

Henri Pesonen & Robert Piché

**Estimation of Linear Systems with Abrupt Changes of the  
Noise Covariances Using Variational Bayes Algorithm**



Tampereen teknillinen yliopisto. Matematiikan laitos.  
Tutkimusraportti 100  
Tampere University of Technology. Department of Mathematics.  
Research Report 100

Henri Pesonen & Robert Piché

## **Estimation of Linear Systems with Abrupt Changes of the Noise Covariances Using Variational Bayes Algorithm**

Tampere University of Technology. Department of Mathematics  
Tampere 2012

ISBN 978-952-15-2923-8  
ISSN 1459-3750

# Estimation of Linear Systems with Abrupt Changes of the Noise Covariances Using Variational Bayes Algorithm

Henri Pesonen and Robert Piché

**Abstract**—The variational Bayes method is applied to the state-space estimation problem with maneuvers or changes in the covariance of the observation noise. The resulting algorithm is an off-line batch method that can be used to provide a baseline performance estimation results for the recursive methods. In addition to batch methods we introduce a heuristic approach to make the algorithm on-line. Through simulations we show how the introduced method achieves the best accuracy out of all compared approximative estimation methods.

**Index Terms**—Linear systems, Bayesian methods, change detection algorithms, fault detection, Kalman filters, smoothing methods

## I. INTRODUCTION

THE estimation of the state described by a linear state-space system requires that we define the parameters of the system, after which we can solve the system optimally with the Kalman filter (KF) if the system noise processes are normally distributed white processes [1]. However, it might be problematic to describe the processes with a Gaussian distribution. Gaussian mixture (GM) distributions are more general models that can take into account several plausible models for the system [2], [3]. For example, a navigation system with maneuvers can be described with one model for the constant velocity motion and another model for the maneuvers [4]. Also, systems with measurement outliers can be described with one model for the good data and another for the bad [5]. Although the most straightforward approach would be to use a single Gaussian model even if there are multiple models, the resulting performance may be degraded. A popular approach for the problem is to use change detection methods [6], [7]. These range from statistical tests [8] to multiple model filtering methods [9]. Also robust filtering methods could be interpreted as methods for detecting changes, or outliers, and adapting the performance accordingly [5], [10]. In the present work we derive the variational Bayes (VB) algorithm for approximating the posterior distribution of the state within a time-window, given that the noise processes are described with a two-component GM distribution. The VB method can be applied as a batch method for computing the posterior distribution offline for the whole track, or as an online method using a moving window.

This note is organized as follows. In Section II we describe the linear state-space model with GM-noise processes. In Section III we formulate the VB algorithm for the problem and in Section IV we test several methods in two sets of simulations. Finally in Section V we conclude the study.

## II. PROBLEM

We model the problem of the abruptly changing linear dynamic system as follows. The system is constructed as a linear state-space model

$$x_k = F_{k-1}x_{k-1} + w_{k-1}, \quad (1)$$

$$y_k = H_k x_k + v_k, \quad (2)$$

$$x_0 \sim \mathcal{N}(x_{0|0}, P_{0|0}), \quad (3)$$

where  $v_k$  and  $w_k$  are mutually independent white noise processes independent of the initial state  $x_k$ , and  $\mathcal{N}(\mu, \Sigma)$  is the normal distribution with the mean  $\mu$  and the covariance  $\Sigma$ . The noise processes are mixtures of two plausible models and are defined as follows. The state noise is modeled as

$$w_k \sim \mathcal{N}(0, Q_k)^{1-\lambda_{k+1}} \mathcal{N}(0, M_k)^{\lambda_{k+1}}, \quad (4)$$

and the observation noise as

$$v_k \sim \mathcal{N}(0, R_k)^{1-\lambda_k} \mathcal{N}(0, W_k)^{\lambda_k}, \quad (5)$$

where  $\lambda_k \in \{0, 1\}$ ,  $k = 1, \dots, N$  are mutually independent Bernoulli-distributed random variables

$$\lambda_k \sim \text{Ber}(\theta). \quad (6)$$

In theory the problem can be solved using the Bayesian framework. The posterior distribution of the state is

$$\begin{aligned} p(x_{0:N} | y_{1:N}) \\ = \sum_{\lambda_1=0}^1 \cdots \sum_{\lambda_N=0}^1 p(x_{0:N} | y_{1:N}, \lambda_{1:N}) p(\lambda_{1:N} | y_{1:N}), \end{aligned} \quad (7)$$

where  $a_{1:N} := [a_1, \dots, a_N]$ . The posterior (7) is a GM distribution and evaluation of its moments is computationally demanding for even small  $N$ . Therefore, we are required to restrict ourself to solving either only parts of the problem, or to approximate the posterior distribution. For example, we could restrict ourselves to computing only the marginal distributions of (7). The marginal  $p(x_k | y_{1:N}, \lambda_{1:N}) = \mathcal{N}(x_{k|N}(\lambda_{1:N}), P_{k|N}(\lambda_{1:N}))$  can be computed recursively using KF for  $k = N$

$$\begin{aligned} & [x_{k+1|k+1}(\lambda_{1:k+1}), P_{k+1|k+1}(\lambda_{1:k+1})] \\ & \leftarrow \text{KalmanStep}(x_{k|k}(\lambda_{1:k}), P_{k|k}(\lambda_{1:k}), y_k, F_k, \\ & \quad Q_k^{1-\lambda_{k+1}} M_k^{\lambda_{k+1}}, H_k, R_k^{1-\lambda_{k+1}} W_k^{\lambda_{k+1}}), \end{aligned} \quad (8)$$

where KalmanStep is one step of the KF algorithm, as given in Algorithm 1.

---

**Algorithm 1**  $[x_{k|k}, P_{k|k}] \leftarrow$   
KalmanStep( $x_{k-1|k-1}, P_{k-1|k-1}, y_k, F_{k-1}, Q_{k-1}, H_k, \mathcal{R}_k$ )

---

- 1:  $x_{k|k-1} \leftarrow F_{k-1}x_{k-1|k-1}$
- 2:  $P_{k|k-1} \leftarrow F_{k-1}P_{k-1|k-1}F_{k-1}^T + Q_{k-1}$
- 3:  $K_k \leftarrow P_{k|k-1}H_k^T(H_kP_{k|k-1}H_k^T + \mathcal{R}_k)^{-1}$
- 4:  $x_{k|k} \leftarrow x_{k|k-1} + K_k(y_k - H_kx_{k|k-1})$
- 5:  $P_{k|k} \leftarrow P_{k|k-1} + K_kH_kP_{k|k-1}$

---

The Rauch-Tung-Striebel smoother (RTS) [11] is a recursive smoothing algorithm for the linear Gaussian state-space model to compute the marginals with  $k < N$

$$\begin{aligned}
& [x_{k|N}(\lambda_{1:N}), P_{k|N}(\lambda_{1:N}), C_{k+1|N}(\lambda_{1:N})] \\
& \leftarrow \text{RTSStep}(x_{k+1|N}(\lambda_{1:N}), P_{k+1|N}(\lambda_{1:N}), x_{k|k}(\lambda_{1:N}), \\
& \quad P_{k|k}(\lambda_{1:N}), F_k, Q_k^{1-\lambda_k} M_k^{\lambda_k}), \quad (9)
\end{aligned}$$

where RTSStep is defined in Algorithm 2. In the algorithm, we compute also the cross-covariance

$$C_{k+1|N}(\lambda_{1:k}) = \text{cov}(x_{k+1}, x_k \mid y_{1:N}, \lambda_{1:k}), \quad (10)$$

as it is required in the VB approximation of the posterior (7) as discussed in Section III.

---

**Algorithm 2**  $[x_{k|N}, P_{k|N}, C_{k+1|N}] \leftarrow$   
RTSStep( $x_{k+1|N}, P_{k+1|N}, x_{k|k}, P_{k|k}, F_k, Q_k$ )

---

- 1:  $x_{k+1|k} \leftarrow F_k x_{k|k}$
  - 2:  $P_{k+1|k} \leftarrow F_k P_{k|k} F_k^T + Q_k$
  - 3:  $G_k \leftarrow P_{k|k} F_k^T P_{k+1|k}^{-1}$
  - 4:  $C_{k+1|N} \leftarrow P_{k+1|N} G_k^T$
  - 5:  $x_{k|N} \leftarrow x_{k|k} + G_k(x_{k+1|N} - x_{k+1|k})$
  - 6:  $P_{k|N} \leftarrow P_{k|k} + G_k(P_{k+1|N} - P_{k+1|k})G_k^T$
- 

The GM component's weight in (7) is obtained by running the bank of  $2^N$  KFs.

$$\begin{aligned}
& p(\lambda_{1:N} \mid y_{1:N}) \propto p(y_{1:N} \mid \lambda_{1:N}) p(\lambda_{1:N}) \\
& = \prod_{k=1}^N p(y_k \mid y_{1:k-1}, \lambda_{1:k}) p(\lambda_k) \quad (11) \\
& = \prod_{k=1}^N \mathbb{N}(y_k \mid H_k x_{k|k-1}(\lambda_{1:k}), S_k(\lambda_{1:k})) (1 - \theta_k)^{1-\lambda_k} \theta_k^{\lambda_k}.
\end{aligned}$$

There exists several methods for making the evaluation of the distributions feasible. Most methods are based on cutting off or merging the branches of the mixture filtering distribution  $p(x_k \mid y_{1:k})$  [2], [3], [5], [9].

### III. VARIATIONAL APPROXIMATION

We use the VB approach to approximate the posterior (7). In the VB method we seek the optimal approximative distributions  $q(\cdot)$  in the factorization

$$\begin{aligned}
& p(x_{1:N}, \lambda_{1:N} \mid y_{1:N}) \\
& \approx q_{x_{1:N}, \lambda_{1:N}}(x_{1:N}, \lambda_{1:N}) =: q_{x_{1:N}}(x_{1:N}) \prod_{k=1}^N q_{\lambda_k}(\lambda_k). \quad (12)
\end{aligned}$$

Distributions  $q_{x_{1:N}}(x_{1:N}), q_{\lambda_1}(\lambda_1), \dots, q_{\lambda_N}(\lambda_N)$  are found such that they minimize the Kullback-Leibler (KL) divergence of the approximative distribution with respect to the posterior

$$\begin{aligned}
& \text{KL}(q(\lambda_{1:N}, x_{1:N}) \parallel p(x_{1:N}, \lambda_{1:N} \mid y_{1:N})) \quad (13) \\
& = \int q(x_{1:N}) \prod_{k=1}^N q(\lambda_k) \log \frac{q(x_{1:N}) \prod_{k=1}^N q(\lambda_k)}{p(x_{1:N}, \lambda_{1:N} \mid y_{1:N})} dx_{1:N} d\lambda_{1:N}.
\end{aligned}$$

In (13) and in the following the notation is simplified by omitting the subscript from the approximative distributions  $q(\cdot)$ . The KL divergence can be minimized using calculus of variations by first fixing  $q(x_{1:N})$  and  $q(\lambda_i)$ ,  $i \in 1 : N \setminus k$ , to minimize (13) with respect to  $q(\lambda_k)$ . As a result, we get

$$\begin{aligned}
& \log q(\lambda_k) \\
& = \mathbb{E}_{x_{1:N}, \lambda_{1:N \setminus k}}(\log p(x_{1:N}, \lambda_{1:N} \mid y_{1:N})) + \text{const.} \quad (14)
\end{aligned}$$

for  $k = 1, \dots, N$ . The expectation  $\mathbb{E}_{x_{1:N}, \lambda_{1:N \setminus k}}(\cdot)$  is evaluated for  $q(x_{1:N}) \prod_{i=1, i \neq k}^N q(\lambda_i)$ . After finding  $q(\lambda_k)$ ,  $k = 1, \dots, N$ , we minimize (13) with respect to  $q(x_{1:N})$  as

$$\begin{aligned}
& \log q(x_{1:N}) \\
& = \mathbb{E}_{\lambda_{1:N}}(\log p(x_{1:N}, \lambda_{1:N} \mid y_{1:N})) + \text{const.}, \quad (15)
\end{aligned}$$

where the expectation  $\mathbb{E}_{\lambda_{1:N}}(\cdot)$  is evaluated for  $\prod_{k=1}^N q(\lambda_k)$ .

To compute (14) and (15), we express the posterior distribution as

$$\begin{aligned}
& p(x_{1:N}, \lambda_{1:N} \mid y_{1:N}) \\
& = p(y_{1:N} \mid x_{1:N}, \lambda_{1:N}) p(x_{1:N}, \lambda_{1:N}) \\
& = \prod_{k=1}^N p(y_k \mid x_k, \lambda_k) p(x_k \mid x_{k-1}, \lambda_k) p(\lambda_k) \quad (16)
\end{aligned}$$

and find its logarithm

$$\begin{aligned}
& \log p(x_{1:N}, \lambda_{1:N} \mid y_{1:N}) \\
& = \sum_{k=1}^N \log p(y_k \mid x_k, \lambda_k) + \log p(x_k \mid x_{k-1}, \lambda_k) + \log p(\lambda_k) \\
& = \sum_{k=1}^N (1 - \lambda_k) \left[ -\frac{1}{2} \log \det R_k - \frac{1}{2} \|y_k - H_k x_k\|_{R_k}^2 \right. \\
& \quad \left. - \frac{1}{2} \log \det Q_{k-1} - \frac{1}{2} \|x_k - F_{k-1} x_{k-1}\|_{Q_{k-1}}^2 + \log(1 - \theta) \right] \\
& \quad + \lambda_k \left[ -\frac{1}{2} \log \det W_k - \frac{1}{2} \|y_k - H_k x_k\|_{W_k}^2 \right. \\
& \quad \left. - \frac{1}{2} \log \det M_{k-1} - \frac{1}{2} \|x_k - F_{k-1} x_{k-1}\|_{M_{k-1}}^2 + \log \theta \right]. \quad (17)
\end{aligned}$$

We compute (14) as

$$\begin{aligned}
& \log q(\lambda_k) = \mathbb{E}_{x_{1:N}, \lambda_{1:N \setminus k}}(\log p(x_{1:N}, \lambda_{1:N} \mid y_{1:N})) + \text{const.} \\
& = (1 - \lambda_k) \mathbb{E}_{x_{k-1:k}} \left( -\frac{1}{2} \log \det R_k - \frac{1}{2} \|y_k - H_k x_k\|_{R_k}^2 \right. \\
& \quad \left. - \frac{1}{2} \log \det Q_{k-1} - \frac{1}{2} \|x_k - F_{k-1} x_{k-1}\|_{Q_{k-1}}^2 + \log(1 - \theta) \right) \\
& \quad + \lambda_k \mathbb{E}_{x_{k-1:k}} \left( -\frac{1}{2} \log \det W_k - \frac{1}{2} \|y_k - H_k x_k\|_{W_k}^2 \right. \\
& \quad \left. - \frac{1}{2} \log \det M_{k-1} - \frac{1}{2} \|x_k - F_{k-1} x_{k-1}\|_{M_{k-1}}^2 + \log(\theta) \right) + \text{const.}
\end{aligned}$$

After some mechanical manipulation, and introducing

$$\begin{aligned}
\log \rho_{k,1} &= -\frac{1}{2} \log \det R_k - \frac{1}{2} \|y_k - H_k x_{k|N}\|_{R_k}^2 \\
&- \frac{1}{2} \text{tr} (H_k^T R_k^{-1} H_k P_{k|N}) - \frac{1}{2} \log \det Q_{k-1} \\
&- \frac{1}{2} \|x_{k|N} - F_{k-1} x_{k-1|N}\|_{Q_{k-1}}^2 \\
&- \frac{1}{2} \text{tr} ((Q_{k-1}^{-1} + F_{k-1} Q_{k-1}^{-1} F_{k-1}^T) \\
&\times (P_{k|N} - 2C_{k|N} F_{k-1}^T + F_{k-1} P_{k-1|N} F_{k-1}^T)) + \log(1 - \theta) \\
\log \rho_{k,2} &= -\frac{1}{2} \log \det W_k - \frac{1}{2} \|y_k - H_k x_{k|N}\|_{W_k}^2 \\
&- \frac{1}{2} \text{tr} (H_k^T W_k^{-1} H_k P_{k|N}) - \frac{1}{2} \log \det M_{k-1} \\
&- \frac{1}{2} \|x_{k|N} - F_{k-1} x_{k-1|N}\|_{M_{k-1}}^2 \\
&- \frac{1}{2} \text{tr} ((M_{k-1}^{-1} + F_{k-1} M_{k-1}^{-1} F_{k-1}^T) \\
&\times (P_{k|N} - 2C_{k|N} F_{k-1}^T + F_{k-1} P_{k-1|N} F_{k-1}^T)) + \log \theta,
\end{aligned}$$

where  $\mathbb{E}_{x_k}(x_k) = x_{k|N}$ ,  $\mathbb{V}(x_k) = P_{k|N}$  and  $\text{cov}(x_k, x_{k-1}) = C_{k|N}$ , the marginal density of the model indicator variable  $\lambda_k$  can be shown to be

$$q(\lambda_k) = (1 - \theta_{k|N})^{1-\lambda_k} \theta_{k|N}^{\lambda_k} = \text{Ber}(\theta_{k|N}) \quad (18)$$

$$\theta_{k|N} = \frac{\rho_{k,2}}{\rho_{k,1} + \rho_{k,2}}. \quad (19)$$

As  $q(\lambda_k)$  is a Bernoulli-distribution, it has the mean  $\mathbb{E}(\lambda_k) = \theta_{k|N}$ .

After finding each of the marginal distributions  $q(\lambda_i)$ , we evaluate the marginal distribution of the state  $q(x_{1:k})$  as

$$\begin{aligned}
\log q(x_{1:N}) &= \mathbb{E}_{\lambda_{1:N}} (\log p(x_{1:N}, \lambda_{1:N} | y_{1:N})) + \text{const.} \\
&= \sum_{k=1}^N -\frac{1}{2} \mathbb{E}_{\lambda_k} (1 - \lambda_k) \|y_k - H_k x_k\|_{R_k}^2 \\
&- \frac{1}{2} \mathbb{E}_{\lambda_k} (\lambda_k) \|y_k - H_k x_k\|_{W_k}^2 \\
&- \frac{1}{2} \mathbb{E}_{\lambda_k} (1 - \lambda_k) \|x_k - F_{k-1} x_{k-1}\|_{Q_{k-1}}^2 \\
&- \frac{1}{2} \mathbb{E}_{\lambda_k} (\lambda_k) \|x_k - F_{k-1} x_{k-1}\|_{M_{k-1}}^2 + \text{const.} \\
&= \sum_{k=1}^N -\frac{1}{2} \|y_k - H_k x_k\|_{\Xi_k}^2 - \frac{1}{2} \|x_k - F_{k-1} x_{k-1}\|_{\Sigma_{k-1}}^2 + \text{const.},
\end{aligned}$$

where

$$\begin{aligned} \Xi_k^{-1} &= \mathbb{E}_{\lambda_k} (1 - \lambda_k) R_k^{-1} + \mathbb{E}_{\lambda_k} (\lambda_k) W_k^{-1} \\ &= (1 - \theta_{k|N}) R_k^{-1} + \theta_{k|N} W_k^{-1} \end{aligned} \quad (20)$$

$$\begin{aligned} \Sigma_{k-1}^{-1} &= \mathbb{E}_{\lambda_k} (1 - \lambda_k) Q_{k-1}^{-1} + \lambda_k (\lambda_k) M_{k-1}^{-1} \\ &= (1 - \theta_{k|N}) Q_{k-1}^{-1} + \theta_{k|N} M_{k-1}^{-1}. \end{aligned} \quad (21)$$

We can see that the density  $q(x_{1:N})$  is a normal distribution and we can compute the marginals

$$q(x_{k-1:k}) = \mathcal{N} \left( \begin{bmatrix} x_{k|N} \\ x_{k-1|N} \end{bmatrix}, \begin{bmatrix} P_{k|N} & C_{k|N} \\ C_{k|N}^T & P_{k-1|N} \end{bmatrix} \right) \quad (22)$$

using KF and RTS-smoother. The set of equations (18) and (22) can be solved by a fixed-point iteration for which the

convergence is guaranteed due to certain convexity properties of the error in the approximative distribution [12, p. 466]. This is the VB method that is summarized in Algorithm 3. Although convergence checks could be performed within the algorithm, we fix the number of iterations to MaxIter to control the computational costs. The resulting algorithm is very close to the EM-method for detecting change in the state transition model [4], [13].

---

**Algorithm 3**  $[x_{0:N|N}, P_{0:N|N}, \theta_{1:N|N}] \leftarrow$   
 $\text{VB}(x_{0|0}, P_{0|0}, \theta_{1:N}, F_{1:N-1}, Q_{0:N-1}, M_{0:N-1}, H_{1:N}, R_{1:N}, W_{1:N})$

---

```

1:  $\theta_{k|N} \leftarrow 0, \quad k = 1, \dots, N$ 
2:  $a^{(1)} \leftarrow -\frac{1}{2} \log \det R_k - \frac{1}{2} \log \det Q_{k-1} + \log(1 - \theta_k)$ 
3:  $a^{(2)} \leftarrow -\frac{1}{2} \log \det W_k - \frac{1}{2} \log \det M_{k-1} + \log \theta_k$ 
4: for  $m = 1, \dots, \text{MaxIter}$  do
5:   for  $k = 0, \dots, N - 1$  do
6:      $\Xi_{k+1}^{-1} \leftarrow (1 - \theta_{k+1|N}) R_{k+1}^{-1} + \theta_{k+1|N} W_{k+1}^{-1}$ 
7:      $\Sigma_k^{-1} \leftarrow (1 - \theta_{k+1|N}) Q_k^{-1} + \theta_{k+1|N} M_k^{-1}$ 
8:      $[x_{k+1|k+1}, P_{k+1|k+1}] \leftarrow$   

       KalmanStep( $x_{k|k}, P_{k|k}, y_k, F_k, \Sigma_k, H_{k+1}, \Xi_{k+1}$ )
9:   end for
10:  for  $j = N - 1, \dots, 0$  do
11:     $[x_{j|N}, P_{j|N}, C_{j+1|N}] \leftarrow$   

      RTSSStep( $x_{j+1|N}, P_{j+1|N}, x_{j|j}, P_{j|j}, F_j, \Sigma_j$ )
12:  end for
13:  for  $i = 1, \dots, N$  do
14:     $\log \rho_{k,1} \leftarrow a^{(1)} - \frac{1}{2} \|y_k - H_k x_{k|N}\|_{R_k}^2$   

     $-\frac{1}{2} \text{tr} (H_k^T R_k^{-1} H_k P_{k|N}) - \frac{1}{2} \|x_{k|N} - F_{k-1} x_{k-1|N}\|_{Q_{k-1}}^2$   

     $-\frac{1}{2} \text{tr} ((Q_{k-1}^{-1} + F_{k-1} Q_{k-1}^{-1} F_{k-1}^T)$   

     $\times (P_{k|N} - 2C_{k|N} F_{k-1}^T + F_{k-1} P_{k-1|N} F_{k-1}^T))$ 
15:     $\log \rho_{k,2} \leftarrow a^{(2)} - \frac{1}{2} \|y_k - H_k x_{k|N}\|_{W_k}^2$   

     $-\frac{1}{2} \text{tr} (H_k^T W_k^{-1} H_k P_{k|N}) - \frac{1}{2} \|x_{k|N} - F_{k-1} x_{k-1|N}\|_{M_{k-1}}^2$   

     $-\frac{1}{2} \text{tr} ((M_{k-1}^{-1} + F_{k-1} M_{k-1}^{-1} F_{k-1}^T)$   

     $\times (P_{k|N} - 2C_{k|N} F_{k-1}^T + F_{k-1} P_{k-1|N} F_{k-1}^T))$ 
16:     $\theta_{k|N} \leftarrow \frac{\rho_{k,2}}{\rho_{k,1} + \rho_{k,2}}$ 
17:  end for
18: end for

```

---

The Algorithm 3 is an offline method for approximating the posterior distribution but can be heuristically modified for online applications. First we choose a window size  $K$ , and then approximate  $p(x_{1:K}, \lambda_{1:K} | y_{1:K})$  using the VB method. Then using  $q(x_K)$  as the prior, we approximate  $p(x_{K+1:2K}, \lambda_{K+1:2K} | y_{1:2K})$  by applying the VB method for the data  $y_{k+1:K}$ . This is repeated at every  $K$ th time step.

#### IV. SIMULATIONS

We simulate two cases of a GPS-positioning problem. In both problems the estimation methods are based on solving the following instance of the system (1)–(3), where

$$F_k = \begin{bmatrix} I_2 & I_2 \\ 0_2 & I_2 \end{bmatrix}, \quad H_k = [I_2 \quad 0_2] \quad (23)$$

$$Q_k = \sigma^2 \begin{bmatrix} \frac{1}{3} I_2 & \frac{1}{2} I_2 \\ \frac{1}{2} I_2 & I_2 \end{bmatrix}, \quad R_k = \begin{bmatrix} 10^2 & \frac{5}{2} \\ \frac{5}{2} & 5^2 \end{bmatrix}, \quad (24)$$

where  $I_2$  and  $0_2$  are  $2 \times 2$  the identity and zero matrices, respectively. If not otherwise mentioned, these parameters are used to generate the simulation data. This “constant velocity” model describes noisy measurements of the planar position of an object whose velocity is a random walk.

#### A. Manoeuvring target

In the example we consider only changes in the state transition model, or maneuvers. We generated 100 tracks with velocities

$$\Delta x_j = \begin{cases} [5 + \nu_j & 0]^T, & j \in [0, 19] \cup [51, 70] \\ [0 & 5 + \nu_j]^T, & j \in [21, 49] \\ [5/\sqrt{2} & 5/\sqrt{2}]^T, & j \in \{20, 50\}, \end{cases} \quad (25)$$

where  $\nu_j \sim N(0, 0.1^2)$  is a white noise process. All the estimation methods model the constant velocity motion with  $\sigma = 0.1$  and the maneuvers with  $M_k = 100 \cdot Q_k$ . The compared methods are the KFs and RTSs using only  $Q_k$  (KF1, RTS1) or  $M_k$  (KF2, RTS2), EM-algorithm (EM) [4], GM filter (GM) with component merging at the each time step [5], the VB algorithm (VB) and the moving window VB (with window size 15) (MWVB), both with 40 iterations. VB and EM methods use the prior  $\theta_k = 0.1$ . From the simulations we investigate the root mean square error (RMSE) and the 95% quantile of the estimation errors, i.e. 95% of the position estimates have error less than the reported 95%-err value. The numbers are reported in Table I. Amongst the online estimation methods, MWVB has the best accuracy and from the offline methods VB has the best accuracy. EM-algorithm seems to be more sensitive to the initial estimates of  $\theta_k = 0$  than VB method, which is the reason for its RMSE performance being worse.

#### B. Change in the observation noise

In the second problem, we generated 100 tracks of 70 time steps. The track is generated with the constant velocity model with  $\sigma = 1$ , and for the observations we simulated batches of observation noise with larger covariance. The observation noise is generated using  $W_k = 25 \cdot R_k$  for time steps  $k \in [20, 30] \cup [50, 60]$ . The compared methods are the KFs and RTSs using only  $R_k$  (KF1, RTS1) or  $W_k$  (KF2, RTS2), EM-algorithm (EM) [4] modified for the problem, GM filter (GM) with component merging at each time step [5], the VB algorithm (VB) and the moving window VB (with window size 15) (WBVB), both with 40 iterations. VB and EM methods use the prior  $\theta_k = 0.1$ . RMSEs and 95% quantile performances are reported in Table I. Again, MWVB has the best accuracy among the online methods and among the offline methods VB has the best accuracy, although the RMSE performance is almost identical to the EM-algorithm.

### V. CONCLUSIONS

A variational Bayes change detection method was described for linear state-space systems with noise processes defined by changing noise covariances. Through simulations it was shown that the method performs very well in offline mode, and that a heuristic online modification of the technique provides good accuracy compared to other methods.

	Test A		Test B	
	RMSE	95%-err	RMSE	95%-err
KF1	13.7	24.9	12.7	41.0
KF2	6.0	12.7	13.5	32.4
GM	7.2	15.1	11.6	37.2
MWVB	4.9	11.6	7.2	19.5
RTS1	9.6	21.5	7.5	21.3
RTS2	3.4	7.1	7.8	17.0
EM	3.6	7.6	5.7	14.1
VB	2.7	6.1	5.6	13.9

TABLE I  
THE SIMULATIONS INDICATE THAT MWVB AND VB METHODS HAVE GOOD ACCURACY AMONG THE ONLINE AND OFFLINE METHODS.

### REFERENCES

- [1] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *Transactions of the ASME-Journal of Basic Engineering*, vol. 82, 1960.
- [2] D. L. Alspach and H. W. Sorenson, “Nonlinear Bayesian estimation using Gaussian sum approximations,” *IEEE Transactions on Automatic Control*, vol. AC-17, no. 4, pp. 439–448, August 1972.
- [3] H. Sorenson and D. Alspach, “Recursive Bayesian estimation using Gaussian sums,” *Automatica*, vol. 7, pp. 465–479, 1971.
- [4] N. Bergman and F. Gustafsson, “Three statistical batch algorithms for tracking manoeuvring targets,” Department of Electrical Engineering, Linköping University, Tech. Rep., 1999.
- [5] D. Peña and I. Guttman, “Optimal collapsing of mixture distributions in robust recursive estimation,” *Communications in Statistics: Theory and Methods*, vol. 18, no. 3, pp. 817–833, 1989.
- [6] F. Gustafsson, *Adaptive Filtering and Change Detection*. John Wiley & Sons Ltd., 2000.
- [7] M. Basseville and I. V. Nikiforov, *Detection of abrupt changes: theory and applications*, ser. Information and system science series. Prentice Hall, NJ, 1993.
- [8] A. S. Willsky, “A survey of design methods for failure detection in dynamic systems,” *Automatica*, vol. 12, pp. 601–611, 1976.
- [9] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, Inc., 2001.
- [10] C. J. Masreliez and R. D. Martin, “Robust Bayesian estimation for the linear model and robustifying the Kalman filter,” *IEEE Transactions on Automatic Control*, vol. 22, no. 3, pp. 361–271, 1977.
- [11] H. E. Rauch, F. Tung, and C. T. Striebel, “Maximum likelihood estimates of linear dynamic systems,” *AIAA Journal*, vol. 3, no. 8, pp. 1445–1450, August 1965.
- [12] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer Science+Business Media, LLC, 2006.
- [13] V. Krishnamurthy and J. B. Moore, “On-line estimation of hidden Markov model parameters based on the Kullback-Leibler information measure,” *IEEE Transactions on Signal Processing*, vol. 41, no. 8, pp. 2557–2573, August 1993.



Tampereen teknillinen yliopisto  
PL 527  
33101 Tampere

Tampere University of Technology  
P.O.B. 527  
FI-33101 Tampere, Finland