(r_k|c_{0:k}, y_{0:k})$  is analytically tractable.

### 5.2 Multiple switching nonlinear dynamic models

In many practical applications of engineering and econometrics, one often deals with nonlinear dynamic systems involving both a continuous value target state and a discrete value regime variable. Such descriptions imply that the system can switch between different nonlinear dynamic regimes, where the parameters of each regime is modulated by the corresponding regime variable. The regime variable also evolves dynamically according to a finite state

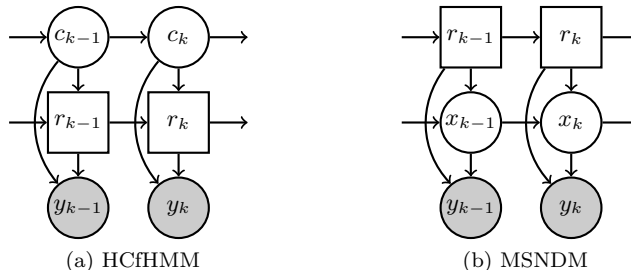


FIGURE 2: Graphical representation of the *hierarchical conditionally finite HMM* (HCfHMM) and *multiple switching nonlinear dynamics model* (MSNDM).

Markov chain. Both the target state and regime variable are latent and are related to the noisy observations. This type of model is referred to as (*Markov*) *regime switching nonlinear dynamic models* or *hybrid nonlinear state-space models*.

The probabilistic description of the above model can be given by

$$\Pi(r_k|r_{k-1}), \quad p_{\theta_{r_k}}(x_k|x_{k-1}, r_k), \quad p_{\theta_{r_k}}(y_k|x_k, r_k), \quad (14)$$

where  $r_k \in S \triangleq \{1, 2, \dots, s\}$ , is a (discrete) regime indicator variable with finite number of regimes (*i.e.*, categorical variable),  $x_k \in \mathbb{R}^{n_x}$  is the (continuous) state variable. As the system can switch between different dynamic regimes, for a given regime variable  $l \in S$ , the corresponding dynamic regime can be characterized by a set of parameters  $\theta_l$ . Both  $x_k$  and  $r_k$  are latent variables, which are related to the measurement  $y_k \in \mathbb{R}^{n_y}$ . The evolution of the regime variable  $r_k$  is commonly modeled by a homogeneous (time invariant) first order Markov chain with *transition probability matrix* (TPM)  $\Pi = [\pi_{ij}]_{ij}$ ,

$$\pi_{ij} \triangleq \mathbb{P}(r_k = j | r_{k-1} = i) \quad (i, j \in S), \quad (15a)$$

$$\pi_{ij} \geq 0; \quad \sum_{j=1}^s \pi_{ij} = 1. \quad (15b)$$

This model is represented graphically in Figure 2(b). Note that the above model can easily be generalized to the cases involving time dependent dynamics and exogenous inputs. For the model, given the densities for the initial state  $\{r_0, x_0\}$  and the measurements up to a time  $k$ , our interest lies in estimating  $\mathbb{P}(r_k|y_{1:k})$  and  $p(x_k|y_{1:k})$  recursively. However, the above posteriors are computationally intractable. In connection to this, we note from (14) that conditioned on the sequence  $x_{0:k}$ ,  $r_k$  follows a finite state hidden Markov model. As a result,  $\mathbb{P}(r_k|x_{0:k}, y_{1:k})$  is analytically tractable [36]. Using this analytical substructure, it is possible to implement an efficient RBPF scheme<sup>4</sup>. We provide below a brief description of RBPF implementation for the same.

<sup>4</sup>It should be noted that the switching linear Gaussian system is a special case of this

### 5.2.1 A RBPF implementation

The initial densities for the state and the regime variables are given by  $p(x_0)$  and  $\mathbb{P}(r_0) \triangleq \mathbb{P}(r_0|x_0)$ , respectively, which can be arbitrary but is assumed to be known. We further assume favorable mixing conditions as in [13].

Suppose that we are at time  $(k-1)$ . We consider the extended target density  $p(r_{k-1}, x_{0:k-1}|y_{1:k-1})$  which can be decomposed as

$$p(r_{k-1}, x_{0:k-1}|y_{1:k-1}) = p(r_{k-1}|x_{0:k-1}, y_{1:k-1})p(x_{0:k-1}|y_{1:k-1}) \quad (16)$$

The posterior propagation of the latent state  $x_{k-1}$  can be targeted through a PF, where  $p(x_{0:k-1}|y_{1:k-1})$  is represented by a set of  $N$  weighted random particles as

$$p(x_{0:k-1}|y_{1:k-1}) \approx \sum_{i=1}^N w_{k-1}^{(i)} \delta(x_{0:k-1} - x_{0:k-1}^{(i)}). \quad (17)$$

Now conditioned on  $\{x_{0:k-1}, y_{1:k-1}\}$ , the regime variable  $r_{k-1}$  follows a finite state-space HMM. Consequently,  $p(r_{k-1}|x_{0:k-1}, y_{1:k-1})$  is analytically tractable<sup>5</sup>, which is represented as

$$q_{k-1|k-1}^{(i)}(l) \triangleq \mathbb{P}(r_{k-1} = l|x_{0:k-1}^{(i)}, y_{1:k-1}), \quad (18)$$

for  $l \in S$  and  $i = 1, \dots, N$ . Now using (17) and (18), the extended target density in (16) can be represented as

$$\left[ x_{0:k-1}^{(i)}, w_{k-1}^{(i)}, \{q_{k-1|k-1}^{(i)}(l)\}_{l=1}^s \right]_{i=1}^N. \quad (19)$$

Now having observed  $y_k$ , we want to propagate the extended target density in (16) to time  $(k)$ . This can be achieved in the following steps (a)–(d):

(a) **Prediction step for conditional HMM filter:** this is obtained as

$$q_{k|k-1}^{(i)}(l) \triangleq \mathbb{P}(r_k = l|x_{0:k-1}^{(i)}, y_{1:k-1}) = \sum_{j=1}^s \pi_{jl} q_{k-1|k-1}^{(i)}(j), \quad (l, j) \in S. \quad (20)$$

(b) **Prediction step for particle filter:** at this stage, generate  $N$  new samples  $x_k^{(i)}$  from an appropriate proposal kernel as  $x_k^{(i)} \sim \pi(x_k|x_{k-1}^{(i)}, y_k)$ . Then set  $x_{0:k}^{(i)} = \{x_{0:k-1}^{(i)}, x_k^{(i)}\}$ , for  $i = 1, \dots, N$ , representing the particle trajectories up to time  $k$ .

---

model structure, where the inference problem is popularly addressed using an alternative RBPF scheme: the propagation of  $\mathbb{P}(r_{0:k}|y_{1:k})$  is targeted by a PF and conditioned on  $r_{0:k}$ ,  $p(x_k|r_{0:k}, y_{1:k})$  is obtained analytically using KF methods

<sup>5</sup>The ‘favorable’ mixing property ensures that  $p(x_{0:k-1}|y_{1:k-1})$  can be well approximated by  $p(x_{k-L:k-1}|y_{1:k-1})$ , for some lag  $L$ . Consequently,  $p(r_{k-1}|x_{0:k-1}, y_{1:k-1}) \approx p(r_{k-1}|x_{k-L:k-1}, y_{1:k-1})$ .

(c) **Update step for conditional HMM filter:** noting that

$$\mathbb{P}(r_k = l | x_{0:k}, y_{1:k}) \propto p(y_k, x_k | r_k = l, x_{0:k-1}, y_{1:k-1}) \times \mathbb{P}(r_k = l | x_{0:k-1}, y_{1:k-1}), \quad (21)$$

we have

$$q_{k|k}^{(i)}(l) \propto p(y_k, x_k^{(i)} | r_k = l, x_{0:k-1}^{(i)}, y_{1:k-1}) q_{k|k-1}^{(i)}(l) \quad (22a)$$

$$\propto p_{\theta_l}(y_k | x_k^{(i)}, r_k = l) p_{\theta_l}(x_k^{(i)} | x_{k-1}^{(i)}, r_k = l) q_{k|k-1}^{(i)}(l). \quad (22b)$$

Now defining

$$\alpha_k^{(i)}(l) \triangleq p_{\theta_l}(y_k | x_k^{(i)}, r_k = l) p_{\theta_l}(x_k^{(i)} | x_{k-1}^{(i)}, r_k = l) q_{k|k-1}^{(i)}(l) \quad (23)$$

we obtain

$$q_{k|k}^{(i)}(l) = \alpha_k^{(i)}(l) / \sum_{j=1}^s \alpha_k^{(i)}(j), \quad (24)$$

for  $l \in S$  and  $i = 1, \dots, N$ .

(d) **Update step for particle filter:** as the continuous state can be recursively propagated using the following relation:

$$p(x_{0:k} | y_{1:k}) \propto p(y_k, x_k | x_{0:k-1}, y_{1:k-1}) p(x_{0:k-1} | y_{1:k-1}), \quad (25)$$

the corresponding weight update equation for the particle filtering is given by

$$w_k^{(i)} = \frac{p(x_k^{(i)}, y_k | x_{0:k-1}^{(i)}, y_{1:k-1})}{\pi_k(x_k^{(i)} | x_{0:k-1}^{(i)}, y_{1:k})} \tilde{w}_{k-1}^{(i)}; \quad \tilde{w}_k^{(i)} = w_k^{(i)} / \sum_{j=1}^N w_k^{(j)}, \quad (26)$$

where  $\{\tilde{w}_k^{(i)}\}_{i=1}^N$  are the normalized weights. The numerator in (26) can be obtained as

$$p\left(\begin{matrix} x_k^{(i)} \\ y_k \end{matrix} \middle| \begin{matrix} x_{0:k-1}^{(i)} \\ y_{1:k-1} \end{matrix}\right) = \sum_{l=1}^s p\left(\begin{matrix} y_k \\ x_k^{(i)} \end{matrix} \middle| \begin{matrix} r_k = l, x_{0:k-1}^{(i)} \\ y_{1:k-1} \end{matrix}\right) \mathbb{P}\left(r_k = l \middle| \begin{matrix} x_{0:k-1}^{(i)} \\ y_{1:k-1} \end{matrix}\right), \quad (27)$$

which is basically given by the normalizing constant of (24). Note that the marginal density  $p(x_k | y_{1:k})$  can be obtained in the particle cloud form as  $p(x_k | y_{1:k}) \approx \sum_{i=1}^N \tilde{w}_k^{(i)} \delta(x_k - x_k^{(i)})$ . The posterior probability of the regime variable can now be obtained as

$$\mathbb{P}(r_k = l | y_{0:k}) = \int \mathbb{P}(r_k = l | x_{0:k}, y_{0:k}) p(x_{0:k} | y_{0:k}) dx_{0:k} \approx \sum_{i=1}^N q_{k|k}^{(i)}(l) \tilde{w}_k^{(i)}. \quad (28)$$

For further details and algorithmic summary, please refer to [38, 42].

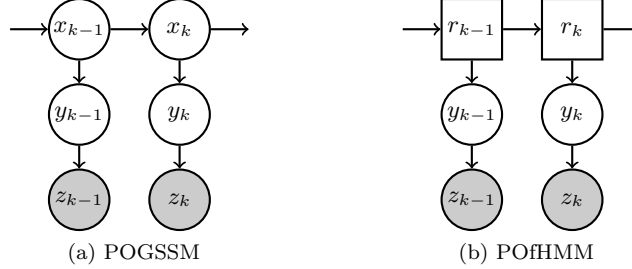


FIGURE 3: Graphical representation of a *partially observed Gaussian state-space model* (POGSSM) and a *partially observed finite HMM* (POfHMM).

### 5.3 Partially observed Gaussian state-space models

Consider the following *partially observed Gaussian state-space models* (POGSSM) [2]

$$x_k = F_k x_{k-1} + G_k v_k, \quad (29a)$$

$$y_k = H_k x_k + H_k^e e_k, \quad (29b)$$

$$z_k \sim p(z_k | y_k), \quad (29c)$$

where  $x_k$  is a latent (Markov) state process with initial density given by a Gaussian with mean  $\hat{x}_0$  and variance  $P_0$  as  $p(x_0) = \mathcal{N}(\hat{x}_0, P_0)$ ;  $y_k$  and  $z_k$  are latent data process and observed data process, respectively. The processes  $x_k$  and  $y_k$  together define a standard linear Gaussian state-space model. Compared to standard LGSS model, however, one observes  $z_k$  instead of  $y_k$ .  $z_k$  is typically a quantized or censored version of  $y_k$ . For example, in a dynamic probit model, we have

$$z_k = \begin{cases} 1, & y_k > 0 \\ 0, & y_k \leq 0, \end{cases} \quad (30)$$

whereas for a dynamic tobit model,

$$z_k = \max(y_k, 0). \quad (31)$$

The graphical representation of the model is shown in Figure 3(a).

For inference purpose, the joint density can be decomposed as

$$p(x_{0:k}, y_{0:k} | z_{0:k}) = p(x_{0:k} | y_{0:k}, z_{0:k}) p(y_{0:k} | z_{0:k}). \quad (32)$$

Then the density  $p(y_{0:k} | z_{0:k})$  can be targeted using a PF in the form of  $N$  weighted particles as  $p(y_{0:k} | z_{0:k}) \approx \sum_{i=1}^N \omega_k^{(i)} \delta(y_{0:k} - y_{0:k}^{(i)})$ . Now for each sequence  $y_{0:k}^{(i)}$ , the conditional density  $p(x_{0:k} | y_{0:k}^{(i)}, z_{0:k})$  is analytically tractable



using a KF. Finally,  $p(x_{0:k}|z_{0:k})$  can be obtained by marginalizing over  $y_{0:k}$  as

$$p(x_{0:k}|z_{0:k}) = \int p(x_{0:k}|y_{0:k}, z_{0:k})p(y_{0:k}|z_{0:k}) dy_{0:k} \approx \sum_{i=1}^N \omega_k^{(i)} p(x_{0:k}|y_{0:k}^{(i)}, z_{0:k}). \quad (33)$$

Thus  $p(x_{0:k}|z_{0:k})$  is obtained as random weighted mixture of Gaussians. For algorithmic details, model extensions and applications, please refer to the original article [2].

#### 5.4 Partially observed finite HMM

Consider the following model described probabilistically as

$$r_k \sim \Pi(r_k|r_{k-1}), \quad y_k \sim p(y_k|r_k), \quad z_k \sim p(z_k|y_k), \quad (34)$$

where the dynamics of  $r_k$  is given by a homogeneous first order finite state Markov chain. Here  $r_k$  and  $y_k$  together describe a *partially observed finite HMM* (POfHMM). However instead of  $y_k$ ,  $z_k$  is actually observed. The graphical representation of the model is shown in Figure 3(b).

Similar to (5.3), the joint density can be decomposed as

$$p(r_k, y_{0:k}|z_{0:k}) = p(r_k|y_{0:k}, z_{0:k})p(y_{0:k}|z_{0:k}). \quad (35)$$

Again, the density  $p(y_{0:k}|z_{0:k})$  can be targeted using a PF in the form of  $N$  weighted particles as  $p(y_{0:k}|z_{0:k}) \approx \sum_{i=1}^N \omega_k^{(i)} \delta(y_{0:k} - y_{0:k}^{(i)})$ . Then for each sequence  $y_{0:k}$ ,  $p(r_k|y_{0:k}, z_{0:k})$  can be obtained analytically using the property of finite HMM. For details, see [2].

#### 5.5 Conditionally conjugate latent process model

The key enabler here is the concept of Bayesian conjugacy. For a latent variable with known observation likelihood, if we can choose a suitable prior (represented with a parametric distribution) such that the prior and posterior belong to the same parametric family, then it is said to be conjugate to the observation likelihood. As a result, the posterior can be obtained in closed form by updating the parameters of the prior distributions. Such conjugate priors are known to exist for the members of the exponential family distributions [6]. Using this conjugacy, one can define a latent process in such a way that both the filtering and prediction distributions are conjugate to the observation distribution in each time step. The latent process so defined is called the conjugate latent process [47].

Now if the state space can be decomposed such that part of the state space follows a conjugate latent process conditioned on the other part, one can then marginalize this part analytically. In the particle filtering framework, this leads to another Rao-Blackwellization approach. One typical application of

the resulting RBPF is a noise adaptive particle filtering for a general state-space model [32, 37, 39]. A brief description of the general approach is given below.

### 5.5.1 Rao-Blackwellized noise adaptive particle filtering

For many practical applications, complete knowledge of the noise processes in the state-space models is unavailable. Typically for such cases, noises are assumed to be generated from a known parametric family of distributions with unspecified parameters. In such setting, one needs to estimate the noise parameters along with the latent state using the available noisy observations. This problem is known as noise adaptive filtering. Moreover, the noise may be non-stationary or state dependent, which prevents off-line tuning. For the above reasons, one needs a sequential method to learn the noise parameters using streamed sensor data.

When the noise distribution is a member of the exponential family, we can model the unknown noise parameter as conjugate latent process. To describe the resulting filtering problem, we first start with a description of the model (including the noise model), which is followed by the problem definition and implementation details.

**Model description:** consider the following general state-space model relating a (Markovian) latent state  $x_k$  to the observation  $y_k$  at time step  $k$  as

$$x_k = f_k(x_{k-1}, u_{k-1}) + g_k(x_{k-1}, u_{k-1})v_k, \quad (36a)$$

$$y_k = h_k(x_k, u_k) + d_k(x_k, u_k)w_k. \quad (36b)$$

Here,  $f_k(\cdot)$ ,  $h_k(\cdot)$ ,  $d_k(\cdot)$  and  $g_k(\cdot)$  are nonlinear functions of the state vector  $x_k$  and the deterministic input  $u_k$  at time  $k$ . Here  $d_k(\cdot)$  is assumed to be invertible whereas  $g_k(\cdot)$  is not, as most motion models in practice, including those with integrators, lead to non-invertible  $g_k(\cdot)$ .

**Noise model:** We define the noise vector  $e_k \triangleq [v_k^T, w_k^T]^T$  as a realization from a distribution belonging to the exponential family as

$$e_k \sim p(e_k | \theta_k) = \rho(e_k) \exp(\eta(\theta_k) \cdot \tau(e_k) - \phi(\theta_k)), \quad (37)$$

where  $\theta_k$  is the vector of unknown parameters,  $\eta(\theta_k)$  and  $\phi(\theta_k)$  are vector and scalar valued functions of the parameters, respectively;  $\rho(e_k)$  and  $\tau(e_k)$  are scalar and vector valued functions of the realization  $e_k$ ; and the symbol ' $\cdot$ ' denotes scalar product of two vectors. Further, following [32], assume that the unknown noise parameters  $\theta_k$  are slowly varying in time. This slowly varying nature can arise, *e.g.*, due to model misspecification [48].

Now for a given realization of the noise  $e_k$ , (37) provides the likelihood function for the unknown noise parameter  $\theta_k$ . Accordingly, we can select a conjugate prior on  $\theta_k$  as

$$p(\theta_k | e_{1:k-1}) = \gamma^{-1}(V_{k|k-1}, \nu_{k|k-1}) \times \exp(\eta(\theta_k)V_{k|k-1} - \nu_{k|k-1}\phi(\theta_k)), \quad (38)$$

where  $V_{k|k-1}$  is a vector of sufficient statistics and  $\nu_{k|k-1}$  is a scalar counter of the effective number of samples in the statistics. The normalization factor  $\gamma(V_{k|k-1}, \nu_{k|k-1})$  is uniquely determined by the statistics  $V_{k|k-1}$  and  $\nu_{k|k-1}$ . Then, the posterior density  $p(\theta_k|e_{1:k})$  is also of the form (38) with statistics

$$V_{k|k} = V_{k|k-1} + \tau(e_k), \quad \nu_{k|k} = \nu_{k|k-1} + 1. \quad (39)$$

This is a convenient result for recursive evaluation of sufficient statistics starting from a prior defined by  $V_0$  and  $\nu_0$ . Further, the predictive distribution of  $e_k$  is then given by

$$p(e_k|e_{1:k-1}) = \int p(e_k|\theta_k)p(\theta_k|e_{1:k-1}) d\theta_k = \frac{\gamma(V_{k|k}, \nu_{k|k})}{\gamma(V_{k|k-1}, \nu_{k|k-1})} \rho(e_k). \quad (40)$$

Now as  $\theta_k$  is assumed to be slowly varying in time, to complete the state-space model, an evolution model for  $\theta_k$  (*i.e.*,  $p(\theta_k|\theta_{k-1})$ ) is required. Given this evolution model, the predictive density of the parameter  $\theta_{k+1}$  is obtained as

$$p(\theta_{k+1}|e_{1:k}) = \int p(\theta_{k+1}|\theta_k)p(\theta_k|e_{1:k}) d\theta_k. \quad (41)$$

If this predictive density conforms to the specification (38), then the noise parameter vector follows a *conjugate latent process*. However, typically the transition model  $p(\theta_{k+1}|\theta_k)$  is unknown. In [32], the evolution is defined implicitly as a Kullback-Leibler norm constraint on the time variability which leads to an exponential forgetting mechanism operating on the sufficient statistics. This keeps the predictive density  $p(\theta_{k+1}|e_{1:k})$  in the same class as in (38) with statistics updated from (39) as

$$\nu_{k+1|k} = \lambda\nu_{k|k} + (1 - \lambda)\nu_u, \quad (42a)$$

$$V_{k+1|k} = \lambda V_{k|k} + (1 - \lambda)V_u, \quad (42b)$$

where  $\lambda \in (0, 1]$  is a forgetting factor and the invariant measure is assumed to be also in the exponential form (38) with statistics  $\nu_u, V_u$ . The details can be found in [32].

**Problem definition:** We are concerned with the evaluation of the joint posterior density  $p(x_k, \theta_k|y_{1:k})$  sequentially over time. Suppose at time  $k - 1$  we have the posterior  $p(x_{k-1}, \theta_{k-1}|y_{1:k-1})$ . With the arrival of a new observation  $y_k$ , we want to update the posterior recursively to  $p(x_k, \theta_k|y_{1:k})$ .

**Joint estimation of state and noise parameters:** The joint posterior density at time  $k$  can be decomposed as follows:

$$p(x_{0:k}, \theta_k|y_{0:k}) = p(\theta_k|x_{0:k}, y_{0:k}) p(x_{0:k}|y_{0:k}). \quad (43)$$

The posterior propagation of the latent state  $x_k$  is targeted through a

PF as  $p(x_{0:k}|y_{0:k}) \approx \sum_{i=1}^n \omega_k^{(i)} \delta(x_{0:k} - x_{0:k}^{(i)})$ . Now for a given  $(x_{0:k}^{(i)}, y_{0:k})$ ,  $p(\theta_k|x_{0:k}, y_{0:k})$  can be considered as the posterior density of the latent parameter  $\theta_k$  and can be propagated analytically using a conjugate latent process as discussed earlier. To see this, note that for a known  $x_k^{(i)}$ , (36a)–(36b) can be transformed into

$$e_k^{(i)} = e(x_k^{(i)}, y_k) = \begin{bmatrix} g_k^\dagger(x_{k-1}^{(i)}, u_{k-1})[x_k^{(i)} - f_k(x_{k-1}^{(i)}, u_{k-1})] \\ d_k^{-1}(x_k^{(i)}, u_k)[y_k - h_k(x_k^{(i)}, u_k)] \end{bmatrix}, \quad (44)$$

where  $g_k^\dagger(x_{k-1}^{(i)}, u_{k-1})$  stands for the Moore-Penrose pseudo-inverse of  $g_k(x_{k-1}^{(i)}, u_{k-1})$ . Consequently,  $p(\theta_k|x_{0:k}^{(i)}, y_{0:k}) = p(\theta_k|e_{0:k}^{(i)})$ . So the parameter posterior follows a conditional conjugate process and the analytical propagation of this posterior can be obtained easily using (37)–(41).

The joint density (43) can be represented as

$$p(x_{0:k}, \theta_k|y_{0:k}) \approx \sum_{i=1}^n \omega_k^{(i)} p(\theta_k|V_{k|k}^{(i)}, \nu_{k|k}^{(i)}) \delta(x_{0:k} - x_{0:k}^{(i)}), \quad (45)$$

where the statistics  $\omega_k^{(i)}$ ,  $V_{k|k}^{(i)}$ ,  $\nu_{k|k}^{(i)}$ , and  $x_{0:k}^{(i)}$  are evaluated as follows. First,  $x_k^{(i)}$  are sampled from a proposal density  $q(x_k|x_{0:k-1}^{(i)}, y_{0:k-1})$ . Second, for the known value  $x_k^{(i)}$ ,  $p(\theta_k|x_{0:k}^{(i)}, y_{0:k})$  is updated using the mapping (44) to  $e_k^{(i)}$ , and the statistics  $V_k^{(i)}$  and  $\nu_k^{(i)}$  are updated using (39). Finally, using the recursive relation  $p(x_{0:k}|y_{0:k}) \propto p(y_k, x_k|x_{0:k-1}, y_{0:k-1}) p(x_{0:k-1}|y_{0:k-1})$ , weights can be updated as

$$\omega_k^{(i)} \propto \frac{p(y_k, x_k|x_{0:k-1}, y_{0:k-1})}{q(x_k|x_{0:k-1}^{(i)}, y_{0:k-1})} w_{k-1}^{(i)},$$

where

$$p(y_k, x_k|x_{0:k-1}, y_{0:k-1}) = \int p(y_k, x_k|\theta_k, x_{0:k-1}, y_{0:k-1}) p(\theta_k|x_{0:k-1}, y_{0:k-1}) d\theta_k,$$

is the marginal predictive distribution of  $x_k$  and  $y_k$ . This marginal distribution can be computed by integrating out the unknown parameters, leading to the predictive distribution of  $x_k$  and  $y_k$ , and consequently  $e_k$  via (44). Note that the predictive distribution  $p(e_k|e_{0:k-1})$  is readily available in the form of (40) for the exponential family. The predictor (46) can then be obtained using the transformation of variables in the probability density functions:

$$p(y_k, x_k|x_{0:k-1}, y_{0:k-1}) = |J(x_k, y_k)| p(e(x_k, y_k)|V_{k|k-1}^{(i)}, \nu_{k|k-1}^{(i)}). \quad (46)$$

where  $J(x_k, y_k)$  is the Jacobian of the transformation (44) and  $p(e_k|\cdot)$  is given by (40). For further details and an algorithmic summary, we refer to [32].

**Remark 1.** *If the noise vector  $e_k$  follows a multivariate Normal distribution, with unknown mean  $\mu_k$  and covariance  $\Sigma_k$ , a Normal-inverse-Wishart distribution defines a conjugate prior.*

**Remark 2.** *This approach is closely related to the works on stationary parameter estimation using RBPF by [9, 16, 45]. Note that the estimation of stationary parameters is a special case of the above approach with  $\lambda = 1$ , which reduces the sufficient statistics update in (39) to the form of [45]. For the station-*

ary parameter, however, this approach is known to suffer from error accumulations as pointed out by [12]. Here, we note that for a sequence of  $\lambda_k < 1, \forall k$ , the posterior density  $p(\theta_k | x_{1:k}, y_{1:k})$  and thus  $p(y_k, x_k | x_{0:k-1}, y_{0:k-1})$  satisfies the exponential forgetting property that is well known to mitigate the path degeneracy problem [13].

---

## 6 Terrain navigational problems

In this section, we outline the problem of positioning an aircraft without the aid of *global positioning systems* (GPS). GPS is vulnerable to jamming, we therefore consider the positioning by using the measurements from alternative sensors. Specifically, one can position the aircraft using only altitude measurements, measurements of the distance down to the ground, measurements of the aircraft's acceleration, and an elevation map of the visited area.

The navigation problem results in an estimation problem with linear dynamics, where the aircraft is assumed to follow a constant velocity model, and a nonlinear measurement equation in the position,

$$x_{k+1} = \begin{bmatrix} I_2 & TI_2 \\ 0 & I_2 \end{bmatrix} x_k + \begin{bmatrix} 0_2 \\ TI_2 \end{bmatrix} (w_k + a_k) \quad (47a)$$

$$y_k = h^{\text{map}}(x_k) + e_k, \quad (47b)$$

where  $T$  is the sampling period,  $w_k$  and  $e_k$  are the independent Gaussian process and the measurement noise, respectively. The first two components of the state are the position and the next two are the velocity in two dimensional Cartesian coordinates (representing north and east). The altitude is well decoupled from these two directions, and is best handled separately. The aircraft's body accelerations  $a_k$ , are measured and enter into the system as input. The inertial sensors in an aircraft are generally good enough to provide very accurate information about accelerations. The function  $h^{\text{map}}$  relates the terrain altitude to position in the form of a digital terrain database and is highly nonlinear. The measurements are obtained by deducting the height measurement of the ground looking (on board) radar from the known altitude of the aircraft (obtained using an altimeter). Furthermore, the level of the ground can be expected to be similar in many local regions of the map, hence the position estimate tends to be a multi-modal distribution, especially after leaving the relatively uninformative planar areas. This makes the PF a natural choice when implementing this type of application (as the alternatives such as extended KF struggles with dealing multimodal distributions). However, terrain navigation is often used in systems with limited computational resources and a straightforward PF implementation is often not feasible.

Fortunately, the model that describes this navigation problem, can easily be identified as an ML/NGSSM; the Rao-Blackwellization conditioned on the position,  $x^n$ , the remaining state-space model describing the velocity,  $x^l$ , becomes linear-Gaussian. Applying a RBPF to this problem using Algorithm 1, is then straightforward. The following model components are identified from (11)

$$\text{as } f^n(x_k^n) = x_k^n \quad F^n(x_k^n) = TI_2 \quad h(x_k^n) = h^{\text{map}}(x_k) \quad (48a)$$

$$f^l(x_k^n) = 0_{2 \times 1} \quad F^l(x_k^n) = I_2 \quad H(x_k^n) = 0, \quad (48b)$$

where  $h^{\text{map}}(x_k)$ , depends only on the position component of the state, *i.e.*, the nonlinear part of the model. Furthermore,  $w_k^n \equiv 0$  (*i.e.*, no process noise enters directly on the nonlinear part),  $w_k^l = w_k$ , and  $e_k$  is the same as in the original problem formulation.

A more accurate measurement noise model would also consider secondary reflections (*e.g.*, from the tree canopies), resulting in a bi-modal Gaussian mixture representation. This could provide further structure to the problem. This application has been studied extensively in [5, 31].

---

## 7 Approximate Rao-Blackwellized nonlinear filtering

The implementation of RBPF requires that a subset of the state vector is analytically tractable conditioned on the rest of the state vector. Then one can marginalize the tractable part to obtain the posterior distribution of the non-tractable part using PF. However, for many applications, this tractability requirement is very restrictive. For example, consider the noise adaptive filtering where the noise class is not from the exponential family. For this case, the unknown noise parameters, in general, do not admit Bayesian conjugate priors. Consequently one cannot implement an RBPF for this case. For such cases, in recent times, different approximations of the RBPF schemes have been envisaged in the literature (see *e.g.*, [1, 11, 27, 41, 46]). The basic idea again is to decompose the whole state space into two (interacting) parts<sup>6</sup> and split the filtering problem into two nested sub-problems.

To illustrate the idea, suppose that the state vector at time  $k$  is given by  $(x_k, z_k)^T$ . The state vector is assumed to follow a Markovian model. Now given the sequence of observations  $\{y_0, y_1, \dots, y_k\}$ , we target to evaluate  $p(x_{0:k}, z_{0:k} | y_{0:k})$  recursively over  $k$ . The target distribution can be decomposed as

$$p(x_{0:k}, z_{0:k} | y_{0:k}) = p(x_{0:k} | y_{0:k}) p(z_{0:k} | x_{0:k}, y_{0:k}) \quad (49)$$

We now propose to handle each sub-problem using PFs. First using a single PF, with  $N_x$  particles, we approximate  $p(x_{0:k} | y_{0:k})$ . Then we run  $N_x$  condi-

---

<sup>6</sup>In principle, the state space can be decomposed into more than two parts to implement an approximate RBPF.

tional PFs, each with  $N_z$  particles, to estimate  $p(z_{0:k}|x_{0:k}, y_{0:k})$ . This is in essence an approximated RBPF, where we use local Monte Carlo approximations instead of analytical solutions. This idea was proposed in [27, 46] and also in [41] in the context of noise adaptive PF. A related but slightly different formulation is considered in [1, 11]. It is important to here emphasize that the resampling steps for the second sub-problem is local in the state space and thus can be parallelized. This is a very generic and promising approach for high dimensional online inference problems. On the one hand it increases the parallelism in PF, on the other hand, it provides flexibility in implementing specialized PFs with different tailored solutions for individual sub-problems.

---

## 8 Concluding remarks

The online (real time) inference problems for general state-space models do not usually, admit closed form solutions. However, thanks to the modern Monte Carlo methods, the inference aim can be achieved to an arbitrary accuracy using *particle filtering* (PF). Although, PF is a very flexible class of methods for such inference problems, it is known to suffer from the curse of dimensionality. For some models, part of the state space admits conditionally tractable structures. The possibility to exploit such conditional tractability, leaves only a part of the state vector to be targeted using a PF. Consequently, Monte Carlo based estimation is confined to a space of lower dimension. This in term, also leads to an efficiency improvement of the (filtering) estimator due to the implications of the *Rao-Blackwell theorem*. The resulting filtering problem is popularly known as *Rao-Blackwellized particle filtering* (RBPF).

In this chapter, we presented a brief review of such RBPF. In particular, we have outlined a range of existing RBPF models that are popular in practice and also presented some recent developments in this context. This is written mainly from the practitioners' point of view and as such, we have not included any theoretical analysis here. Our coverage is primarily aimed at providing the model structures admitting such conditional tractabilities. Nonetheless, the coverage is by no means exhaustive and partially biased towards our interest. Many other such RBPF may possibly be constructed, *e.g.*, using the exact filtering sublayer as provided in Beneš and Daum [36].

---

## Acknowledgment

The authors would like to thank COOP-LOC, funded by SSF and CADICS, funded by Swedish Research Council (VR) for the financial supports.

---

## ***Bibliography***

---

- [1] M. O. Ahmed, P. T. Bibalan, N. de Freitas, and S. Fauv. Decentralized, adaptive, look-ahead particle filtering. *arXiv:1203.2394v1*, 2012.
- [2] C. Andrieu and A. Doucet. Particle filtering for partially observed Gaussian state space models. *Journal of the Royal Statistical Society Series B*, 64(4):827–836, 2002.
- [3] C. Andrieu and S. J. Godsill. A particle filter for model based audio source separation. In *Proceedings of the 2000 International Workshop on Independent Component Analysis and Blind Signal Separation*, Helsinki, Finland, June 2000.
- [4] A. Bagchi. *Optimal Control of Stochastic Systems*. Prentice Hall Inc., NJ, USA, 1993.
- [5] N. Bergman, L. Ljung, and F. Gustafsson. Terrain navigation using Bayesian statistics. *IEEE Control Systems Magazine*, 19(3):33–40, June 1999.
- [6] J. M. Bernardo and A. F. M. Smith. *Bayesian Theory*. John Wiley & Sons, 1994.
- [7] D. Blackwell. Conditional expectation and unbiased sequential estimation. *The Annals of Mathematical Statistics*, 18(1):105–110, March 1947.
- [8] O. Cappé, S. J. Godsill, and E. Moulines. An overview of existing methods and recent advances in sequential Monte Carlo. *Proceedings of the IEEE*, 95(5):899–924, 2007.
- [9] C. M. Carvalho, M. Johannes, H. F. Lopes, and N. Polson. Particle learning and smoothing. *Statistical Science*, 25(1):88–106, 2010.
- [10] R. Chen and J. Liu. Mixture Kalman filters. *Journal of the Royal Statistical Society Series B*, 62(3):493–508, 2000.
- [11] T. Chen, T. B. Schön, H. Ohlsson, and L. Ljung. Decentralized particle filter with arbitrary state decomposition. *IEEE Transactions on Signal Processing*, 59(2):465–478, February 2011.
- [12] N. Chopin, A. Iacobucci, J. M. Marin, K. L. Mengersen, C. P. Robert, R. Ryder, and C. Schafer. On particle learning. *arXiv: 1006.0554v2*, 2010.



- [13] D. Crisan and A. Doucet. A survey of convergence results on particle filtering methods for practitioners. *IEEE Transactions on Signal Processing*, 50(3):736–746, 2002.
- [14] F. Daum, R. Co, and J. Huang. Curse of dimensionality and particle filters. In *Proceedings of IEEE Aerospace Conference*, Big Sky, USA, 2003.
- [15] P. Del Moral. *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer Verlag New York, 2004.
- [16] P. M. Djuric and J. Miguez. Sequential particle filtering in the presence of additive Gaussian noise with unknown parameters. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1621–1624, Orlando, Florida, 2002.
- [17] R. Douc, O. Cappé, and E. Moulines. Comparison of resampling schemes for particle filtering. In *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis*, Zagreb, Croatia, 2005.
- [18] A. Doucet and A. M. Johansen. A tutorial on particle filtering and smoothing: Fifteen years later. In *Oxford Handbook of Nonlinear Filtering*. Oxford University Press, 2011.
- [19] A. Doucet, S. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10:197–208, 2000.
- [20] N. J. Gordon, D. J. Salmond, and A. F. M. Smith. Novel approach to nonlinear/ non Gaussian Bayesian state estimation. *IEE Proc. F-Radar and Signal Processing*, 140(2):107–113, April 1993.
- [21] F. Gustafsson. Particle filter theory and practice with positioning applications. *IEEE Aerospace and Electronic Systems Magazine*, 25(7):53–81, March 2010.
- [22] F. Gustafsson. *Statistical Sensor Fusion*. Studentlitteratur, Malmo, Sweden, 2010.
- [23] F. Gustafsson, F. Gunnarson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P.-J. Nordlund. Particle filters for positioning, navigation, and tracking. *IEEE Transactions on Signal Processing*, 50(2):425–437, February 2002.
- [24] A. Gut. *An Intermediate Course in Probability*. Springer-Verlag, 1995. ISBN 0-387-94507-5.
- [25] G. Hendeby. *Performance and Implementation Aspects of Nonlinear Filtering*. Dissertations no 1161, Linköping Studies in Science and Technology, March 2008.

- [26] G. Hendebey, R. Karlsson, and F. Gustafsson. The Rao-Blackwellized Particle Filter: A Filter Bank Implementation. *EURASIP Journal on Advances in Signal Processing*, 2010(724087), 2010.
- [27] A. M. Johansen, N. Whiteley, and A. Doucet. Exact approximation of Rao-Blackwellised particle filters. In *Proceedings of 16th IFAC Symposium on Systems Identification (SYSID)*, Brussels, Belgium, 2012.
- [28] E. L. Lehmann. *Theory of Point Estimation*. Probability and Mathematical Statistics. John Wiley & Sons, Ltd, 1983. ISBN 0-471-05849-1.
- [29] M. Montemerlo, S. Thrun, D. Koller, and W. Wegbreit. FASTSLAM: A factored solution to the simultaneous localization and mapping problem. In *Proceedings of the AAAI National Conference on Artificial Intelligence*, Edmonton, Canada, 2002.
- [30] K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge, Massachusetts, 2012.
- [31] P.-J. Nordlund. *Sequential Monte Carlo Filters and Integrated Navigation*. Licentiate thesis no 945, Department of Electrical Engineering, Linköpings universitet, Sweden, May 2002.
- [32] E. Ozkan, V. Smidl, S. Saha, C. Lundquist, and F. Gustafsson. Marginalized adaptive particle filtering for non-linear models with unknown time-varying noise parameters. *Automatica*, 49(6):1566–1575, 2013.
- [33] P. B. Quang, C. Musso, and F. LeGland. An insight into the issue of dimensionality in particle filtering. In *Proceedings of 13th International Conference on Information Fusion (FUSION)*, Edinburgh, Scotland, 2010.
- [34] L. R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of IEEE*, 77(2):257–286, 1989.
- [35] C. R. Rao. Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, 37: 81–91, 1945.
- [36] B. Ristic, S. Arulampalam, and N. Gordon. *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House, 2004.
- [37] S. Saha and F. Gustafsson. Particle filtering with dependent noise processes. *IEEE Transactions on Signal Processing*, 60(9):4497–4508, 2012.
- [38] S. Saha and G. Hendebey. Online inference in Markov modulated nonlinear dynamic systems: a Rao-Blackwellized particle filtering approach. *ArXiv: 1311.6486v1*, 2013.

- [39] S. Saha, E. Ozkan, F. Gustafsson, and V. Smidl. Marginalized particle filters for Bayesian estimation of Gaussian noise parameters. In *Proceedings of 13th International Conference on Information Fusion (FUSION)*, Edinburgh, Scotland, 2010.
- [40] S. Saha, U. Orguner, and F. Gustafsson. Non-linear filtering based on observations from Student's t processes. In *Proceedings of the IEEE Aerospace Conference*, Big Sky, USA, 2012.
- [41] S. Saha, G. Hendeby, and F. Gustafsson. Noise Adaptive Particle Filtering: A Hierarchical Perspective. *Working paper (part of the paper was presented in ISBA Regional Meeting and International Workshop/Conference on Bayesian Theory and Applications (IWCBTA), Varanasi.)*, 2013.
- [42] Saikat Saha and Gustaf Hendeby. Rao-Blackwellized particle filter for Markov modulated nonlinear dynamic systems. In *IEEE Workshop on Statistical Signal Processing*, Gold Coast, Australia, June 2014.
- [43] T. Schön, F. Gustafsson, and P. J. Nordlund. Marginalized particle filter for mixed linear/nonlinear state space models. *IEEE Transactions on Signal Processing*, 53(7):2279–2289, 2005.
- [44] C. Snyder, T. Bengtsson, P. Bickel, and J. Anderson. Obstacles to high-dimensional particle filtering. *Monthly Weather Review*, 136(12):4629–4640, 2008.
- [45] G. Storvik. Particle filters for state space models with the presence of unknown static parameters. *IEEE Transactions on Signal Processing*, 50(2):281–289, 2002.
- [46] C. Vergé, C. Dubarry, P. Del Moral, and E. Moulines. On parallel implementation of sequential Monte Carlo methods: the island particle model. *arXiv:1306.3911v1*, 2013.
- [47] P. Vidoni. Exponential family state space models based on a conjugate latent process. *Journal of the Royal Statistical Society. Series B*, 61(1): 213–221, 1999.
- [48] M. West and P. J. Harrison. *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, New York, USA, 1997.