NON-LINEAR NOISE ADAPTIVE KALMAN FILTERING VIA VARIATIONAL BAYES

Simo Särkkä

Aalto University, 02150 Espoo, Finland

Jouni Hartikainen

Rocsole Ltd., 70211 Kuopio, Finland

ABSTRACT

We consider joint estimation of state and time-varying noise covariance matrices in non-linear stochastic state space models. We propose a variational Bayes and Gaussian non-linear filtering based algorithm for efficient computation of the approximate filtering posterior distributions. The formulation allows the use of efficient Gaussian integration methods such as unscented transform, cubature integration and Gauss-Hermite integration along with the classical Taylor series approximations. The performance of the algorithm is illustrated in a simulated application.

Index Terms— variational Bayes, unknown noise covariance, adaptive filtering, non-linear Kalman filtering

1. INTRODUCTION

In this paper, we propose a method for Bayesian inference on the state x_k and noise covariances Σ_k in heteroscedastic non-linear stochastic state space models (see, e.g., [1]) of the form

$$\begin{aligned} x_k &\sim \mathrm{N}\left(f(x_{k-1}), Q_k\right) \\ y_k &\sim \mathrm{N}\left(h(x_k), \Sigma_k\right) \\ \Sigma_k &\sim p(\Sigma_k \,|\, \Sigma_{k-1}), \end{aligned} \tag{1}$$

where $x_k \in \mathbb{R}^n$ is the state at time step k, and $y_k \in \mathbb{R}^d$ is the measurement, Q_k is the known process noise covariance and Σ_k is the measurement noise covariance. The non-linear functions $f(\cdot)$ and $h(\cdot)$ form the dynamic and measurement models, respectively, and the last equation defines the Markovian dynamic model for the dynamics of the unknown noise covariances Σ_k. We aim at computing the joint posterior (filtering) distribution of the states and noise covariances $p(x_k, \Sigma_k \,|\, y_{1:k})$. Although the formal Bayesian solution to this problem is well-known (see, e.g., [1]), it is computationally intractable and we can only approximate it.

In a recent article, Särkkä and Nummenmaa [2] introduced the variational Bayesian (VB) adaptive Kalman filter (VB-AKF), which can be used for estimating the measurement noise variances along with the state in linear state space models. In this paper, we extend the method to allow estimation of the full noise covariance matrix and non-linear state space models. The idea is similar to what was recently used by Piché et al. [3] in the context of outlier-robust filtering, which in turn is based on the linear results of [4]. VB methods have been applied to parameter identification in state space models also in [5, 6, 7] and various other (Bayesian) approaches have can be found, for example, in references [8, 9, 10, 11, 12].

1.1. Gaussian Filtering

If the covariances in the model (1) were known, the filtering problem would reduce to the classical non-linear (Gaussian) optimal filtering problem [13, 8, 14, 1]. This non-linear filtering problem can be solved in various ways, but one quite general approach is the Gaussian filtering approach [8, 15, 16], where the idea is to assume that the filtering distribution is approximately Gaussian. That is, we assume that there exist means m_k and covariances P_k such that $p(x_k \mid y_{1:k}) \approx \mathrm{N}(x_k \mid m_k, P_k)$.

The Gaussian filter prediction and update steps can be written as follows [15]:

- Prediction:

$$\begin{aligned} m_k^- &= \int f(x_{k-1}) \, \mathrm{N}(x_{k-1} \,|\, m_{k-1}, P_{k-1}) \, \mathrm{d}x_{k-1} \\ P_k^- &= \int (f(x_{k-1}) - m_k^-) \, (f(x_{k-1}) - m_k^-)^T \\ &\quad \times \mathrm{N}(x_{k-1} \,|\, m_{k-1}, P_{k-1}) \, \mathrm{d}x_{k-1} + Q_k. \end{aligned} \tag{2}$$

- Update:

$$\begin{aligned} \mu_k &= \int h(x_k) \, \mathrm{N}(x_k \,|\, m_k^-, P_k^-) \, \mathrm{d}x_k \\ S_k &= \int (h(x_k) - \mu_k) \, (h(x_k) - \mu_k)^T \\ &\quad \times \mathrm{N}(x_k \,|\, m_k^-, P_k^-) \, \mathrm{d}x_k + \Sigma_k \\ C_k &= \int (x_k - m^-) \, (h(x_k) - \mu_k)^T \\ &\quad \times \mathrm{N}(x_k \,|\, m_k^-, P_k^-) \, \mathrm{d}x_k \\ K_k &= C_k \, S_k^{-1} \\ m_k &= m_k^- + K_k \, (y_k - \mu_k) \\ P_k &= P_k^- - K_k \, S_k \, K_k^T. \end{aligned} \tag{3}$$

978-1-4673-1026-0/12/$31.00

With different selections for the Gaussian integral approximations, we get different filtering algorithms [16] such as the unscented Kalman filter (UKF) [17], Gauss-Hermite Kalman filter (GHKF) [15], cubature Kalman filter (CKF) [18], and various others [19, 20, 21] along with the classical methods [13, 8].

1.2. Variational Approximation

In this paper, we approximate the joint filtering distribution of the state and covariance matrix with the free-form variational Bayesian (VB) approximation (see, *e.g.*, [22, 23, 24, 5]):

$$p(x_k, \Sigma_k \,|\, y_{1:k}) \approx Q_x(x_k)\, Q_\Sigma(\Sigma_k), \tag{4}$$

where $Q_x(x_k)$ and $Q_\Sigma(\Sigma_k)$ are the yet unknown approximating densities. The VB approximation can be formed by minimizing the Kullback-Leibler (KL) divergence between the true distribution and the approximation:

$$\begin{aligned}&\mathrm{KL}[Q_x(x_k)\, Q_\Sigma(\Sigma_k) \,||\, p(x_k, \Sigma_k \,|\, y_{1:k})] \\ &= \int Q_x(x_k)\, Q_\Sigma(\Sigma_k) \log\left(\frac{Q_x(x_k)\, Q_\Sigma(\Sigma_k)}{p(x_k, \Sigma_k \,|\, y_{1:k})}\right) \mathrm{d}x_k\, \mathrm{d}\Sigma_k.\end{aligned}$$

Minimizing the KL divergence with respect to the probability densities, we get the following equations:

$$\begin{aligned}Q_x(x_k) &\propto \exp\left(\int \log p(y_k, x_k, \Sigma_k \,|\, y_{1:k-1})\, Q_\Sigma(\Sigma_k)\, \mathrm{d}\Sigma_k\right) \\ Q_\Sigma(\Sigma_k) &\propto \exp\left(\int \log p(y_k, x_k, \Sigma_k \,|\, y_{1:k-1})\, Q_x(x_k)\, \mathrm{d}x_k\right).\end{aligned} \tag{5}$$

The solutions to these equations can be found by a fixed-point iteration for the sufficient statistics of the approximating densities.

2. VARIATIONAL BAYESIAN ADAPTATION OF NOISE COVARIANCE

2.1. Estimation of Full Covariance in Linear Case

We start by considering the linear state space model with unknown covariance as follows:

$$\begin{aligned}p(x_k \,|\, x_{k-1}) &= \mathrm{N}(x_k \,|\, A_k\, x_{k-1}, Q_k) \\ p(y_k \,|\, x_k, \Sigma_k) &= \mathrm{N}(x_k \,|\, H_k\, x_k, \Sigma_k),\end{aligned} \tag{6}$$

where A_k and H_k are some known matrices. We assume that the dynamic model for the covariance is independent of the state and of the Markovian form $p(\Sigma_k \,|\, \Sigma_{k-1})$, and set some restrictions to it shortly. In this section we follow the derivation in [2], and extend the scalar variance case to the full covariance case.

Assume that the filtering distribution of the time step $k-1$ can be approximated as product of Gaussian distribution and inverse Wishart (IW) distribution as follows:

$$\begin{aligned}&p(x_{k-1}, \Sigma_{k-1} \,|\, y_{1:k-1}) = \\ &\qquad \mathrm{N}(x_{k-1} \,|\, m_{k-1}, P_{k-1})\, \mathrm{IW}(\Sigma_{k-1} \,|\, \nu_{k-1}, V_{k-1}),\end{aligned} \tag{7}$$

where the densities, up to non-essential normalization terms, can be written as [25]:

$$\begin{aligned}\mathrm{N}(x \,|\, m, P) &\propto |P|^{-1/2} \exp\left(-\frac{1}{2}(x-m)^T\, P^{-1}\,(x-m)\right) \\ \mathrm{IW}(\Sigma \,|\, \nu, V) &\propto |\Sigma|^{-(\nu+n+1)/2} \exp\left(-\frac{1}{2}\mathrm{tr}\left(V\,\Sigma^{-1}\right)\right).\end{aligned}$$

That is, in the VB approximation (4), $Q_x(x_k)$ is the Gaussian distribution and $Q_\Sigma(\Sigma_k)$ is the inverse Wishart distribution.

We now assume that the dynamic model for the covariance is of such form that it maps an inverse Wishart distribution at the previous step into inverse Wishart distribution at the current step. This gives (cf. [2])

$$\begin{aligned}p(\Sigma_k \,|\, y_{1:k-1}) &= \mathrm{IW}(\Sigma_k \,|\, \nu_k^-, V_k^-) \\ p(x_k \,|\, y_{1:k-1}) &= \mathrm{N}(x_k \,|\, m_k^-, P_k^-),\end{aligned}$$

where ν_k^- and V_k^- are certain parameters (see Section 2.2), and m_k^- and P_k^- are given by the standard Kalman filter prediction equations:

$$\begin{aligned}m_k^- &= A_k\, m_{k-1} \\ P_k^- &= A_k\, P_{k-1}\, A_k^T + Q_k.\end{aligned} \tag{8}$$

Because the distribution and the previous step is separable, and the dynamic models are independent we thus get the following joint predicted distribution:

$$\begin{aligned}&p(x_k, \Sigma_k \,|\, y_{1:k-1}) = \\ &\qquad \mathrm{N}(x_k \,|\, m_k^-, P_k^-)\, \mathrm{IW}(\Sigma_k \,|\, \nu_k^-, V_k^-).\end{aligned} \tag{9}$$

We are now ready to form the actual VB approximation to the posterior. The integrals in the exponentials of (5) can now be expanded as follows (cf. [2]):

$$\begin{aligned}&\int \log p(y_k, x_k, \Sigma_k \,|\, y_{1:k-1})\, Q_\Sigma(\Sigma_k)\, \mathrm{d}\Sigma_k \\ &\quad = -\frac{1}{2}(y_k - H_k\, x_k)^T \langle \Sigma_k^{-1} \rangle_\Sigma (y_k - H_k\, x_k) \\ &\qquad - \frac{1}{2}(x_k - m_k^-)^T \left(P_k^-\right)^{-1} (x_k - m_k^-) + C_1 \\ &\int \log p(y_k, x_k, \Sigma_k \,|\, y_{1:k-1})\, Q_x(x_k)\, \mathrm{d}x_k \\ &\quad = -\frac{1}{2}(\nu_k^- + n + 2)\, \log|\Sigma_k| - \frac{1}{2}\mathrm{tr}\left\{V_k^-\, \Sigma_k^{-1}\right\} \\ &\qquad - \frac{1}{2}\langle (y_k - H_k\, x_k)^T\, \Sigma_k^{-1}\, (y_k - H_k\, x_k) \rangle_x + C_2,\end{aligned} \tag{10}$$

where $\langle\cdot\rangle_\Sigma = \int(\cdot)\, Q_\Sigma(\Sigma_k)\,\mathrm{d}\Sigma_k$, $\langle\cdot\rangle_x = \int(\cdot)\, Q_x(x_k)\,\mathrm{d}x_k$, and C_1, C_2 are some constants. If we have that $Q_\Sigma(\Sigma_k) = \mathrm{IW}(\Sigma_k \mid \nu_k, V_k)$, then the expectation in the first equation of (10) is

$$\langle \Sigma_k^{-1} \rangle_\Sigma = (\nu_k - n - 1) V_k^{-1}. \tag{11}$$

Furthermore, if $Q_x(x_k) = \mathrm{N}(x_k \mid m_k, P_k)$, then the expectation in the second equation of (10) becomes

$$\begin{aligned} &\langle (y_k - H_k\, x_k)^T\, \Sigma_k^{-1}\, (y_k - H_k\, x_k) \rangle_x \\ &\quad = \mathrm{tr}\left\{ H_k\, P_k\, H_k^T\, \Sigma_k^{-1} \right\} \\ &\quad + \mathrm{tr}\left\{ (y_k - H_k\, m_k)\, (y_k - H_k\, m_k)^T\, \Sigma_k^{-1} \right\}. \end{aligned} \tag{12}$$

By substituting the expectations (11) and (12) into (10) and matching terms in left and right hand sides of (5) results in the following coupled set of equations:

$$\begin{aligned} S_k &= H_k\, P_k^-\, H_k^T + (\nu_k - n - 1)^{-1}\, V_k \\ K_k &= P_k^-\, H_k^T\, S_k^{-1} \\ m_k &= m_k^- + K_k\, (y_k - H_k\, m_k^-) \\ P_k &= P_k^- - K_k\, S_k\, K_k^T \\ \nu_k &= \nu_k^- + 1 \\ V_k &= V_k^- + H_k\, P_k\, H_k^T + (y_k - H_k\, m_k)\, (y_k - H_k\, m_k)^T. \end{aligned} \tag{13}$$

The first four of the equations have been written into such suggestive form that they can easily be recognized to be the Kalman filter update step equations with measurement noise covariance $(\nu_k - n - 1)^{-1}\, V_k$.

2.2. Dynamic Model for Covariance

In analogous manner to [2], the dynamic model $p(\Sigma_k \mid \Sigma_{k-1})$ needs to be chosen such that when it is applied to an inverse Wishart distribution, it produces another inverse Wishart distribution. Although, the explicitly construction of the density is hard, all we need to do is to postulate a transformation rule for the sufficient statistics of the inverse Wishart distributions at the prediction step. Using similar heuristics as in [2], we arrive at the following dynamic model:

$$\begin{aligned} \nu_k^- &= \rho\, (\nu_{k-1} - n - 1) + n + 1 \\ V_k^- &= B\, V_{k-1}\, B^T, \end{aligned} \tag{14}$$

where ρ is a real number $0 < \rho \le 1$ and B is a matrix $0 < |B| \le 1$. A reasonable choice for the matrix is $B = \sqrt{\rho}\, I$, in which case parameter ρ controls the assumed dynamics: value $\rho = 1$ corresponds to stationary covariance and lower values allow for higher time-fluctuations. The resulting multidimensional variational Bayesian adaptive Kalman filter (VB-AKF) is shown in Algorithm 1.

Predict: Compute the parameters of the predicted distribution as follows:

$$\begin{aligned} m_k^- &= A_k\, m_{k-1} \\ P_k^- &= A_k\, P_{k-1}\, A_k^T + Q_k \\ \nu_k^- &= \rho\, (\nu_{k-1} - n - 1) + n + 1 \\ V_k^- &= B\, V_{k-1}\, B^T, \end{aligned}$$

Update: First set $m_k^{(0)} = m_k^-$, $P_k^{(0)} = P_k^-$, $\nu_k = 1 + \nu_k^-$, and $V_k^{(0)} = V_k^-$ and the iterate the following, say N, steps $i = 1, \ldots, N$:

$$\begin{aligned} S_k^{(i+1)} &= H_k\, P_k^-\, H_k^T + (\nu_k - n - 1)^{-1}\, V_k^{(i)} \\ K_k^{(i+1)} &= P_k^-\, H_k^T\, [S_k^{(i+1)}]^{-1} \\ m_k^{(i+1)} &= m_k^- + K_k^{(i+1)}\, (y_k - H_k\, m_k) \\ P_k^{(i+1)} &= P_k^- - K_k^{(i+1)}\, S_k^{(i+1)}\, [K_k^{(i+1)}]^T \\ V_k^{(i+1)} &= V_k^- + H_k\, P_k^{(i+1)}\, H_k^T \\ &\quad + (y_k - H_k\, m_k^{(i+1)})\, (y_k - H_k\, m_k^{(i+1)})^T \end{aligned}$$

and set $V_k = V_k^{(N)}$, $m_k = m_k^{(N)}$, $P_k = P_k^{(N)}$.

Algorithm 1: The multidimensional Variational Bayesian Adaptive Kalman Filter (VB-AKF) algorithm

2.3. Extension to Non-Linear Models

In this section we extend the results in the previous section into non-linear models of the form (1). We again start with the assumption that the filtering distribution is approximately product of a Gaussian term and inverse Wishart (IW) term as in Equation (7). The prediction step can be handled in similar manner as in the linear case, except that the computation of the mean and covariance of the state should be done with the Gaussian filter prediction equations (2) instead of the Kalman filter prediction equations (8). The inverse Wishart part of the prediction remains intact. After the prediction step, the approximation is again a product of Gaussian and inverse Wishart distributions as in Equation (9).

The expressions corresponding to (10) now become:

$$\begin{aligned} &\int \log p(y_k, x_k, \Sigma_k \mid y_{1:k-1})\, Q_\Sigma(\Sigma_k)\,\mathrm{d}\Sigma_k \\ &\quad = -\frac{1}{2}(y_k - h(x_k))^T \langle \Sigma_k^{-1} \rangle_\Sigma (y_k - h(x_k)) \\ &\qquad - \frac{1}{2}(x_k - m_k^-)^T \left(P_k^-\right)^{-1} (x_k - m_k^-) + C_1 \\ &\int \log p(y_k, x_k, \Sigma_k \mid y_{1:k-1})\, Q_x(x_k)\,\mathrm{d}x_k \\ &\quad = -\frac{1}{2}(\nu_k^- + n + 2)\, \log|\Sigma_k| - \frac{1}{2}\mathrm{tr}\left\{ V_k^-\, \Sigma_k^{-1} \right\} \\ &\qquad - \frac{1}{2}\langle (y_k - h(x_k))^T\, \Sigma_k^{-1}\, (y_k - h(x_k)) \rangle_x + C_2. \end{aligned} \tag{15}$$

The expectation in the first equation is still given by the equation (11), but the resulting distribution in terms of x_k is intractable in closed form due to the non-linearity $h(x_k)$. Fortunately, the approximation problem is exactly the same as encountered in the update step of Gaussian filter and thus we can directly use the equations (3) for computing Gaussian approximation to the distribution.

The simplification (12) does not work in the non-linear case, but we can rewrite the expectation as

$$\begin{aligned} &\langle (y_k - h(x_k))^T \, \Sigma_k^{-1} \, (y_k - h(x_k)) \rangle_x \\ &\quad = \mathrm{tr} \left\{ \langle (y_k - h(x_k)) \, (y_k - h(x_k))^T \rangle_x \, \Sigma_k^{-1} \right\}, \end{aligned} \tag{16}$$

where the expectation can be separately computed using some of the Gaussian integration methods in [16]. Because the result of the integration is just a constant matrix, we can now substitute (11) and (16) into (15) and match the terms in equations (5) in the same manner as in linear case. This results in equations which consist of the Gaussian filter update step (3) with measurement noise $\Sigma_k = (\nu_k - n - 1)^{-1} \, V_k$ along with the following two additional equations:

$$\begin{aligned} \nu_k &= \nu_k^- + 1 \\ V_k &= V_k^- + \int (y_k - h(x_k)) \, (y_k - h(x_k))^T \\ &\qquad \times \mathrm{N}(x_k \,|\, m_k, P_k) \, \mathrm{d}x_k. \end{aligned} \tag{17}$$

2.4. The Adaptive Filtering Algorithm

The general filtering method for the full covariance and non-linear state space model is shown in Algorithm 2. Various useful special cases and extensions can be deduced from the equations:

- The *Gaussian integration method* [15, 16, 17, 18, 26, 19, 20, 21] will result in different variants of the algorithm. For example, the Taylor series based approximation could be called VB-AEKF, unscented transform based method VB-AUKF, cubature based VB-ACKF, Gauss-Hermite based VB-AGHKF and so on.

- The *diagonal covariance case*, which was considered in [2], can be recovered by updating only the diagonal elements in the last equation of the algorithm and keeping all other elements in the matrices $V_k^{(i)}$ zero. The matrix B in the prediction step then needs to be diagonal also. Although the inverse Wishart parametrization does not reduce to the inverse Gamma parametrization, the formulations are equivalent.

- *Non-additive dynamic models* can be handled by simply replacing the state prediction with the non-additive counterpart.

Predict: Compute the parameters of the predicted distribution as follows:

$$\begin{aligned} m_k^- &= \int f(x_{k-1}) \, \mathrm{N}(x_{k-1} \,|\, m_{k-1}, P_{k-1}) \, \mathrm{d}x_{k-1} \\ P_k^- &= \int (f(x_{k-1}) - m_k^-) \, (f(x_{k-1}) - m_k^-)^T \\ &\qquad \times \mathrm{N}(x_{k-1} \,|\, m_{k-1}, P_{k-1}) \, \mathrm{d}x_{k-1} + Q_k \\ \nu_k^- &= \rho \, (\nu_{k-1} - n - 1) + n + 1 \\ V_k^- &= B \, V_{k-1} \, B^T, \end{aligned}$$

Update: First set $m_k^{(0)} = m_k^-$, $P_k^{(0)} = P_k^-$, $\nu_k = 1 + \nu_k^-$, and $V_k^{(0)} = V_k^-$ and precompute the following:

$$\begin{aligned} \mu_k &= \int h(x_k) \, \mathrm{N}(x_k \,|\, m_k^-, P_k^-) \, \mathrm{d}x_k \\ T_k &= \int (h(x_k) - \mu_k) \, (h(x_k) - \mu_k)^T \, \mathrm{N}(x_k \,|\, m_k^-, P_k^-) \, \mathrm{d}x_k \\ C_k &= \int (x_k - m^-) \, (h(x_k) - \mu_k)^T \, \mathrm{N}(x_k \,|\, m_k^-, P_k^-) \, \mathrm{d}x_k \end{aligned}$$

Iterate the following, say N, steps $i = 1, \ldots, N$:

$$\begin{aligned} S_k^{(i+1)} &= T_k + (\nu_k - n - 1)^{-1} \, V_k^{(i)} \\ K_k^{(i+1)} &= C_k \, [S_k^{(i+1)}]^{-1} \\ m_k^{(i+1)} &= m_k^- + K_k^{(i+1)} \, (y_k - \mu_k) \\ P_k^{(i+1)} &= P_k^- - K_k^{(i+1)} \, S_k^{(i+1)} \, [K_k^{(i+1)}]^T \\ V_k^{(i+1)} &= V_k^- + \int (y_k - h(x_k)) \, (y_k - h(x_k))^T \\ &\qquad \times \mathrm{N}(x_k \,|\, m_k^{(i+1)}, P_k^{(i+1)}) \, \mathrm{d}x_k. \end{aligned}$$

and set $V_k = V_k^{(N)}$, $m_k = m_k^{(N)}$, $P_k = P_k^{(N)}$.

Algorithm 2: The Variational Bayesian Adaptive Gaussian Filter (VB-AGF) algorithm

3. NUMERICAL RESULTS

3.1. Multi-Sensor Bearings Only Tracking

As an example, we consider the classical multi-sensor bearings only tracking problem with coordinated turning model [14], where the state $x = (u, \dot{u}, v, \dot{v}, \omega)$ contains the 2d location (u, v) and the corresponding velocities $(\dot{u}, \dot{v})$ as well as the turning rate ω of the target. The dynamic model was the coordinated turn model and the measurements consisted of bearings reported by four sensors with unknown (joint) noise covariance matrix.

We simulated a trajectory and measurements from the model and applied different filters to it. We tested various

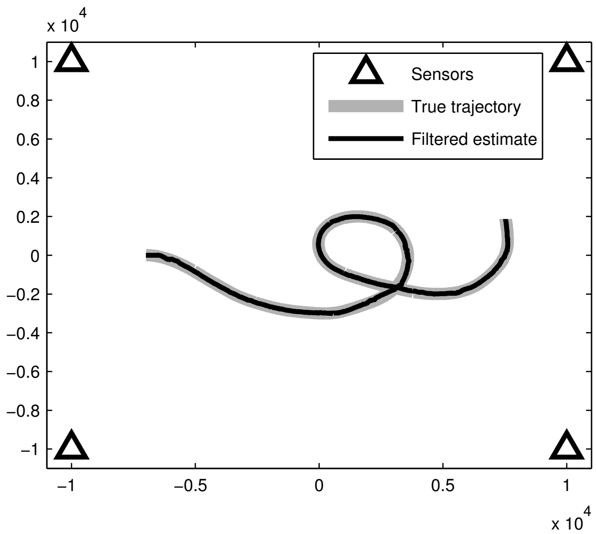

Fig. 1. The simulated trajectory and the estimate obtained with VB-ACKF.

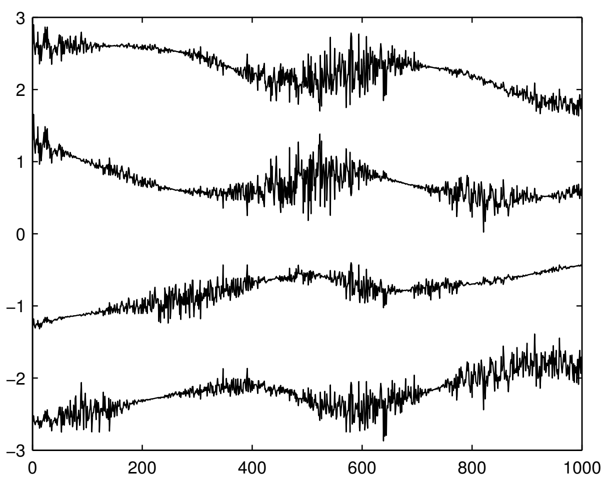

Fig. 2. The simulated measurements.

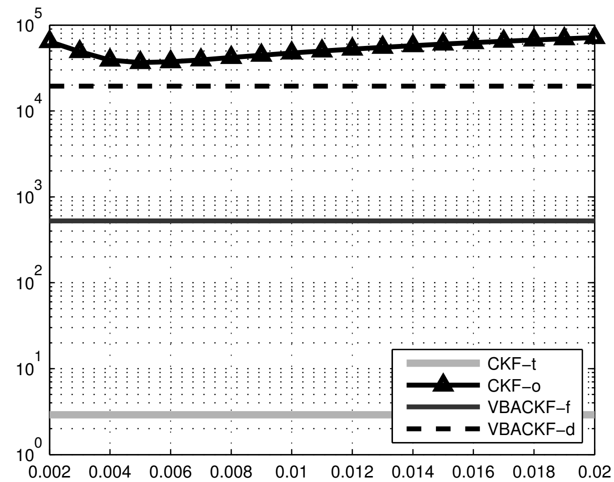

Fig. 3. Root mean squared errors (RMSEs) for different methods.

Gaussian integration based methods (VB-AEKF, VB-AUKF, VB-ACKF, VB-AGHKF) and because the results were quite much the same with different Gaussian integration methods, we only present the results obtained with VB-ACKF. Figure 1 shows the simulated trajectory and the VB-ACKF results with the full covariance estimation. In the simulation, the variances of the measurement noises as well as the cross-correlations varied smoothly over time. The simulated measurements are shown in Figure 2.

Figure 3 shows the root mean squared errors (RMSEs) for CKF with the true covariance matrix (CKF-t), CKF with a diagonal covariance matrix with diagonal elements given by the value on the x-axis (CKF-o), CKF with full covariance estimation (VBCKF-f), and CKF with diagonal covariance estimation (VBCKF-d). As can be seen in the figure, the results of filters with covariance estimation are indeed better than the results of any filter with fixed diagonal covariance matrix. The filter with the known covariance matrix gives the lowest error, as would be expected, and the filter with full covariance estimation gives a lower error than the filter with diagonal covariance estimation.

4. CONCLUSION AND DISCUSSION

In this paper, we have presented a variational Bayes and nonlinear Gaussian (Kalman) filtering based algorithm for joint estimation of state and time-varying noise covariances in nonlinear state space models. The performance of the method has been illustrated in a simulated application.

There are several extensions that could be considered as well. For instance, we could try to estimate the process noise covariance in the model. However, it is not as easy as it sounds, because the process noise covariance does not appear in the equations in such simple conjugate form as the measurement noise covariance. Another natural extension would be the case of smoothing (cf. [3]). Unfortunately the current dynamic model makes things challenging, because we do not know the actual transition density at all. This makes the implementation of a Rauch–Tung–Striebel type of smoother impossible—although a simple smoothing estimate for the state can be obtained by simply running the RTS smoother over the state estimates while ignoring the noise covariance estimates completely. However, it would be possible to construct an approximate two-filter smoother for the full state space model, but even in that case we need to put some more constraints to the model, for example, assume that the covariance dynamics are time-reversible.

5. REFERENCES

[1] S. Särkkä, *Bayesian Filtering and Smoothing*, Cambridge University Press, 2013.

[2] S. Särkkä and A. Nummenmaa, "Recursive noise adaptive Kalman filtering by variational Bayesian approximations," *IEEE Transactions on Automatic Control*, vol. 54(3), pp. 596–600, 2009.

[3] R. Piché, S. Särkkä, and J. Hartikainen, "Recursive outlier-robust filtering and smoothing for nonlinear systems using the multivariate Student-t distribution," in *Proceedings of MLSP*, 2012.

[4] G. Agamennoni, J.I. Nieto, and E.M. Nebot, "An outlier-robust Kalman filter," in *IEEE Int. Conf. on Robotics and Automation (ICRA)*, May 2011, pp. 1551–1558.

[5] V. Smidl and A. Quinn, *The Variational Bayes Method in Signal Processing*, Springer, 2006.

[6] M. J. Beal and Z. Ghahramani, "The variational Kalman smoother," Tech. Rep. TR01-003, Gatsby Unit, 2001.

[7] H. Valpola, M. Harva, and J. Karhunen, "Hierarchical models of variance sources," *Signal Processing*, vol. 84(2), pp. 267–282, 2004.

[8] P. Maybeck, *Stochastic Models, Estimation and Control, Volume 2*, Academic Press, 1982.

[9] X.-R. Li and Y. Bar-Shalom, "A recursive multiple model approach to noise identification," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 30(3), July 1994.

[10] G. Storvik, "Particle filters in state space models with the presence of unknown static parameters," *IEEE Transactions on Signal Processing*, vol. 50(2), pp. 281–289, 2002.

[11] P. M. Djuric and J. Miguez, "Sequential particle filtering in the presence of additive Gaussian noise with unknown parameters," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Orlando, FL*, 2002.

[12] S. Särkkä, "On sequential Monte Carlo sampling of discretely observed stochastic differential equations," in *Proceedings of the Nonlinear Statistical Signal Processing Workshop*, September 2006, [CDROM].

[13] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, 1970.

[14] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*, Wiley Interscience, 2001.

[15] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Transactions on Automatic Control*, vol. 45, no. 5, pp. 910–927, May 2000.

[16] Y. Wu, D. Hu, M. Wu, and X. Hu, "A numerical-integration perspective on Gaussian filters," *IEEE Transactions on Signal Processing*, vol. 54, no. 8, pp. 2910–2921, August 2006.

[17] S. J. Julier, J. K. Uhlmann, and H. F. Durrant-Whyte, "A new method for the nonlinear transformation of means and covariances in filters and estimators," *IEEE Transactions on Automatic Control*, vol. 45, no. 3, pp. 477–482, March 2000.

[18] I. Arasaratnam and S. Haykin, "Cubature Kalman filters," *IEEE Transactions on Automatic Control*, vol. 54, no. 6, pp. 1254–1269, June 2009.

[19] M. Nørgaard, N. K. Poulsen, and O. Ravn, "New developments in state estimation for nonlinear systems," *Automatica*, vol. 36, no. 11, pp. 1627 – 1638, 2000.

[20] T. Lefebvre, H. Bruyninckx, and J. De Schutter, "Comment on "a new method for the nonlinear transformation of means and covariances in filters and estimators" [and authors' reply]," *IEEE Transactions on Automatic Control*, vol. 47, no. 8, pp. 1406–1409, August 2002.

[21] M. P. Deisenroth, M. F. Huber, and U. D. Hanebeck, "Analytic moment-based Gaussian process filtering," in *Proceedings of the 26th International Conference on Machine Learning*, 2009.

[22] T. S. Jaakkola, "Tutorial on variational approximation methods," in *Advanced Mean Field Methods – Theory and Practice*, M. Opper and D. Saad, Eds. 2001, pp. 129–159, MIT Press.

[23] M. J. Beal, *Variational Algorithms for Approximate Bayesian Inference*, Ph.D. thesis, Gatsby Computational Neuroscience Unit, University College London, 2003.

[24] H. Lappalainen and J. W. Miskin, "Ensemble learning," in *Advances in Independent Component Analysis*, M. Girolami, Ed. 2000, pp. 75–92, Springer-Verlag.

[25] A. Gelman, J. B. Carlin, H. S. Stern, and D. R. Rubin, *Bayesian Data Analysis*, Chapman & Hall, 2004.

[26] J. H. Kotecha and P. M. Djuric, "Gaussian particle filtering," *IEEE Transactions on Signal Processing*, vol. 51, no. 10, pp. 2592–2601, October 2003.